Proceedings of the 17th International Conference on Knowledge Management

"Knowledge, Uncertainty and Risks: From individual to global scale"

23-24 June 2022

Editor

Prof. Dr.-Ing. Peter Heisig
University of Applied Sciences - FH Potsdam

Preface

At the time of writing this Call-for-Papers most of us are sitting at home in our home offices undertaking research projects and teaching our students with online tools. All stakeholders have had to adapt quickly and had to master changes which sometimes happened on a weekly basis. The global dimension of this pandemic and the fight undertaken by the different actors have evolved nearly in real-time given the global media coverage. While this pandemic came within days and weeks around the globe, another change is more subtle and even more hidden to direct human perception except those who are affected directly at the moment by the change of the climate in different regions.

From both developments we can see how the gathering of data, its transformation into information and the development of knowledge as well as the discussions about the 'right' actions are difficult processes with several iterations and setbacks on very different levels in society. From individual behavior towards organizational behavior national and international or global behaviors, all levels are interwoven and interdependent. Finally, this crisis shows us the uncertainties and risks we face as individuals, as organizations and as society if we don't have the right information and knowledge or we don't believe the information and knowledge provided by authorities and other stakeholders.

Therefore, the overall theme of the 17th edition of the ICKM conference will be on “Knowledge, Uncertainty and Risks: From individual to global scale” at different levels of analysis and agency.

Table of Contents

Acknowledgements

Every event is a team effort and an international event is a huge effort by a distributed team around the globe. Every member contributes from spreading the word in her/his community via reviewing submissions made to the Call-for-Papers up to supporting the organization of the conference.

Our big thank you goes to all members of the conference organizing committee, all reviewers and authors contributing to the 17th ICKM 2022 in Potsdam!

Conference Organizing Committee

General Chair:
Dr. Suliman Hawamdeh

Conference Chair:
Dr. Peter Heisig

Co-Chairs:
Dr. Daniel Alemneh, Dr. Jeff Allen, Andreas Matern, Waltraut Ritter, Deborah Swain

ICKM Program Committee

Naresh Agarwal, Dr. Daniel Alemneh, Rose Baker, Franz Barachini, Dr. Andreas Brandner, Hsia-Ching Chang, Alton Chua, Kimiz Dalkir, Gertrudes Dandolini, Patricia de Sá Freire, Dursun Delen, Pranathy Enamela, Anwarul Islam, Yong Mi Kim, Jay Liebowitz, Joyline Makani, Tereza Merlo, Dr. Günther Neher, Dr. Angela Schreyer, Christian Stary, Dr. Julia Maria Struß, Oksana Zavalina, Vyacheslav Zavalin.

As the local conference host, I would like to extend my great thanks to all the local colleagues, members of FHP staff and the students from our Department of Information Sciences (FB5) for their help and support to make this event on our campus possible.

I am very grateful to everybody for engaging in such an excellent manner and particular to those who put even more hours in as to be expected.

FHP Staff

Ulrike Weichelt
Robert Rother
Dorothee Gümpel
Dörte Behrend
Anna-Maria Lentz
Reimar Kobi
Axel Thiele
Benjamin Schippling
Julia Maria Struß
Andreas Lüdtke
Katharina Krüth
Leonie Schatter
Katrin Gabriel
Daniel Wolter
Stephan Alexander Schier
Christoph Klein
André Schubert
Günther Neher
Alexander Doss
Iris Mahnke
Birgit Bröse
Manuela Delfanti
Ingrun Griesa
Maximilian Budwill
Julia Lee
Lars Walter
Anja Zeltner

FB5 Students

Dominic Budisantoso
Felix Lohmann
Sascha Mähler
Leon Raasch
Laura Schubert
Tim Krause
Antonia Schubert
Bruno Seifert
Santiago Pietrek
Johanna Küsters
Elias Teut
Tim Klages
Juliane Höving
Dmytro Shulpiak
Johannes Hercher

Keynote Speeches

Conventional and systemic Risks: Implications for individual and societal governance

PROF. DR. DR. ORTWIN RENN
Scientific Director at the Institute for Advanced Sustainability Studies (IASS) in Potsdam, Germany

Ortwin Renn is scientific director at the International Institute for Advanced Sustainability Studies (IASS) in Potsdam (Germany) and professor for environmental sociology and technology assessment at the University of Stuttgart. He also directs the non-profit company DIALOGIK, a research institute for the investigation of communication and participation processes. Renn is Adjunct Professor for "Integrated Risk Analysis" at Stavanger University (Norway), Honorary Professor at the Technical University Munich and Affiliate Professor for "Risk Governance" at Beijing Normal University. His research interests include risk governance (analysis perception, communication), stakeholder and public involvement in environmental decision making, transformation processes in economics, politics and society.

Ortwin Renn has a doctoral degree in social psychology from the University of Cologne. Renn is a member of the German National Academy of Sciences "Leopoldina", the Berlin-Brandenburg Academy of Sciences (Berlin), and of the Board of Directors of the German National Academy of Technology and Engineering (Acatech). His most prominent English publication is the book "Risk Governance. Coping with Uncertainty in a Complex World (London: Earthscan 2008).

Understanding risk under conditions of inherent Uncertainty

DAVE SNOWDEN
Director & Founder – The Cynefin Centre Cognitive Edge, UK

In this presentation he will draw on the EU field guide to example questions how we measure and manage risk under conditions of inherent uncertainty where our predictive capacity is limited. He will explore key ideas from the field guide, including the use of human sensor networks for weak signal detection and the generation of anticipatory triggers to draw human attention to outliers that they would normally ignore.

Snowden is the founder of one of the five schools of sense-making, and takes an approach which draws heavily on natural science, in particular complex adaptive systems theory, cognitive neuroscience and the biological end of anthropology. He is the creator of the Cynefin framework and lead author of the recently published EU Field Guide to managing complexity (and chaos).

He is the author of Complex Acts of Knowing which is in the top ten of cited papers in knowledge management and is a visiting Chair at the Centre of Systems Studies at the University of Hull.

Don't forget the knowledge risks in conjunction with the green transformation

PROF. DR. SUSANNE DURST
Professor of Management, Department of Business Administration, Tallinn University of Technology, Estonia

Research on knowledge risks, their consequences and potential ways of handling them is underdeveloped in general. Consequently, an understanding of its link with green transformation is missing too. Taking into account the emerging need for a green transformation, the keynote introduces knowledge risks from the perspective of green transformation and provides some possible solutions of handling those risks in organizations.

Susanne Durst is a Full Professor of Entrepreneurship at Tallinn University of Technology, Estonia, and Full Professor of Business Administration at the University of Skovde, Sweden. She holds a doctorate in Economics from Paris-Sud University, France. Her research interests include small business management, knowledge (risk) management, responsible digitalization and sustainable business development. She has been conducting several national and international research projects on knowledge (risk) management, SME business transfers, innovation management and learning. Her work has been recognised through different awards, including an Emerald Literati Award in 2020, and has been published in international peer-reviewed journals. Before joining academia, she worked with private organizations of different sizes and industries.

Track – Knowledge Management I

Level of Adoption and Use of KM Practices in Supply Chains of Manufacturing Companies

TOMAS CHERKOS KASSANEH *

Department of Management and Engineering, University of Padova, Stradella S. Nicola, 3, 36100, Vicenza, Italy & Bahir Dar Institute of Technology, Bahir Dar University, Bahir Dar, Poly-peda 01 0026, Ethiopia
E-mail: tomascherkos.kassaneh@studenti.unipd.it

ETTORE BOLISANI † AND ENRICO SCARSO

Department of Management and Engineering, University of Padova Stradella S. Nicola, 3, 36100, Vicenza, Italy
†E-mail: ettore.bolisani@unipd.it & enrico.scarso@unipd.it

JUAN GABRIEL CEGARRA NAVARRO

Faculty of Business Studies, Technical University of Cartagena, Calle real, 3, 30201, Cartagena, Spain
E-mail: Juan.Cegarra@upct.es

Business competition increasingly involves entire supply chains (SCs) rather than individual firms. Thus, knowledge, which is a strategic resource for companies, needs to be managed properly not only in single firms, but also at inter-firm level and in SCs, so that companies can coordinate and co-manage resources, activities, and innovative efforts in collaboration with their partners for improved competitiveness. To this end, knowledge management (KM) practices and their adoption for the effective management of SC inter-firm relationships can be of great help, but the current studies on this aspect are very fragmented, and there is still insufficient knowledge of what practices are adopted to manage knowledge in inter-firm relationships and how intensely these are used. This study contributes to filling this gap by measuring the adoption and intensity of KM practices and comparing their use at the intra- and inter-firm level. A survey of a sample of European large and medium-sized manufacturing firms was used. A descriptive statistical analysis and a correlation analysis were performed on the collected data. The study confirms that KM practices are less used at the inter-firm level. Also, a relative higher variation is depicted among firms in the use of the proposed practices at inter-firm applications. On the other hand, the correlation analysis revealed that the more intensely firms used the practices internally, the more likely the firms will intensely employ them to manage knowledge in the SC inter-firm relationships. The study provides information about how the different KM practices are used by firms. Improving the awareness of that for business managers and consultants, both at firm and inter-firm, can facilitate introduction and development of KM programs.

1. Introduction

In recent decades, particularly in the current knowledge-based and digitalization era, there is increasing awareness that knowledge is a strategic resource of companies which must be managed properly not only in single firms but also across the supply chains (SCs) to face the complex business challenges. This is due to two main reasons. First, it has become evident that competition tends to involve entire SCs rather than individual firms (Attia, 2015; Li et al., 2006; Shakerian et al., 2016); therefore, managers should focus not only on the planning and operations of internal activities, but also on how the different capabilities, resources, and processes of all the firms in an SC can be profitably integrated and coordinated. Another reason is that knowledge in SC connections can require special measures to be managed (Rodríguez-Enríquez et al., 2015; Thomas et al., 2017). Thus, since competition affects every company in an SC, it is important to address solutions from a collective rather than an individual point of view. Therefore, managing knowledge in SCs is even more important than in a company itself, to face the competition challenges and achieve sustainable growth that adds value to the entire SC. The adoption of appropriate knowledge management (KM) practices is also recognized to be a way to improve innovation performance (Inkinen et al., 2015).

In recent years, there has been a progressive shift of focus from traditional intraorganizational KM to interorganizational KM, i.e., the application of KM approaches to manage the relationships with external partners (i.e., suppliers, customers, service providers, etc.) (Agostini et al., 2020; Chen, 2021; Tesavrita et al., 2017; Van Wijk et al., 2008). In essence, knowledge generated in any part of a SC and flowing through inter-firm connections must be managed properly for achieving higher business value (Rodríguez- Enríquez et al., 2015; Thomas et al., 2017), undertaking joint knowledge creation and problem solving, sharing knowledge among the appropriate SC members with the adequate level of protection, and so on.

The development of "KM-based SCs" is an opportunity to achieve better value for customers (Patil & Kant, 2013; Wadhwa & Saxena, 2005), to promote better use of resources in knowledge-intensive and multi-cultural enterprises (Samuel et al., 2011), and to improve the sustainability of business activities (Martins et al., 2019). The recent cases and trends in the global economy (for example, the impact of the COVID-19 pandemic on supply and distribution chains, and the increasing demand for sustainable products and productions - Kumar et al., 2020), showed that KM practices supported with technologies (Centobelli et al., 2019; H. Inkinen et al., 2017) can help to reduce the knowledge gaps that are essential to manage purchases, supplies, and sales, and ensure a traceable and transparent environment.

While KM and SC are two substantial research fields, the application and impact of KM practices to inter-firm relationships and SCs (Cerchione & Esposito, 2016; Chen, 2021; Kassaneh et al., 2021) has not been sufficiently addressed. Of course, during the past 20 years (Kassaneh et al., 2021), key contributions have been published and these studies (Butt, 2021; Chen, 2021; Chen et al., 2018; Lim et al., 2017; Marra et al., 2012) support that there is a growing interest in applying KM to SCs. The literature is, however, fragmented, and the research has sometimes taken diverging directions. Importantly, the level of adoption and use of KM practices by firms at both inter- and intra-level, has not been well addressed so far, and the existing studies (Centobelli et al., 2019; Cerchione & Esposito, 2017) are limited in scope or extension.

This study aims to contribute to fill this research gap by examining how firms exercise KM practices internally and in inter-firm relationships (in SCs). It specifically focuses on measuring the adoption and intensity of use of KM practices, and a comparison between intra- and inter-firm applications. This is also vital for current and future managers of SCs, that can learn more about where and how KM practices, not only in their organizations but also in their relationships with external partners. Providing better awareness and effective understanding of the possible KM practices that are adopted in SCs, to facilitate the creation, delivery, sharing, and protection of knowledge across SC partners (manufacturers, suppliers, customers, service providers, etc.) for common strategic goals, is crucial.

2. Research framework and the context

The starting point of this study is a recently proposed triple-categorization of KM practices (Kassaneh et al., 2021) which was initially developed from an in-depth systematic literature review and a subsequent review based on feedbacks and discussions with KM experts. That study defined KM practices as ‘‘a group of activities regarding the use of practical methods to manage knowledge as a resource and/or applications of IT systems for KM and/or use of other supporting management actions that can also support the fruitful adoption and development of KM for better performance of firms’’. KM practices in SCs explicitly focus on “KM for collaboration between different companies and trading partners, involving the various stakeholders in an SC”. Based on this conceptualization, the study also systematically classified the practices into three primary categories: KM methods, KM applications of IT, and KM-enabling management actions (figure 1). “KM methods” refer to practices that are explicitly and directly targeted to the management of knowledge resources in a company, such as: approaches to learning and exchanging knowledge contents; practical or mental toolboxes for favouring the systematization and access to knowledge resources; organizational arrangements which can be employed to facilitate knowledge sharing among people, etc. The second category, “KM applications of IT” refers to the use of IT systems for supporting the management of knowledge contents in some form – for example, explicit knowledge in the case of database repositories and automatic analysis, or tacit knowledge in the case of communication- enabling systems. The third category, “KM-enabling management actions”, includes purposeful KM-related managerial activities that may not directly refer to the management of knowledge but, anyway, can help to set the appropriate organizational context that facilitates the application of KM - for example, organizing a KM office, providing KM training, etc.

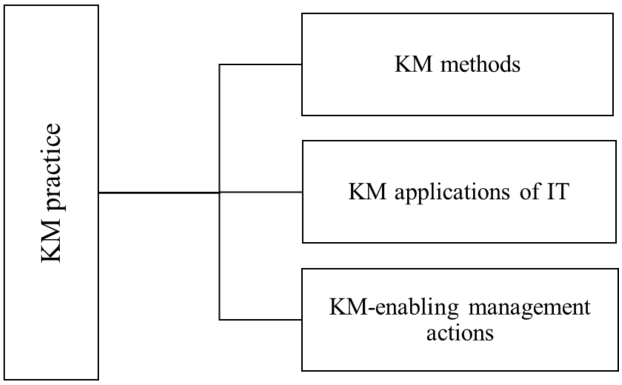


Fig. 1. Triple-category systematic classification of KM practices

Based on this conceptualization, numerous practices are identified and categorized as shown in the taxonomic scheme of table 1 (Kassaneh et al., 2021). Unlike other models, this classification provides a complete picture of KM practices that encompasses many of the core KM aspects (Edwards, 2015) and shows an all-inclusive view of KM (Dalkir, 2011). More importantly, it includes specific list of practices under each category that can also be used to the application to the case of SCs for collaborative learning among the trading partners.

Table 1. Taxonomic scheme for the triple-part categorized KM-practices

Category for KM-Practices	Conventional names of practices	
KM Methods	Community of practice (CoP) Knowledge domain mapping Lessons learnt Knowledge Cafe´s Peer Assist	Enterprise Social Network (ESN) Analysis Case based reasoning Online knowledge searches Brainstorming Mentoring and coaching
KM applications of IT	Email and voice mail Chat rooms and bulletin board systems (BBS) Wikis Database systems and shared folders Video conferencing Enterprise Resource Planning (ERP)	Enterprise social media platforms Simple knowledge organization system (SKOS) Cloud computing Customer Relationship Management (CRM) systems Artificial Intelligence (AI) systems Supplier Relationship Management (SRM) systems
KM-enabling management actions	Project teams Knowledge strategy plans Knowledge development Collaborative KM Top management support Open sharing structure KM assessment Knowledge networking	Knowledge protection Knowledge communication KM investments KM training KM recognition Rewarded knowledge sharing Knowledge retention Trust building

As mentioned, the aim of this study is to analyze adoption and intensity of use of these practices among firms, based on empirical evidence. Therefore, the following research questions (RQs) were formulated:

- RQ1. What is the firms' degree of adoption of KM practices both at intra- and interfirm level?
- RQ2. To what extent (i.e., intensity) firms exercise these KM practices (both at intra- and inter-firm application)?
- RQ3. What homogeneity or difference exists among the firms, across the various practices, and comparing the intra- and inter-firm application?

From a scientific perspective, the achievement of this objective could fill the literature gaps highlighted above and could provide an up-to-date overview of the KM practices used in firms in general and their applications to the case of interfirm relationships in SCs. From a practical viewpoint, a study on KM practices, that facilitates companies' KM effort to properly interact with the SC partners for a collaborative learning, will provide lessons to managers on what practices are mostly used in SCs.

3. Method: survey approach and the measures

3.1. Unit of investigation and Survey approach

It is not easy to detect and examine the KM practices in a SC across all the possible involved firms and inter-firm relationships. Therefore, considering that there are some companies which often have a leading role in an SC (for example, a large manufacturer with its suppliers or customers) and they may also have the power to select what KM practices, methods and technologies be used to exchange and manage knowledge in SCs, it was decided to focus, as unit of analysis for this study, on the KM practices used by these leading or focal firms. These practices where investigated in relation to its application internally as well as in relationship with their main suppliers and customers. This was considered convenient for data collection, to make the study feasible and realistically implementable.

A survey with a structured questionnaire was used to collect data. Each item in the questionnaire was designed to measure the intensity of use of practices by leading firms in three different application situations: internally, in relationship with suppliers, and with customers. This can help to examine the practices used by firms in collaboration with their SC partners (at least with their main suppliers and customers). All measures were based on five-point verbal frequency scale (Flynn et al., 1990) (from 0 = don't used at all to 4 = used very frequently). To understand how knowledge is managed differently in the internal organizational context compared to the external inter-firm environment, the survey investigated both aspects. Indeed, the first point from the scale (0 value) does not show a frequency of use, rather it is dedicatedly introduced to know which practices are not adopted by a firm.

The study used a "convenience sampling" which is a popular approach (Pham & Ho, 2017; Zhang & Liu, 2021), especially given the exploratory and descriptive nature of this research and the difficulty to reach respondents (in the pandemic time) using a proper random sample representing all the companies of the population. As mentioned, the survey was carried out on a sample of 60 large and medium sized European manufacturing firms located in Italy, Spain, Germany, Switzerland, Netherlands, and UK. The sample included companies with at least 100 employees (which was considered a minimum size for a really leading firm in an SC) and, in the end, mainly comprised (75%) firms with more than 249 employees. In addition, only manufacturing industries were considered because, given that the focus of this study is KM in SCs, it was deemed more likely to find complete supply chains in these sectors.

Target respondents were KM managers/officers, if any in the company, or alternatively CEOs, general or plant managers, and SC and operations managers. Generally, as shown in the table 2, the majority of respondents are working in a position where they are familiar to the overall (including inter-firm) KM activities of the company, which makes them suitable for the study.

Table 2. Classification of respondents by their roles/job positions in the companies

Roles/job positions	%
CEOs/General managers/Plant managers	18.3%
KM managers/officers	20.0%
SC and operations managers	18.3%
R&D/project/innovation managers/directors	21.7%
IT/IS/ technology managers	6.7%
Marketing/sales managers	6.7%
Human Resource managers	8.3%

3.2. Descriptive statistics and measures

The firm's differentiation index and intensity of use index were used as a measure for the adoption and use of practices respectively. As explained in the previous studies (Centobelli et al., 2017; Cerchione & Esposito, 2017), the index of differentiation (IoD) is defined as the ratio of the number of KM practices adopted by the firm to the total number of possible KM practices (in the proposed classification). It ranges from zero, if no KM practice is adopted by the firm, to one hundred, if the firm adopted all the listed KM practices. The intensity of use index (IoUI) is the mean score value of the frequency of use (calculated based on the responses to the questionnaire) of all the KM practices adopted by the individual firm. The IoUI shows how intensely the firms are using the practices. Here, the two indexes were calculated for each investigated firm in each category of KM practice and at each application level.

For example, the scores given (frequency of use of each practice) for the ten KM-Methods in the relationship with suppliers is calculated for each firm, and the mean score value is taken as "IoUI of the firm with suppliers". Then, the IoUI values referring to the three of levels of application (i.e., IoUI internally, IoUI with suppliers and IoUI with customers) are averaged and considered as "the IoUI of the firm". The results are presented and discussed in next chapter. Also, the standard deviation (SD) and the coefficient of variation (CV), which is determined by using mean (an average across all firms) and SD, are determined at each application level. Furthermore, using the firms' IoUI values, a correlation analysis is conducted to see the relationship between the firms' intensity of use of practices at intra- and inter-firm level.

4. Survey results and discussion

4.1. Index of differentiation

Through the firms' adoption levels of practices at each application level (i.e., internally, with suppliers or with customers) it is possible to make comparisons between the internal and the inter-firm adoption levels. The index of differentiation of the category "KM methods" (table 3) ranges from 30% (a firm-F20 that adopted only 3 practices out of 10) to 100% (firms that adopted all the practices, of which 28% of investigated firms at inter-firm level and half of surveyed firms at internal use) with a mean of 78.5 at inter-firm and 88.3 at intra-firm level.

Similarly, IoD for the second category of practices ranges from 30.6% (firms, F27 and F35, that adopt only 4 out of 12 KM applications of IT practices) to 100% (firms that adopted all the KM-IT practices, of which 13% of firms at inter-firm level and 21% of firms at internal use) with a mean of 67.5 at inter-firm and 74 at intra-firm level. Also, for the third category, IoD ranges from 27.1% (firms, F20 and F33, that adopted only 5 out of 16 KM-enabling management actions) to 100% (firms that adopted all management actions, of which 15% of firms at inter-firm level and 30% of firms at internal use) with a mean of 70.8 at inter-firm and 79.4 at intra-firm level.

Generally, the firms' level of adoption of practices ranges from the minimum values of 27.1 to 100 percent (on average, about 33% of firms at internal use and 19% of firms at inter-firm application) but it is not a real heterogeneity (high variation) among firms, but rather it is due to the presence of some firms which adopted few practices, especially at interfirm level. In fact, the mean values are always above 67, implying that majority firms adopted most of the proposed practices, and a relatively lower CV values (26.1, 32.5, 33.7 for methods, IT practices and supporting management actions respectively). Comparing the intra- vs inter-firm IoD, the respective mean score values does not show a significant difference.

This can signal that firms adopting practices for internal KM are also adopting the practices to manage their knowledge in interfirm relationships. In addition, the adoption level of KM practices is generally, and irrespective of the intensity of use (see below), not marginal, so it represents the basis for future successful introduction and development of KM programs.

Table 3. Firms' IoD (adoption level) of KM practices (%)

Firm	KM Methods		KM-applications of IT		KM-enab. Management actions		Firm	KM Methods		KM-applications of IT		KM-enab. Management actions	
	Internal	Inter-firm	Internal	Inter-firm	Internal	Inter-firm		Internal	Inter-firm	internal	Inter-firm	internal	Inter-firm
F1	100	86.7	100	88.9	93.7	75.0	F31	100	96.7	75.0	83.3	100	100
F2	100	63.3	83.3	86.1	100	97.9	F32	100	100.0	100.0	100	100	97.9
F3	70	50.0	58.3	41.7	68.7	39.6	F33	60	40.0	41.7	33.3	31.2	27.1
F4	90	76.7	58.3	47.2	68.7	60.4	F34	90	90.0	50.0	50.0	75	75.0
F5	100	100	100	80.6	100	100	F35	90	83.3	33.3	30.6	31.2	29.2
F6	90	53.3	66.7	44.4	43.7	35.4	F36	100	90.0	66.7	61.1	68.7	58.3
F7	80	90.0	83.3	77.8	87.5	91.7	F37	80	50.0	83.3	55.6	87.5	56.3
F8	100	86.7	58.3	58.3	75	72.9	F38	90	76.7	41.7	55.6	75	52.1
F9	90	66.7	83.3	69.4	100	81.3	F39	70	36.7	58.3	41.7	68.7	43.8
F10	100	83.3	66.7	55.6	87.5	75.0	F40	90	70.0	66.7	50.0	81.2	64.6
F11	100	93.3	91.7	91.7	100	100	F41	100	100	91.7	97.2	93.7	93.8
F12	100	100	100	100	100	100	F42	100	93.3	83.3	52.8	100	85.4
F13	100	100	100	100	87.5	87.5	F43	60	53.3	75.0	61.1	75	58.3
F14	70	70.0	50.0	55.6	50	45.8	F44	90	83.3	100	100	93.7	79.2
F15	100	76.7	58.3	50.0	31.2	39.6	F45	100	100	100	91.7	100	100
F16	100	100	58.3	58.3	50	50.0	F46	90	63.3	66.7	44.4	100	91.7
F17	90	66.7	75.0	58.3	100	85.4	F47	100	56.7	41.7	33.3	68.7	45.8
F18	100	100	58.3	58.3	93.7	89.6	F48	50	50.0	66.7	50.0	68.7	70.8
F19	60	53.3	91.7	91.7	93.7	93.8	F49	100	100	83.3	83.3	93.7	93.8
F20	50	30.0	41.7	36.1	31.2	27.1	F50	100	100	100.0	100	100	100
F21	50	46.7	41.7	38.9	43.7	31.3	F51	100	100	83.3	83.3	93.7	93.8
F22	90	86.7	75.0	72.2	43.7	39.6	F52	70	63.3	66.7	58.3	56.2	47.9
F23	100	100	66.7	52.8	75	45.8	F53	100	80.0	100.0	80.6	100	60.4
F24	100	100	100	94.4	93.7	93.8	F54	80	80.0	58.3	47.2	87.5	70.8
F25	90	90.0	91.7	88.9	87.5	87.5	F55	100	100	100	100	100	100
F26	80	53.3	41.7	44.4	56.2	43.8	F56	60	60.0	75.0	75.0	50	50.0
F27	60	40.0	33.3	30.6	68.7	39.6	F57	90	76.7	66.7	52.8	81.2	56.3
F28	100	100	83.3	83.3	62.5	62.5	F58	100	86.7	83.3	69.4	100	100
F29	100	100	100	100	100	100	F59	100	93.3	83.3	72.2	100	95.8
F30	100	100	100	100	100	91.7	F60	80	73.3	83.3	83.3	87.5	68.8

4.2. *Intensity of use*

4.2.1 KM Methods

The IoUI (table 4) ranges from 7.5 to 91.7 with a mean of 45.2 at inter-firm level and 58.4 at intra-firm level. Both at firm and interfirm level, the IoUI is in general lower (particularly with suppliers - 37.5 and customers - 39.8), compared to the IoD (above 78.5 as mentioned previously). Also, figure 2 shows that the firms' intensity of use of KM methods with suppliers and customers is far lower than the internal use. Most firms exercise the KM practices primarily for internal use.

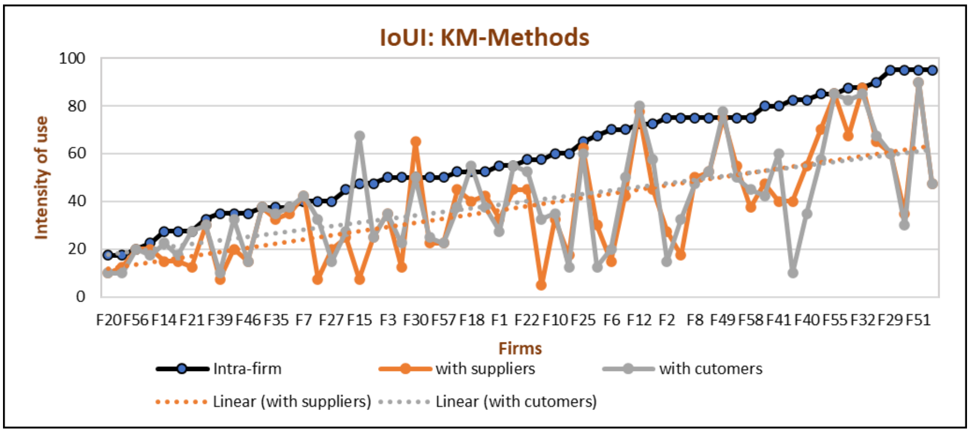


Fig. 2. Intensity of use of KM-Methods - % for each company

For some firms the result is opposite, i.e., the IoUI with suppliers and customers is above the internal intensity of use. This may depend on the fact that some practices may be more applicable with suppliers or customers than the internal use. In any case, the overall IoUI is higher for internal KM.

Table 4. Summary of statistical values of IoUI of KM-Methods (%)

	Internal	With suppliers	With customers
Min	17.5	7.5	10
Max	95	90	90
Mean	58.4	37.5	39.8
CV	37.8	58.2	53.9

Looking at the CV values (table 4), a higher variation is for KM methods with suppliers (58.2) which implies a relative heterogeneity of firms. At the same time, there is a positive relationship among the firms' use of KM methods internally and with suppliers and customers.

4.2.2 KM application of IT

IoUI ranges from 12.5 to 95.8 internally and from 11.1 to 91.7 at inter-firm level, with a mean of 50.8 at intra-firm and 41.1 at inter-firm application (table 5). Similarly, to the results for KM-Methods, there is a lower average IoUI for IT practices with suppliers (35.1) and customers (37.3) compared to internal use (50.8) (figure 3).

Table 5. Summary of statistical values of IoUI of KM applications of IT (%)

	Internal	With suppliers	With customers
Min	12.5	8.3	10.4
Max	95.8	83.3	95.8
Mean	50.8	35.1	37.3
CV	43.3	60.3	57.2

A lower CV of KM-IT practices internally indicates a relative homogeneity among firms. Like the KM methods, a higher variation in the intensity of use of KM-IT practices is shown with suppliers (60.3). Figure 3 also shows that a higher intensity of use of KM-IT practices with suppliers and customers is registered in six firms compared to their internal level. This may not be expected and needs further investigation. A possible explanation is that these companies adopted KM- IT under the influence of their suppliers and customers.

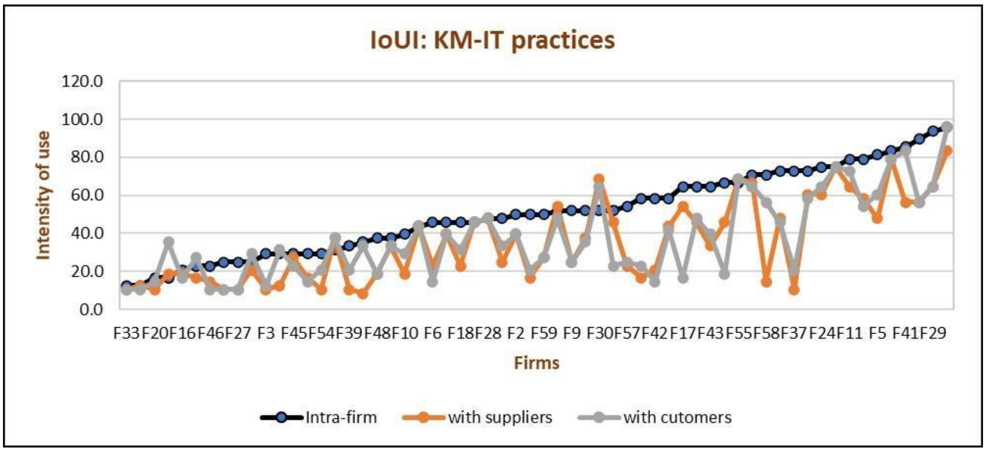


Fig. 3. Intensity of use of KM applications of IT - % for each company

4.2.3 KM-enabling management actions

The firms have almost a similar average intensity of use of KM-enabling management actions compared to the use of KM-IT practices, both at firm and inter-firm level (table 5 and 6). IoUI of KM-enabling management actions ranges from 7.8 to 88 at inter-firm and from 7.8 to 95.3 at intra-firm level with a mean of 40.5 and 50.8 respectively. Also, as depicted in figure 4, even though the IoUI is in general higher at intra-firm than inter-firm level, some firms have a higher IoUI with suppliers and/or customers than internally.

Table 6. Summary of statistical values of IoUI of KM-enabling management actions (%)

	Internal	With suppliers	With customers
Min	7.8	6.3	6.3
Max	95.3	87.5	87.5
Mean	50.8	35.0	35.8
CV	45.8	64.4	64.4

The CV (table 6) shows some heterogeneity (higher CV values) of companies, especially in the inter-firm case (CV is 64.4 with suppliers and customers). This can be explained by the presence of a group of firms with very low IoUI (from 6.3 to 25) and another group with a very high IoUI (more than 60). In comparison to the other practices, the IoUI (table 4, 5 and 6) is lowest in this category.

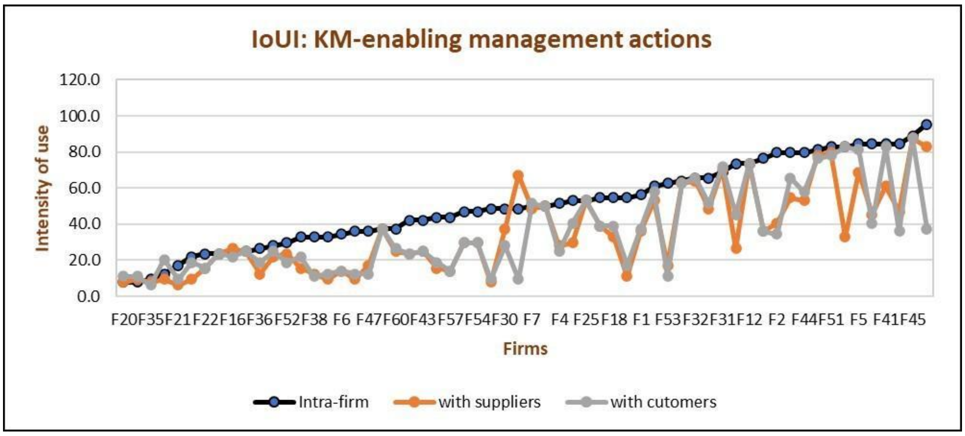


Fig. 4. Intensity of use of KM-enabling management actions - % for each company

To sum up, even though there are some firms which have a higher intensity of use of practices at each category, in general, the firms' intensity of use of practices at inter-firm level is low (on average below 45) compared to their level of adoption (on average above 67).

4.2.4 Correlation analysis

To examine the relationship between the IoUI of KM practices internally and at inter-firm level, a correlation analysis is conducted using the IoUI values of firms for each category of KM practice (figures 5, 6 and 7). A notable result is that a higher and positive correlation is found among the IoUI of firms in the three of the relationships, i.e., the relationship between the use of practices at intra- and inter-firm level (r=0.85 for methods, r=0.91 for IT practices and r=0.92 for management actions).

In general (and especially in the case of KM-enabling management actions), the more the firms intensively used KM practices at intra-firm level, the more the firms will intensively use them to manage knowledge in the relationship with suppliers and clients in SCs.

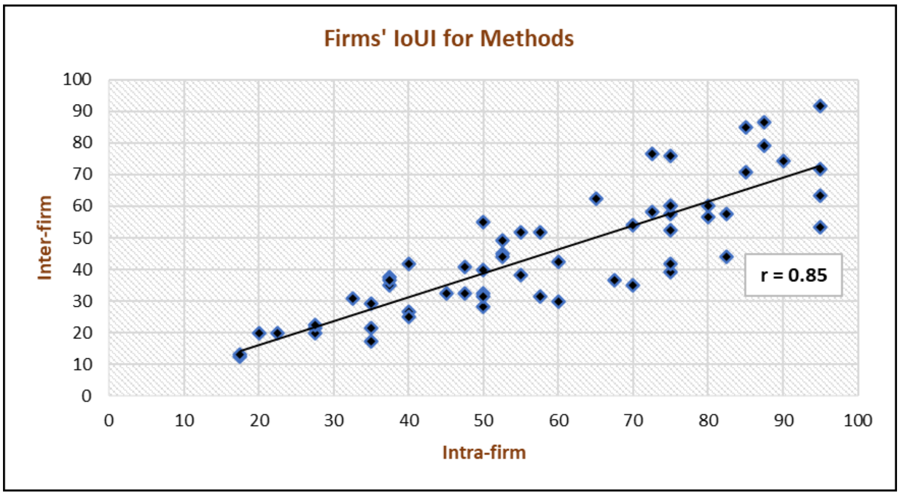


Fig. 5. Correlation between IoUI of KM-Methods at firm and inter-firm level (%)

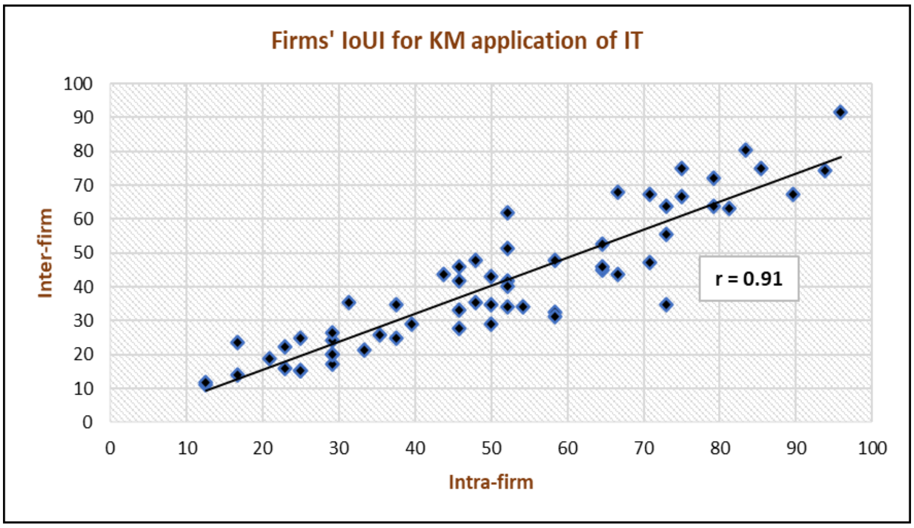


Fig. 6. Correlation between IoUI of KM applications of IT at firm and inter-firm level (%)

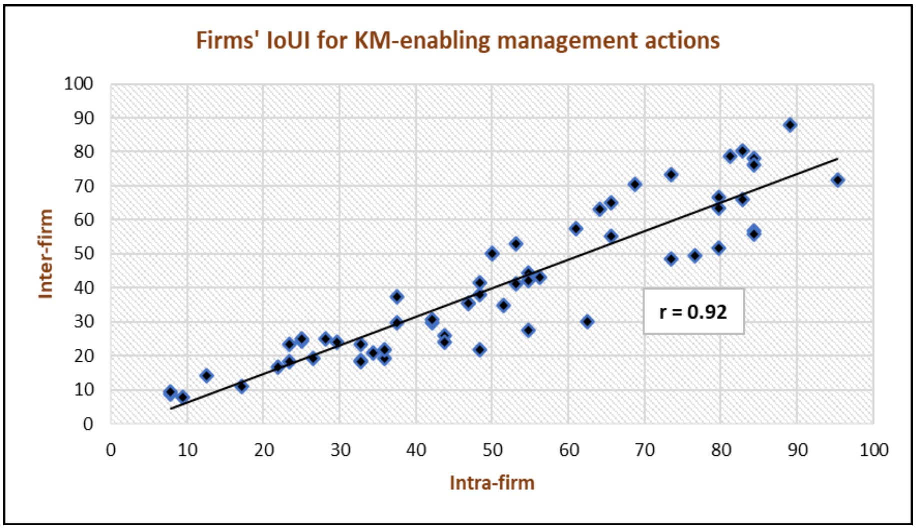


Fig. 7. Correlation between IoUI of KM-enabling management actions at firm and inter-firm level (%)

5. Conclusions

This study aimed to investigate KM practices used by firms not only to manage their knowledge internally but also in the case of inter-firm relationships in SCs. It was found that firms do adopt KM practices, but with a very varied level of adoption and intensity of use. Also, by comparing the use of KM practices internally and at inter-firm levels (i.e., in relationship with suppliers and customers), it was discovered that the latter are characterized by a lower intensity of use. This signals that, while it is common opinion that managing knowledge with SC partners may be important to get full benefits from a collective inter-firm learning process and collaboration, there is a certain “sluggishness” or difficulty to extend internal KM practices to external relationships. In any case, the correlation analysis confirms that the more the firms used the practices internally, the more likely the firms will intensely practice them with partners. It can be concluded that the intensive use of KM practices internally tends to help or lead companies to extend the KM adoption to inter-firm relationships.

This is the general situation, although there are some differences depending on the specific practice (namely, there are cases where a KM practice is used more with partners than internally). There are also differences from a company to another: a partially surprising result is that a few firms show a higher intensity of use of practices (in all the three categories) with suppliers and customers than internally. The reason for that needs further investigation, maybe with other methods.

5.1. Limitations and research opportunities

While the results presented here provide some interesting preliminary insights into the examined issue, there is a need for further empirical validation for assessing their general applicability and relevance. An analysis of a more statistically representative sample of firms can be proposed for a significant generalization of the results. It can be estimated that the total population of companies in the selected countries amounts of approximately 15000 firms, and our sample was of 60. At the current stage, it was not possible to measure the statistical significance properly and, in any case, this was a convenience sample. All these measurements can be done in a future study. It is also essential to go in deeper details about the specific application of KM practices to SCs, to understand the reasons of specific differences or to explain some of the counterintuitive results obtained in this investigation. For this, other methods, for example qualitative case-studies, can be useful here. In addition, the study was restricted to the analysis of KM practices used by firms mainly in the relationship with the main suppliers and customers of a firm, but it did not consider dyads or supply networks, which may require another study. Finally, since the paper shows that the general firms’ intensity of use of practices specifically at the inter-firm level was found low, investigation of the potential barriers hindering the successful implementation and intense use of the KM practices by firms in collaboration with their SC partners can be important.

Acknowledgements

This study contributes to the project “BOLI_BIRD2121_01” funded by the University of Padua.

References

Agostini, L., Nosella, A., Sarala, R., Spender, J. C., & Wegner, D. (2020). Tracing the evolution of the literature on knowledge management in inter-organizational contexts: a bibliometric analysis. *Journal of Knowledge Management*, *24*(2), 463–490. https://doi.org/10.1108/JKM-07-2019-0382

Attia, A. (2015). Testing the effect of marketing strategy alignment and triple-A supply chain on performance in Egypt. *EuroMed Journal of Business*, *10*(2), 163–180. https://doi.org/10.1108/EMJB-07-2014-0020

Butt, A. S. (2021). Termination of personal relationship and their effects on knowledge sharing in supply chains. *Knowledge and Process Management*, *28*(1), 18–28. https://doi.org/10.1002/kpm.1643

Centobelli, P., Cerchione, R., & Esposito, E. (2017). Knowledge management systems: The hallmark of SMEs. *Knowledge Management Research and Practice*, *15*(2), 294–304. https://doi.org/10.1057/s41275-017-0054-x

Centobelli, P., Cerchione, R., & Esposito, E. (2019). Measuring the use of knowledge management systems in supply firms. *Measuring Business Excellence*, *23*(4), 426–441. https://doi.org/10.1108/MBE-12-2018-0107

Cerchione, R., & Esposito, E. (2016). A systematic review of supply chain knowledge management research: State of the art and research opportunities. *International Journal of Production Economics*, *182*, 276–292. https://doi.org/10.1016/j.ijpe.2016.09.006

Cerchione, R., & Esposito, E. (2017). Using knowledge management systems: A taxonomy of SME strategies. *International Journal of Information Management*, *37*(1), 1551–1562. https://doi.org/10.1016/j.ijinfomgt.2016.10.007

Chen, L. (2021). Measuring knowledge management in supplier development: A knowledge chain perspective. *Knowledge and Process Management*, *28*(4), 377–387. https://doi.org/10.1002/kpm.1690

Chen, L., Ellis, S., & Holsapple, C. (2018). A knowledge management perspective of supplier development: Evidence from supply chain scholars and consultants. *Knowledge and Process Management*, *25*(4), 247–257. https://doi.org/10.1002/kpm.1566

Dalkir, K. (2011). Knowledge management in theory and practice. In *Knowledge Management in Theory and Practice* (Second). MIT Press. https://doi.org/10.4324/9780080547367

Edwards, J. S. (2015). Knowledge Management Concepts and Models. In E. Bolisani & M. Handzic (Eds.), *Advances in Knowledge Management: Celebrating Twenty Years of Research and Practice* (pp. 25–44). Springer International Publishing. https://doi.org/10.1007/978-3-319-09501-1_2

Flynn, B. B., Kakibara, S. S., Schroeder, R. G., Bates, K. A., & Flynn, E. J. (1990). *Empirical Research Methods in Operations Management*. *9*(2).

Inkinen, H., Kianto, A., Vanhala, M., & Ritala, P. (2017). Assessing the universality of knowledge management practices. *Journal of Knowledge Management*, *21*(6). https://doi.org/10.1108/JKM-09-2016-0394

Inkinen, H. T., Kianto, A., & Vanhala, M. (2015). Knowledge management practices and innovation performance in Finland. *Baltic Journal of Management*, *10*(4), 432–455. https://doi.org/10.1108/BJM-10-2014-0178

Kassaneh, T. C., Bolisani, E., & Cegarra-Navarro, J. G. (2021). Knowledge management practices for sustainable supply chain management: A challenge for business education. *Sustainability (Switzerland)*, *13*(5), 1–15. https://doi.org/10.3390/su13052956

Kumar, A., Luthra, S., Mangla, S. K., & Kazançoğlu, Y. (2020). COVID-19 impact on sustainable production and operations management. *Sustainable Operations and Computers*, *1*(July), 1–7. https://doi.org/10.1016/j.susoc.2020.06.001

Li, S., Ragu-Nathan, B., Ragu-Nathan, T. S., & Subba Rao, S. (2006). The impact of supply chain management practices on competitive advantage and organizational performance. *Omega*, *34*(2), 107–124. https://doi.org/10.1016/j.omega.2004.08.002

Lim, M. K., Tseng, M. L., Tan, K. H., & Bui, T. D. (2017). Knowledge management in sustainable supply chain management: Improving performance through an interpretive structural modelling approach. *Journal of Cleaner Production*, *162*, 806–816. https://doi.org/10.1016/j.jclepro.2017.06.056

Marra, M., Ho, W., & Edwards, J. S. (2012). Supply chain knowledge management: A literature review. *Expert Systems with Applications*, *39*(5), 6103–6110. https://doi.org/10.1016/j.eswa.2011.11.035

Martins, V. W. B., Rampasso, I. S., Anholon, R., Quelhas, O. L. G., & Leal Filho, W. (2019). Knowledge management in the context of sustainability: Literature review and opportunities for future research. *Journal of Cleaner Production*, *229*, 489–500. https://doi.org/10.1016/j.jclepro.2019.04.354

Patil, S. K., & Kant, R. (2013). A fuzzy DEMATEL method to identify critical success factors of knowledge management adoption in supply chain. *Journal of Information and Knowledge Management*, *12*(3). https://doi.org/10.1142/S0219649213500196

Pham, Q. T., & Ho, B. T. (2017). Impact factors of knowledge sharing intention of it employees in Vietnam: An integrated approach. *A Knowledge Management Approach for Ensuring the Success of IT Industries in Vietnam*, *April*, 71–92. https://doi.org/10.46223/hcmcoujs.econ.en.7.1.193.2017

Rodríguez-Enríquez, C. A., Alor-Hernández, G., Sánchez-Ramírez, C., & Córtes-Robles, G. (2015). Supply chain knowledge management: A linked data-based approach using SKOS. *Dyna*, *82*(194), 27–35. https://doi.org/10.15446/dyna.v82n194.54463

Samuel, K. E., Goury, M. L., Gunasekaran, A., & Spalanzani, A. (2011). Knowledge management in supply chain: An empirical study from France. *Journal of Strategic Information Systems*, *20*(3), 283–306. https://doi.org/10.1016/j.jsis.2010.11.001

Shakerian, H., Dehnavi, H. D., & Shateri, F. (2016). A Framework for the Implementation of Knowledge Management in Supply Chain Management. *Procedia - Social and Behavioral Sciences*, *230*(May), 176–183. https://doi.org/10.1016/j.sbspro.2016.09.022

Tesavrita, C., Suryadi, K., Wiratmadja, I. I., & Govindaraju, R. (2017). Intra-organizational and inter-organizational knowledge sharing in collaborative learning process: A conceptual framework for SME. *2017 4th International Conference on Industrial Engineering and Applications, ICIEA 2017*, 187–191. https://doi.org/10.1109/IEA.2017.7939204

Thomas, A., Dorrington, P., Costa, F., Loudon, G., Francis, M., & Fisher, R. (2017). Organisational learning capability in SMEs: An empirical development of innovation in the supply chain. *Cogent Business and Management*, *4*(1), 1–20. https://doi.org/10.1080/23311975.2017.1364057

Van Wijk, R., Jansen, J. J. P., & Lyles, M. A. (2008). Inter- and intra-organizational knowledge transfer: A meta-analytic review and assessment of its antecedents and consequences. *Journal of Management Studies*, *45*(4), 830–853. https://doi.org/10.1111/j.1467-6486.2008.00771.x

Wadhwa, S., & Saxena, A. (2005). Knowledge management based supply chain: an evolution perspective. *Global Journal of E-Business and Knowledge Management*, *2*(2), 13–29.

Zhang, C., & Liu, L. (2021). The effect of job crafting to job performance. *Knowledge Management Research and Practice*, *19*(2), 253–262. https://doi.org/10.1080/14778238.2020.1762517

First Observations from an Analysis of KM Job Adverts from English-Speaking Countries and Germany

Peter Heisig*
Günther Neher †
Department of Information Sciences,
University of Applied Sciences Potsdam,
Kiepenheuerallee 5, 14469 Potsdam, Germany
E-mail: peter.heisig@fh-potsdam.de
†E-mail: guenther.neher@fh-potsdam.de

Extended Abstract

The systematic management of knowledge has become an essential element in a multitude of organizations in any sector and of any size (Drucker, 1993; Al-Hawamdeh, 2003). The responsibility for such tasks has been assigned to roles labelled "Knowledge Manager" or "Chief Knowledge Officer" (Davenport and Prusak, 1998; McKeen and Staples, 2003; Riempp, 2004; Dalkir, 2017).

From the literature about KM and job advertisements we could only identify two previous studies. Majid and Waa Bee (2003) focused on Singapore including 150 job ads from newspapers and LIS websites collected in 2001-2002 and Harper (2013) who analyzed 165 job ads from UK-based websites gathered over six months in 2011. Based on the limited geographical reach, the limited sample size and the

This paper presents the first results from an analysis of recent job advertisements from online platforms from English and German-speaking countries. The full texts of about 4100 job adverts carrying the terms "knowledge manager" or "knowledge management" either in in the job title or in the short description have been analyzed.

Our main research aim was to identify the most common KM job titles, providing a basis for further analysis of tasks and required competencies to compare the data samples. In particular, the research will address the following research questions (RQ):

- RQ1: What are the **job titles** used in job adverts related to KM?
- RQ2: Which **tasks** are mainly associated with different KM job titles (roles)?
- RQ3: Which **competencies** & **skill** sets are typically required in job adverts?

This research aims to inform KM practitioners and HR managers as well as the KM community about landscape of KM roles, main tasks and skills required on the labour markets in the German-speaking (AU, CH, DE) countries and three English-speaking (AUS, UK, USA) countries.

For the present study, job advertisements were collected from sites of the job portal indeed.com for Germany[1], United Kingdom[2], USA[3] and Australia[4] and additionally stepstone.de[5]. These portals were crawled once a week in the time period from 25.12.2021 to 04.06.2022. Search terms used were "knowledge manager", "knowledge management" for all portals and additionally "Wissensmanagement" and "Wissensmanager" for the german portals. For the data acquisition as well as the later data analysis the software tool RapidMiner[6] was used.

A surprising result emerges from the first comparison of the job adverts from Germany and the English-Speaking countries which leads to questions about the differences in staffing of KM positions. In

[1] https://de.indeed.com
[2] https://uk.indeed.com
[3] https://www.indeed.com
[4] https://au.indeed.com
[5] https://www.stepstone.de
[6] https://rapidminer.com

the area of knowledge management, job titles are very heterogenous in all four labor markets as the analysis of the denominations of the job adverts showed for our sample while the job title “knowledge manager” or “Wissensmanager” is the dominant denomination. The sample suggests that the Chief Knowledge Officer role has disappeared from the organizational chart altogether in all four countries.

Looking at the variants of job titles an interesting result emerges from comparing the job adverts from Germany and the three English-Speaking countries (Fig. 1). In job adverts from Germany the UNSKILLED “working student” is in quite high demand while in the English-speaking countries it is the knowledge management “SPECIALIST”. How can we explain this observation? Does this mirror the importance organizations assign to the KM function in the different countries? We would like to discuss this observation with the delegates from academia and industry at the 17th ICKM in order to further analyze our extended data set.

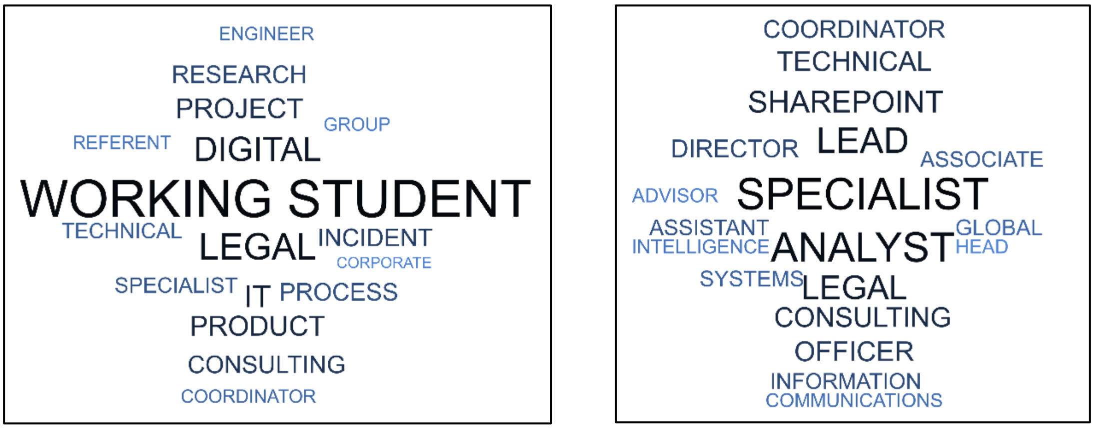


Fig. 1: Terms occurring in Job titles together with KM or WM, visualized as "word cloud".
left: German dataset, right: combined datasets for all english-speaking countries

References

Al-Hawamdeh, S. (2003) Knowledge management: Cultivating knowledge professionals. Rollinsford, N.H.: Chandos Publishing.

Dalkir, K. (2017) Knowledge Management in Theory and Practice. Cambridge M.A.: MIT Press, 3rd. Edition.

Davenport, T.H. and Prusak, L. (1998) Working Knowledge. How Organizations manage what the know. Boston M.A.: Harvard Business School Press

Drucker, P.A. (1993) A Post-Capitalist Society, New York: HarperCollins.

Harper, R. (2013) Knowledge Management through the Lens of Library and Information Science: A Study of Job Advertisements, Library Trends, Vol. 61, No. 3, pp. 703-734

Majid, S., Bee, Waa Wong (2003) Competencies Sought by Information and Knowledge Employers: An Analysis of Job Advertisements, Journal of Information & Knowledge Management, Vol. 2, No. 3, pp. 253-260.

McKeen, J.D. and Staples, D.S. (2003) Knowledge Managers: Who they are and what they do, in: Holsapple, C.W. (Ed.) Handbook on Knowledge Management. Knowledge Matters, Vol.1, Berlin, Heidelberg, New York: Springer, p. 20-41

Riempp, G. (2004) Integrierte Wissensmanagement-Systeme. Architektur und praktische Anwendung. Berlin, Heidelberg, New York: Springer

Soft Skills Development in Knowledge Management

KENDRA ALBRIGHT
Kent State University, USA

ALBINA KRYMSKAYA
Saint Petersburg State University of Culture, Russia

Soft skills have become increasingly important in the workplace, particularly in knowledge management (KM). Developing a knowledge culture in an organization depends on the soft skills of leaders. KM literature, however, does not often discuss the development and teaching of soft skills that can facilitate the success of knowledge management in organizations.

This presentation will present the challenge of teaching and learning soft skills as well as a proposed framework to address those challenges. Currently, there is debate about whether soft skills can be learned or whether they are an innate part of individual personality. Where it was once believed that emotions were interruptions in the cognitive processes of the brain that needed to be controlled, they are now considered to be underlying motivations that direct our activities either consciously or unconsciously. This perspective informs the basis for Emotional Intelligence (EI) theory, which offers a framework of competencies that includes the individual skills necessary to manage emotions and increase personal effectiveness. Active learning is an approach to teaching that engages students in the learning process through instructor-guided activities (e.g., case studies, self-reflection, etc.). EI competencies offer a framework that, when combined with active learning, can address the challenges of teaching and learning soft skills.

This presentation will provide an overview of soft skills that are important in KM. An EI competencies framework will be described for how it can be applied in combination with active learning to facilitate soft skills development. Case studies will be used to demonstrate this approach.

Analysis of knowledge risks identified in the Brazilian Space Program: A methodological framework for verification & mitigation of risks in the light of knowledge governance

JOAO SÉRGIO BESERRA DE LIMA
Universidade de Brasília, Doctoral candidate, joaosergio.lima@outlook.com
https://orcid.org/0000-0003-2346-3844

LILLIAN MARIA ARAUJO DE REZENDE ALVARES
Universidade de Brasília, PhD, lillianalvares@unb.br
https://orcid.org/0000-0002-8920-0150

RODRIGO LEONARDI
Brazilian Space Agency, PhD, rodrigo.leonardi@aeb.gov.br
https://orcid.org/0000-0002-5071-0831

Summary

Brazilian scientific and technological research in the space area covers a diversity of themes so that its results are applied in several areas, such as earth observation, fire control, climate change, telecommunications, and security, among others. By providing useful products and services for society, economy, and industry, one perceives its relevance and the need to improve Brazil's space research. Thus, given its social and economic importance, this project aims to structure, based on scientific and situated knowledge, a framework that could help to mitigate the knowledge risks identified in the diverse types of organizations and sectors that are part of the Brazilian Space Program. This work in progress will try to address this problem by providing a methodological framework based on Knowledge Governance tools and it is intended to homogenize and mitigate the knowledge risks identified in the diverse types of organizations that are part of the Brazilian Space sector. The methodological approach to be used in the research will be predominantly qualitative, with possible quantitative treatment when analyzing the collected data used to develop the framework. We willidentify, analyze and catalogue, based on the distinct categories found in the literature, the types of knowledge risks observed in organizations and sectors that are part of the Brazilian Space Program, to propose a methodological model that is capable of providing mechanisms for mitigating these risks. The originality of this project stems from its attempt to catalog the different knowledge risks identified in the literature and correlate them with the Brazilian Space Program, to contribute to the study of the production cycle and preservation of knowledge generated by Brazilian scientific and technological research in the space area.

Keywords: Knowledge Governance. Knowledge Risks. Brazilian Space Program. Methodological Framework.

1. INTRODUCTION

Technological advances and the resulting increase in access to information have caused major transformations in social and organizational relations in recent decades. According to Alvarez and Caregnato (2018), the amount of scientific and technological information recorded after World War II, as well as the progress of information and communication technologies caused an "explosion of information" that influenced the increase in knowledge production.

Several social, technological, economic, and cultural transformations have come together and given rise to a new form of society, called network society, where knowledge and information have become fundamental elements in all modes of development so that the production process became to be always based on some degree of knowledge and information processing (CASTELLS, 2019).

According to Calazans (2006) "information has evolved through and along with history, modifying meanings and impacting individuals, societies and organizations". It happens that to the large volume of data and scientific information available, there is no corresponding realization of the information analysis stage (ALVARES *et al.*, 2018). Thus, there is a greater discrepancy between the availability and the ability to deal with information.

Although it has occurred more frequently in recent decades, this discrepancy is not a recent phenomenon and can be observed in the historical-disciplinary course of Information Science. Otlet (1989), in his classic work, Document Treaty, already dealt with the theme, including predicting the need for new procedures::

> To make available the daily amount of information, scientific articles, [...] to keep pamphlets, reports, prospectuses, official documents; to find the materials distributed by the books, to make a homogeneous whole of these incoherent masses, it is necessary new procedures, quite different from those of the old librarianship. [...]. (OTLET, 1989, p. 6).

In the conception of Leite and Pinho Neto (2014), the contribution of technology as a facilitator in the process of access and use of information cannot be questioned, however, it is necessary to reflect on the social condition of disseminating data to contribute to the creation of knowledge. In this regard, the authors say:

> The larger question is related to whether the possibilities of access and sharing of information contribute producing advances and scientific advances, which in turn will be able to generate and produce greater social development (LEITE; PINHO NETO, 2014, p. 34–35)

The need to manage the information, as well as the knowledge acquired through it, has given rise to the term Knowledge Management, which means:

> The set of activities aimed at promoting organizational knowledge, enabling organizations and their employees to always use the best information and the best available knowledge, to achieve organizational objectives and maximize competitiveness. (ALVARENGA NETO, 2008, p. 151).

More specifically, regarding Brazilian scientific and technological research in the space area, it is important to point out that it drives the growth of the space industry, which in turn offers more and more products and services so it is necessary to encourage and improve it. According to the AEB (2012, p. 7):

> [...] it is imperative to prioritize the development and mastery of critical space technologies, indispensable to industrial advancement and the achievement of the necessary national autonomy in such a strategic activity. This domain is only achieved with intense and effective synergistic participation of the government, research centers universities and industries

Given the importance that information and knowledge have become in production processes, it is necessary to know the informational risks involved in these processes in different branches of the production chain, as well as to seek ways to identify and eventually mitigate them.

Thus, this project proposes a reflection on the preservation of knowledge generated by Brazilian space research, using practices related to Knowledge Governance, to ensure that it remains accessible over time, through the mitigation of knowledge risks that may be identified.

To this end, it is proposed to structure a methodological framework that assists in mitigating the Knowledge Risks identified in the different types of organizations and sectors that are part of the Brazilian Space Program (PEB).

1.1. RESEARCH PROBLEM

Knowledge exists and is transferred in organizations spontaneously, regardless of formal management processes on it. (DAVENPORT; PRUSAK, 1998).

With the large volume of data and the speed of information generated and available about space research in Brazil, there is the possibility of losing knowledge, either due to its obsolescence or the infeasibility of using existing knowledge, so it is necessary to search for procedures that can optimize the application and perpetuation of its results.

In addition to the obsolescence of the knowledge generated, another difficulty faced by national space research is the issue of the loss of knowledge resulting from the removal of qualified personnel. According to the AEB (2012, p. 14):

> [...] in the main research institutes in the space area, a repressed short-term demand for the hiring of specialists to remake the losses of qualified personnel who have departed or retired in recent years.

The Brazilian technological development is surrounded by many highly qualified professionals, who have been providing the unprecedented realization of design and implementation of a series of national projects. The cycle of knowledge generation necessary for this achievement must be consolidated, otherwise, it may get lost.

In this sense, it is expected that the research can answer the following question: How to identify and mitigate the Knowledge Risks verified in the Brazilian Space Program?

1.2. OBJECTIVES

This research deals with the application of Knowledge Governance by Entities of the Brazilian Space Program responsible for national space research and has the following objectives:

General objective

Structure, from scientific and situated knowledge, a methodological framework that helps in mitigating the knowledge risks identified in the different types of organizations and sectors that are part of the Brazilian Space Program.

Specific objectives

a) To present the Brazilian Space Program, detailing its characteristics, history, development, future perspectives, and its actors, coming from academia, government, and industry.
b) Identify and analyze, based on the different categories found in the literature, the types of knowledge risks observed in organizations and sectors that are part of the Brazilian Space Program.
c) To propose a conceptual framework for identify and mitigate knowledge risks in the Brazilian Space Program.
d) Check the proposed model.

2. DEVELOPMENT

To justify the proposal of this research, it will be explained below its repercussions for Information Science, Knowledge Management, and the research object (Brazilian Space Research).

JUSTIFICATION FOR INFORMATION SCIENCE

For Information Science, the research will contribute to the study of the cycle of production and preservation of knowledge generated by Brazilian scientific and technological research in the space area.

Information Science, characterized by its interdisciplinarity, studies the need for knowledge and information management in organizations. Borko (1968), in particular, mentions that Information Science aims to provide a theoretical body that would lead to improvements to various institutions and procedures dedicated to the accumulation and transmission of knowledge. It happens that, at the time, according to the author, these institutions would be inadequate to address the need for communication, given: "the tremendous growth of Science and Technology and the accelerated pace at which new knowledge becomes obsolete" (BORKO, 1968, p. 4).

Otlet (1937), in his speech at the World Documentation Congress, gave the human element a prominent role in the cycle of operations and documentation products. In this cycle, men "alternately takes ideas from reality or introduces ideas into reality; between reality and the idea intervenes, increasingly, the documents which, in turn, serve the preparation of new documents".

Thus, the need to manage information and knowledge is a theme that can be observed in the classical literature of Information Science, permeating its historical path, its concepts, approaches, and objects of study.

JUSTIFICATION FOR KNOWLEDGE MANAGEMENT

Providing procedures related to the transmission of knowledge in an appropriate way, preventing new knowledge from becoming old-fashioned, and continuing to be used by institutions is one of the objectives of Knowledge Management.

The Brazilian National Institute of Space Research (INPE), the main government institution of national space research, recognizes the importance of Knowledge Management. According to INPE (2016, p. 83):

> In the area of training and capacitation of Human Resources, it is essential to establish the knowledge management of a multidisciplinary community such as that of the Institute, with the maintenance and updating of the policy of preservation and conservation of knowledge. The current context of INPE has as one of its great challenges to establish a culture and an environment dedicated to organizational learning and/or innovation, also seeking to encourage and stimulate the continued education of its employees.

As for increased interest in the topic of Knowledge Management, according to Hislop (2013, p. 1) "the explosion of interest in Knowledge Management among academics, policymakers, consultants and entrepreneurs began [...] in the mid-1990s".

Still, about the increased interest in the subject, Marteleto (2009) made the following considerations:

> In the literature of Information Science, in the studies of its researchers and students, and professional practice, there is a substantive and even overwhelming increase in literature and discourse on "knowledge and information management" with a modeling and application focus on organizations, especially business. (2009, p. 17)

Based on this statement, Marteleto (2009) argues that the new fields linked to Information Science, such as Knowledge Management, are due to a new information regime, which in turn may suggest that these are redundant themes so that a rupture in the approach given to these new themes is necessary. As a solution, the author suggests that research eats how to reflect on the convergence between market society and technological rationality.

The importance of research for Knowledge Management lies in the opportunity to establish guidelines that enable its implementation and optimize the results achieved by research units related to the development of the Brazilian Space Program through the mitigation of possible Knowledge Risks that may be verified by these units.

JUSTIFICATION FOR THE BRAZILIAN SPACE RESEARCH

Brazilian scientific and technological research in the space area covers a diversity of themes so that its results are applied in several areas, such as land observation, fire control, climate change, telecommunications, and security, among others.

It is also an important agent in fostering the economy. According to the AEB (2012, p. 9), "the global space market grows at an annual average of 6%, thanks to the emergence of new demands for applications and services".

By providing useful products and services for society, economy, and industry, one perceives its relevance and the need to improve Brazil's space research. Also according to the AEB, concerning the economic aspect, it is worth noting that space activities move more than US$ 280 billion per year.

Given its social and economic importance, the purpose of pointing out guidelines that can optimize availability, combat obsolescence, and understand how the preservation of the technical-scientific memory of knowledge generated by Brazilian spatial research occurs. It is in this sense that research becomes important for the Brazilian space program.

2.1. PRELIMINARY INDICATIONS OF LITERATURE REVIEW

BRAZILIAN SPACE RESEARCH

The English-language literature on Brazil's space activities is small but growing (MOLTZ, 2015). Thus, when searching for literature on the history of the Brazilian Space Program, it was necessary to use descriptors in both English and Portuguese.

The ability of a nation to generate knowledge and convert it into wealth and social development depends on the action of several institutional agents that generate and apply knowledge, according to Benite (2006). Brazilian space research involves several of these agents, so that the search for improvement of its results, converts into social and economic development. Also according to the author (2006, p. 26):

> The exploration of outer space is an ancient dream of man. Numerous archaeological records expressed throughout the ages by paintings, sculptures, artifacts, and buildings show that the desire to reach space was part of the imagination of several peoples since the emergence of the first civilizations.

Regarding national space research and its history, according to Alvares et al. (2018), space activities in Brazil began after the launches of Sputnik 1 satellites in 1957 by the Soviet Union and Explorer 1 in 1958 by the United States.

In 1960, the Brazilian Interplanetary Society (SIB) decided, during the Inter-American Meeting of Space Research, to propose the creation of a civilian space research institution in Brazil, which culminated, in August of the same year, in the creation of the Group of Organization of the National Commission of Space Activities (GOCNAE), which can be considered as the beginning of space activities in Brazil, according to INPE (2016).

Concerning the need for Brazilian national space industrial and technological research and production, according to the AEB (2005), the importance of training in the field of space technology, which in its complete cycle covers launch centers, launch vehicles, satellites, and payloads, is relevant for the future of the country, as it is necessary to develop its own technologies, without deference from third parties. In this sense, it is stated:

> only countries that dominate space technology can have autonomy in the elaboration of scenarios of global evolution capable of taking into account both the impacts of human action and those of natural phenomena. AEB (2005, p. 7)

Much of the national space research is developed by the National Institute of Space Research, which has its own graduate program (master's and doctorate), seeking knowledge generation and training of high-level human resources, according to INPE (2016).

The research developed by INPE in its postgraduate courses covers several areas such as astrophysics, applied computing, engineering and space technologies, spatial geophysics, meteorology, remote sensing, and earth system science. According to Alvares et al. (2018, p. 258):

> As the country did not count at the time of the creation of INPE with the preparation of specialists in space research and development, the Institute began the training of specialized human resources in 1968, with the creation of the first master's course. From that date until 2016, the graduate of INPE graduated 2,123 masters and 794 doctors in the various courses of the Institute.

National space research is also an important tool for fostering the economy. According to the AEB (2012, p. 9) "the global space market grows at an annual average of 6%, thanks to the emergence of new demands for applications and services".

Thus, by providing useful products and services for society, economy, and industry, one perceives its relevance and the need to improve Brazil's Space research.

KNOWLEDGE MANAGEMENT

One of the first definitions of Knowledge Management is by Nicholas L. Henry, elaborated in 1974. This occurred in the 1970s, and it was perceived that at that time the concepts related to the KM were already relevant to organizations. In the author's words, KM is "public policy for the production, dissemination, accessibility, and use of information applied in the formulation of public policies" (HENRY, 1974, p. 189).

According to Dalkir (2011), the ability to manage knowledge is crucial in the knowledge economy, so its creation and dissemination have become important factors of competitiveness, making knowledge a commodity inserted in the products of organizations.

Also according to Dalkir (2011), the advent of access to information, caused unlimited sources of available information and knowledge, which characterized the emergence of the Age of Knowledge to the detriment of the Industrial Age, so that an organization in the Age of Knowledge is able to learn, retain and act on the basis of the best information, knowledge, and know-how available.

In this sense, Davenport and Prusak (1998) state that knowledge exists and is transferred in organizations spontaneously, regardless of formal management processes on it. When formalized, the management that acts in this knowledge lacks specific methods to encourage these spontaneous exchanges. KM is based on improving the organization's existing resources in a knowledge-oriented way.

Cong and Pandya (2003) state that the concept and practices of Knowledge Management have occurred in organizations for a long time, especially informally. Nevertheless, there is a lack of consensus in the proposition of a definition, generating distortions of understanding in several areas where KM is applied. Thus, the terms data and information are often confused with the term knowledge, although they have distinct meanings. Therefore, to understand the concept of KM, first, the distinctions between data, information, and knowledge must be made to clarify their differences and convergences.

Regarding these differences and convergences, Cong and Pandya (2003, p. 26) provide the following definitions:

> In general, data are raw facts. For data to be of value, however, they must be processed (put in a given context) to obtain information, which decision can be made. Knowledge is perceived as meaningful information. The relationship between data, information and knowledge is recursive and depends on the degree of “organisation” and “interpretation”. Data and information are distinguished by their “organisation”, and information and knowledge are differentiated by “interpretation” (Bhatt 2001). So knowledge is neither data nor information. Knowledge is an understanding, and one gains knowledge through experience, reasoning, intuition and learning.

Concerning knowledge, Cong and Pandya (2003) state that it is a derivation of information, when it applies to it, identification of consequences, and creation of connections. Finally, they claim that some experts include wisdom and discernment in their definitions of knowledge so that wisdom would be the use of accumulated knowledge.

As for the increased interest in the theme of Knowledge Management, according to Hislop (2013, p. 1) "the explosion of interest in Knowledge Management among academics, public policy formulators, consultants and entrepreneurs began [...] in the mid-1990s."

Knowledge Management has a multidisciplinary and interdisciplinary nature so that interest in the subject can be observed in different areas of knowledge.

The multidisciplinary nature of Knowledge Management, according to Dalkir (2011) on one hand is advantageous, because it allows different fields of knowledge to identify, establish an understanding, and practice KM actions. On the other hand, according to the author, this diversity results in some challenges, particularly for the most skeptical, who do not accept Knowledge Management as a distinct field of Information Management. It is in this context that it is necessary to establish premises that allow the correct distinction to be made to the terms.

Regarding the understanding and establishment of assumptions about the term Knowledge Management, Alvarenga Neto (2008, p. 2) states that

> [...] knowledge exists only in the human mind and between minds. Knowledge outside this context is seen as information and the KM gains ground from the understanding that its terminology is metaphorical, since knowledge is inherent to human beings and is not transferred or shared easily or spontaneity

Despite establishing as a premise that it is metaphorical terminology, Alvarenga Neto (2002) understands that Knowledge Management occurs through activities seen as a process or system of results optimization to achieve the objectives of the organization and increase its competitiveness so that these activities are understood as practices related to Knowledge Management, different from the processes related to Information Management, but identified in the interdisciplinary objectives of Information Science.

KNOWLEDGE RISKS

Knowledge Risks and their management have developed as a new area of research related to Knowledge Management and intellectual capital, but despite the scientific interest in the subject has grown in recent years, it is still not possible to affirm that there is a unified knowledge so that the need to research it more rigorously (BRATIANU, 2013; DURST, 2019; DURST; HENSCHEL, 2020; MASSINGHAM, 2010).

About this, say Durst and Zieba (2017, p. 52):

> several researchers have started to examine various types of knowledge risks, such as risk of knowledge loss, knowledge leakage, or knowledge hiding. This is a promising development, but given the fact that the study of knowledge risks is a recent phenomenon these studies have addressed very specific issues and thus, produced a fragmented understanding of the topic

Below, some of the Risks most frequently identified in the scientific literature will be explained.

Knowledge loss

Knowledge loss can be defined as a situation when an organization loses a part or all of its crucial knowledge as a consequence of for example employee leaving a company, employee poaching, or some technical faults (DURST; ZIEBA, 2019).

It occurs when an individual with valuable knowledge exits an organization. The problem is increasing due to workforce mobility and our ageing society and it is one of the most important knowledge risks for any organization since it influences directly the dynamics of organizational knowledge (BRATIANU, 2013; MASSINGHAM, 2018).

Knowledge leakage

Knowledge leakage is a metaphor for that knowledge that escapes from certain confinement or boundary and is used by others, with possible negative effects for the owners of that knowledge, especially when it comes to competitive advantage. The risk of knowledge leakage is high when it happens with new ideas and technologies and when there is a certain secret concerning the new products or services a company is preparing for the market. (BRATIANU, 2013)

Intentional knowledge leakage refers to the focal firm's loss of private knowledge to partners through opportunistic activities such as private learning and unauthorized imitation (JIANG *et al.*, 2013).

Knowledge spillover

According to Ferenhof, (2016), there is a slight difference between knowledge spillover and knowledge leakage, relating to the way the knowledge exceeds company boundaries:

> Knowledge spillover, as stated by Albornoz et al. (2009), may happen when companies transfer knowledge that encompasses both technology and know-how of its foreign affiliates by some kind of leak and which is then absorbed by those firms. On the other hand, as stated by Durst and Ferenhof (2014), knowledge leakage may occur in situations when parts of their business functions are outsourced, or parts of the core knowledge are transferred to others. (FERENHOF, 2016, p. 47)

Knowledge spillover takes place when valuable knowledge spills out of the organization to competitors who use this knowledge to gain competitive advantage (DURST; ZIEBA, 2019). As Bratianu (2013) remarks, knowledge spillover is considered to have a certain influence on innovativeness and growth of companies or industries.

> Knowledge spillover is a metaphor for the knowledge that spills over the boundary of a company towards other companies as users of that knowledge, which may lead to the risk of reducing or losing the competitive advantage (Durst & Zieba, 2017; Inkpen, 2000; Tseng et al., 2011). (BRATIANU, 2013, p. 598).

In addition to the risks explained above, during the research the other risks found in the literature will also be addressed, such as knowledge attrition, knowledge waste, knowledge hiding, knowledge hoarding, as well as the other ones that may be identified.

2.2. METHODOLOGY

This session presents the methodology that will be used in the research, so that its characterization, description of the universe researched, its procedures and stages, and the instruments of data collection will be explained.

This project is related to the Knowledge Management Research Project in Public Organizations, in the information organization research line of the Information Sciences course of the University of Brasília. To discuss the subject, concerning the methodological aspect, the scientific method that will be used in the development of the research will be presented, explaining how it will be possible to achieve the objectives of the study by applying the proposed methodology.

The methodological approach to be used in the research will be predominantly qualitative, with possible quantitative treatment when analyzing the collected data. Qualitative analysis is justified since the subject of the proposed study is a specific area, which are the different types of organizations and sectors that are part of the Brazilian Space Program, which will provide part of the elements of the investigation to describe the reality found from the observation of the occurrence of different types of Knowledge Risks.

About the type of approach, the inductive method will be used, because it is a process by which, based on private data, a general truth not contained in the examined parties is inferred (MARCONI; LAKATOS, 2010). In this sense, the research will start from the observation of the possible Knowledge Risks found in the Brazilian Space Program, so that it is possible to identify and analyze them and propose a conceptual framework for identifying and mitigating these risks. Thus, any generalization will not be aprioristic, but rather "verified from concrete facts sufficiently confirming this reality" (GIL, 2008, p. 10).

Regarding the type of procedure, we intend to use the monographic method, which consists of the "study of certain individuals, professions, conditions, institutions, groups or communities, to obtain generalizations" (MARCONI; LAKATOS, 2010). It is thus understood that the result may be applied to similar institutions and/or processes. This understanding can be corroborated by Gil's definition:

> The monographic method assumes that the study of an in-depth case can be considered representative of many or even all similar cases. These cases can be individuals, institutions, groups, communities, etc. (GIL, 2008)

Concerning technical procedures related to the data sources, a bibliographic survey will be made in the scientific literature and will be based on articles, specialized books, and documents.

The different organizations that make up the Brazilian Space Program, composed of sectors of government, universities, and companies in the Space sector, were defined as the universe of research.

To achieve the proposed objectives, we intend to follow the steps explained in Frame 1, as follows:

Frame 1 - Methodological Procedures

General objective	Specific objectives	Procedures
Structure, from scientific and situated knowledge, a methodological framework that helps in mitigating the knowledge risks identified in the different types of organizations and sectors that are part of the Brazilian Space Program.	To present the Brazilian Space Program, detailing its characteristics, history, development, future perspectives, and its actors, coming from academia, government and industry	- Bibliographic survey in databases related to space technology.
	Identify and analyze, based on the different categories found in the literature, the types of knowledge risks observed in organizations and sectors that are part of the Brazilian Space Program.	- Bibliographic survey in databases related to Knowledge Management to verify the Knowledge Risks identified in the scientific literature - Investigation through questionnaires and interviews applied to the research universe, which will be the source of data collection.
	To propose a conceptual framework for identify and mitigate knowledge risks in the Brazilian Space Program.	- Analysis of the results found, revisiting the identified risks - Elaboration of the conceptual framework
	Check the proposed model.	- Framework validation

Source: Author (2022)

3. FINAL REMARKS

The identification of knowledge risks involved in the scope of Brazilian space research, with the aim of pointing out guidelines for optimizing results through risk mitigation is an unprecedented proposition in the scientific literature of Information Science.

Until the current phase of the research, it has been possible to relate different Knowledge Risks identified in institutions linked to the Brazilian Space Program. It is expected, however, that with the model that will be proposed, it will be possible to identify them more clearly, and, most importantly, mitigate them.

In this sense, taking as a premise the importance of the possibility of results that can be achieved by national space research, it is expected that it will be possible to offer a framework that can be implemented, so that the Brazilian Space Program can be optimized.

It is expected, in turn, that when implementing the model that will be proposed, there will be a real possibility of improving the knowledge generation cycle that permeates the development of Brazilian space research, avoiding the negative effects of Knowledge Risks, whether this is loss resulting from early obsolescence or even the lack of knowledge transmission that could be able to perpetuate the informational competences of the researchers involved.

Finally, as the guidelines that will be proposed because of the research will be established with a theoretical basis focused on national space program, it is expected that it will be possible to adapt them to other S&T research units, respecting their limitations and diversity of themes.

Acknowledgments

The author gratefully acknowledges the financial support from FAP-DF - Brazilian Federal District Research Support Foundation, for the financial incentive that will allow me to participate in person at the 17th International Conference on Knowledge Management 2022

REFERENCES

AGÊNCIA ESPACIAL BRASILEIRA (AEB). **Programa Nacional de Atividades Espacial: PNAE / 2005 - 2014**. Brasília: Ministério da Ciência, Tecnologia e Inovação, Agência Espacial Brasileira, 2005.

AGÊNCIA ESPACIAL BRASILEIRA (AEB). **Programa Nacional de Atividades Espaciais: PNAE 2012 - 2021**. Brasília: Ministério da Ciência, Tecnologia e Inovação, Agência Espacial Brasileira, 2012.

ALVARENGA NETO, R. C. D. **Gestão do Conhecimento em organizações**. São Paulo: Saraiva, 2008.

ALVARES, L. M. A. de R. *et al.* Análise da pesquisa espacial brasileira sob a ótica da Biblioteca Digital Brasileira de Teses e Dissertações. **Ciência da Informação**, *[S. l.]*, v. 45, n. 3, p. 251–266, 2018. Disponível em: http://revista.ibict.br/ciinf/article/view/4064.

ALVAREZ, G. R.; CAREGNATO, S. E. A ciência da informação e sua contribuição para a avaliação do conhecimento científico. **BIBLOS**, *[S. l.]*, v. 31, n. 1, p. 09–26, 2018. Disponível em: https://doi.org/10.14295/biblos.v31i1.5987.

BENITE, E. V. **A tecnologia espacial como ferramenta para o monitoramento ambiental: um caminho para o desenvolvimento sustentável**. Brasília: *[s. n.]*, 2006.

BORKO, H. Information science: What is it? **American Documentation**, *[S. l.]*, v. 19, n. 1, p. 3–5, 1968. Disponível em: https://doi.org/10.1002/asi.5090190103.

BRATIANU, C. A Holistic Approach to Knowledge Risk. **Management Dynamics in the Knowledge Economy**, *[S. l.]*, v. 6, n. 4, p. 593–607, 2013. Disponível em: https://doi.org/10.25019/MDKE/6.4.06.

CALAZANS, A. T. S. Conceitos e uso da informação organizacional e informação estratégica.

Transinformação, Campinas, v. 18, n. 1, p. 63–70, 2006. Disponível em: https://doi.org/10.1590/S0103-37862006000100006.

CASTELLS, M. **A Sociedade em Rede**. Tradução Roneide Venancio Majer. 20. ed. São Paulo: Paz e Terra, 2019.

CONG, X.; PANDYA, K. V. Issues of Knowledge Management in the Public Sector. **Electronic Journal of Knowledge Management**, *[S. l.]*, v. 1, n. 2, p. 25–33, 2003. Disponível em: http://www.ejkm.com/issue/download.html?idArticle=17.

DALKIR, K. Knowledge management in theory and practice. *[S. l.]*, p. 485, 2011.

DAVENPORT, T. H.; PRUSAK, L. **Conhecimento empresarial : como as organizações gerenciam o seu capital intelectual**. Rio de Janeiro: Campus, 1998.

DURST, S. How far have we come with the study of knowledge risks? **VINE Journal of Information and Knowledge Management Systems**, *[S. l.]*, v. 49, n. 1, p. 21–34, 2019. Disponível em: https://doi.org/10.1108/VJIKMS-10-2018-0087

DURST, S.; HENSCHEL, T. **Knowledge Risk Management**. 1. ed. *[S. l.]*: Springer International Publishing, 2020. (Management for Professionals).*E-book*. Disponível em: https://doi.org/10.1007/978-3- 030-35121-2

DURST, S.; ZIEBA, M. Knowledge risks - towards a taxonomy. **International Journal of Business Environment**, *[S. l.]*, v. 9, n. 1, p. 51, 2017. Disponível em: https://doi.org/10.1504/IJBE.2017.084705

DURST, S.; ZIEBA, M. Mapping knowledge risks: towards a better understanding of knowledge management. **Knowledge Management Research & Practice**, *[S. l.]*, v. 17, n. 1, p. 1–13, 2019. Disponível em: https://doi.org/10.1080/14778238.2018.1538603.

FERENHOF, H. A. Recognizing Knowledge Leakage and Knowledge Spillover and Their Consequences. **International Journal of Knowledge and Systems Science**, *[S. l.]*, v. 7, n. 3, p. 46–58, 2016. Disponível em: https://doi.org/10.4018/IJKSS.2016070104.

GIL, A. C. **Métodos e técnicas de pesquisa social**. 6a. ed. São Paulo: Atlas, 2008.

HENRY, N. L. Knowledge Management: A New Concern for Public Administrationitle. **Public Administration Review**, *[S. l.]*, p. 189–196, 1974.

HISLOP, D. **Knowledge management in organizations: a critical introduction**. 3. ed. Oxford: Oxford University Press, 2013.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS (INPE). **Plano Diretor do INPE 2016-2019**. São José dos Campos: Instituto Nacional de Pesquisas Espaciais, 2016. *E-book*. Disponível em: http://www.inpe.br/noticias/arquivos/pdf/Plano_Diretor_2016-2019.pdf.

JIANG, X. *et al.* Managing knowledge leakage in strategic alliances: The effects of trust and formal contracts. **Industrial Marketing Management**, *[S. l.]*, v. 42, n. 6, p. 983–991, 2013. Disponível em: https://doi.org/10.1016/j.indmarman.2013.03.013

LEITE, J. S. G.; PINHO NETO, J. A. S. de P. O Pensamento, a Análise e a Reflexão em Tempos de Excesso e Obsolescência da Informação. **Informação & Tecnologia**, *[S. l.]*, v. 1, n. 2, p. 34–41, 2014. Disponível em: http://www.periodicos.ufpb.br/ojs/index.php/itec/article/view/22393.

MARCONI, M.; LAKATOS, E. M. **Fundamentos de metodologia científica.** 7a. ed. São Paulo: Atlas, 2010.

MARTELETO, R. M. A pesquisa em Ciência da Informação no Brasil: marcos institucionais, cenários e perspectivas. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 14, n. spe, p. 19–40, 2009. Disponível em: https://doi.org/10.1590/S1413-99362009000400003.

MASSINGHAM, P. Knowledge risk management: a framework. **Journal of Knowledge Management**, *[S. l.]*, v. 14, n. 3, p. 464–485, 2010. Disponível em: https://doi.org/10.1108/13673271011050166

MASSINGHAM, P. R. Measuring the impact of knowledge loss: a longitudinal study. **Journal of Knowledge Management**, *[S. l.]*, v. 22, n. 4, p. 721–758, 2018. Disponível em: https://doi.org/10.1108/JKM-08-2016-0338

MOLTZ, J. C. Brazil's space program: Dreaming with its feet on the ground. **Space Policy**, London, v. 33, n. P1, p. 13–19, 2015. Disponível em: https://doi.org/10.1016/j.spacepol.2015.05.001.

OTLET, P. **Documentos e documentação: discurso pronunciado no Congresso de Documentação Universal, Paris, 1937**. Rio de Janeiro: Imprensa Nacional (Separata), 1937. *E-book*. Disponível em: http://www.conexaorio.com/biti/otlet/.

OTLET, P. **Traité de documentation. Le livre sur le livre**. Bruxelles: Mundaneum, 1989. *E-book*. Disponível em: https://archive.org/details/OtletTraitDocumentationUgent/page/n1.

Track – Knowledge Management II

Organizational memory in the startup deployment phase

Firmansyah David
Department of Information Technology, Padang Institute of Technology, Jalan Gajah Mada Kandis Nanggalo,25143, Indonesia
E-mail : firman@itp.ac.id

Dede Wira Trise Putra†
Department of Information Technology, Padang Institute of Technology, Jalan Gajah Mada Kandis Nanggalo,25143, Indonesia
E-mail : dedewtp339@yahoo.com†

The study in this paper aims to describe the mechanism of knowledge recall in the context of organizational memory. We approached a startup involved in the development of a digital marketplace and observed how the founders interacted with the accumulated knowledge. With observation as the research method and longitudinal data collected, we found the founding group in the way they understood organizational memory. There are three mechanisms of remembering organizational memory including identification, adaptation, and correction. The identification mechanism is the highest level of recall ability followed by adaptation and correction. This study implies that founders with different backgrounds have different ways of remembering organizational memory.

1. INTRODUCTION

The implementation stage is a critical point in executing the proposed value for IT-based startups. In this phase, the founders should, at least, have a common understanding of the direction and goals of the organization as well as its main tasks and functions. In this line, the founders must have gone through the process of integrating knowledge as the resolutions of different perspectives and views (Kenney and Gudergan, 2006; Liao, 2018). Nevertheless, the main challenge of this stage is the founders should maintain the goal of the startup as the results of integration of knowledge (Chang and Lee, 2008).

Individuals' ability to 'retain' common knowledge is demonstrated by the way they make that knowledge explicit (O'Toole, 2011). For example, founders demonstrate their understanding of a process by knowing the milestones of the various processes which are being undertaken. However, the key factor for successful market deployment depends on the founder's strategy - such as maintaining, developing and understanding the value proposed, niche and technology platforms (Brunn et al., 2002). This may be easy if the founders consist of people who have the same professional knowledge/background, making it easier to control startup implementation. On the other hand, it becomes more challenging if the founders consist of individuals from different backgrounds or professions. For this reason, seeing how organizational members maintain a predetermined value proposition and in accordance with agreed business logic and processes becomes a discourse on organizational theory at a startup.

By visiting knowledge-based theory and organizational memory, this paper aims to see how founders in a multi-field startup execute and maintain the agreed values and plan. The study in this paper describes the behavior of founders in relation to the strategies they take to keep up with changes in the highly dynamic startup environment. The paper consists of three parts. The first part is a literature review discussing the theoretical framework. The second part is research setting and the last part is discussion and conclusion.

2. DEVELOPMENT

Organizational memory has become a key asset in startup development and innovation (Mariano and Cassey, 2016). In organizational theory, organizational memory is highly related to the dynamics of knowledge transfer - do organizational members succeed in transferring and acquiring knowledge from their colleagues and counterparts (Muskat and Deery, 2017). One might say that knowledge resides in an organization either by explicit or tacit mechanisms (Harvey, 2012) which can be labeled as organizational memory (Mariano and Cassey, 2016). Nevertheless, Muscat and Deery (2017) suggested that tacit knowledge becomes difficult knowledge or rarely turns into explicit knowledge in the post-event phase and this is the cause of organizational memory failure. Although organizational memory has been embodied in various repositories and Knowledge Management Technology supports knowledge acquisition and retention (Argote, 2013; Adobor et al., 2019); however, an uncertain startup environment requires good dynamic skills and founders need to be able to identify which knowledge to use to tackle the situation (Gonzalez et al., 2017).

Founders at startups will certainly face uncertain conditions in achieving their vision and mission. Chien and Tsai (2012) show that organizations that have dynamic capabilities can improve organizational memory and this shows that knowledge sources and learning mechanisms have a positive effect on the level of dynamic capabilities. In line with this, Tseng and Lee (2014) argue that dynamic capabilities are an intermediary between knowledge management and organizational performance, and then dynamic capabilities will in turn increase the competitiveness of an organization. For example, the support of IT is also the mediating variable in leveraging organizational dynamics capabilities via knowledge management systems (Wang et al., 2007). Eventually, a good implementation of a knowledge management system will increase knowledge transfer effectiveness and later on to improve organizational performances.

Since organizational memory and dynamic capabilities are very important to improve organizational performance, the correlation between the two is also important to study. Founders as members of the organization have a very important role in remembering and selecting and using the knowledge that has been collected in accordance with the conditions of the startup's external environment. The existence of tools such as visuals can lead the founder to 'back' to the memory of the organization (Bera et al., 2011). The mechanisms for 'bringing back' organizational memory may vary depending on how the founders identified and integrated knowledge into the ongoing process. Therefore, the research in our paper attempts to synthesize knowledge identification mechanisms under uncertain conditions. In this paper, we approach a startup engaged in the digital marketplace to uncover the mechanisms for identifying organizational memory.

3. SETTING, DATA AND METHODS

We approached a startup with a vision to build a digital organic food market in Indonesia. Its mission is to create a marketplace that integrates organic communities including producers, retailers, entrepreneurs and consumers. Startups have formulated their value proposition, developed an e-marketplace platform and until now the startup is testing the platform. Dynamic capacity is demonstrated by how the founders discuss and present their knowledge of the business model, target market, and customer relations. This study is a longitudinal study using observational methods. To conduct the study, the authors engaged with a group of IT founders and developers to collect long-term behavioral data (Kawulich, 2005). Respondents consist of seven founders and eight respondents from IT developers. To collect data, the authors used a covert method in which respondents were not aware that they were being observed (Costa, 2020). Data was collected in the period June 2020 - May 2021 resulting from notes and conversations on social media and groups. To capture the knowledge accumulation process for the organizational memory data were collected into three stages. The first stage - the first of the value proposition stage. The founders agree on their vision and mission and all meetings are documented in the meeting minutes. The second stage is at the stage of developing the marketplace platform. The data collected is related to how the founders from various backgrounds perceive the form and application. The last stage is the platform development stage. In this phase, data is collected in relation to the way the founders monitor implementation and recall organizational memory.

4. RESULTS

Our study aims to provide empirical evidence on how founders identify or remember certain knowledge so as to adapt to the unpredictable changes of the startup environment. The condition in which founders recall the so-called accumulated knowledge or organizational memory is shown by the expression and behaviors. From the observational data we get four mechanisms in extracting organizational memory. The four mechanisms are identification, submission, adaptation and correction. The four mechanisms were seen in the development and implementation phase of the marketplace platform. Even more interesting, these four mechanisms emerge from all founders and IT professionals where the stored knowledge is reminded by each founder.

4.1. Identification

The first mechanism that is most widely displayed in the data is the identification mechanism. This means that founders recognize the predictable process that will occur and founders are able to interpret it according to the values, vision and mission of the startup organization. One of the founders with an agricultural background was able to identify the process of implementing a marketplace platform in accordance with the organization's vision and mission. The founder seems to be able to recognize the knowledge that was agreed upon in the previous phase. The founder suggests,

"I mean so Mr _. On the _______________ website, it is best not to explicitly display the ___________narrative. In my opinion, it's really just a matter of communication strategy to show that even though _marketplace is created and owned by ________________, the ultimate goal or vision is to support the achievement of true organic farming ideals. Our company is only a legal entity in an effort to comply with existing regulations."

This identification mechanism shows that the founders are able to see two sides, both from organic farming and from information technology support. Based on the narrative above, this founder understands the knowledge defined as a marketplace and its acquisition with an organic farming business model. An IT background founder show his ability to recognise the value of product by saying,

*"Before I forget and go up to print the sticker, I plan to direct this QR Code link to a special page that *contains traceability information* for all partners who work with PasarOrganik"*

The keyword stated by this founder is 'traceability' which is the noble values of the organic food product supply chain. Thus, the identification process is carried out by consciously returning to the values, vision, and mission that are continuously instilled at the beginning of the discussion of value proportions.

4.2. Adapting

The second mechanism is the condition in which the founder recognizes and identifies but is not fully in-depth in understanding knowledge. In this line, The founder recognizes the knowledge that has been socialized but the founder adapts the new knowledge. This is also known as verification of socialized knowledge. A founder from the background of organic agriculture suggests that,

"Interesting issue: what about organic farming which encourages farmers to develop seeds and seedlings independently without relying on the seed industry? Will our marketplace also support and facilitate the marketing of organic seeds planted twice by breeders through the marketplace platform?"

In this direction, the founder seeks to adapt the embedded knowledge to the new knowledge that occurs in the implementation process. Adaptation also means that the founders acquire knowledge both in terms of organic farming and in terms of Information Technology. The existence of a two-way acquisition of new and old knowledge makes the adaptation process take longer or even faster depending on how the founder responds and concludes. The adaptation mechanism is indicated by the founders' efforts to combine or revise the knowledge that has been memorized with other sources of knowledge. Other founders show their perspectives in adapting new knowledge to old ones, *"Maybe it's not just greedy but maybe the business model is no longer "profitable" so it's a burden to partners. This is the true side effect of "burning money". In addition to competition with the same business model. The saturation of partner and market turnover is not immediately addressed by renewing the business model but using the traditional method, namely incentives or mere incentives."*

4.3. Correction

The third mechanism is correction. In this case, the founders were involved in correcting the knowledge embedded as organizational memory. This can be found, usually, in discussion between founders who have different backgrounds. From the case, it showed the two different backgrounds which are agriculture and IT contributes to the discourse of value and norms of how the digital marketplace should be. An IT founder give perception of how hard to sell product online,

"I admit... selling online is not as easy as it used to be... customers are getting smarter, slashing prices are happening... plus the country border is gone...it takes more than just uploading photos and descriptions."

The founders act to 'correct' the knowledge of their peers and incorporate their logic into what they know by virtue of their profession. The direction of correction can be in the form of positive construction, namely by increasing knowledge to become richer and dynamic or in a negative direction where there is debate and there is no common ground and must be resolved by the third founder. A founder suggests that,

"I think we show our good intentions in our concrete actions, sir _. For example, later in our marketplace system, we really ensure that a percentage of our profits will be distributed to all organic market partners automatically and openly."

Another example of a correction mechanism was pointed out by a founder,

"Agreed, but what I mean is how we can encourage partners to innovate in terms of packaging and processing, then we market it."

5. FINAL REMARKS

The study in this paper aims to describe the mechanism of remembering and using knowledge in the context of organizational memory. The bottom line is that under uncertain conditions, founders can use their organizational memory to evaluate initial values and to achieve their organizational goals. It is necessary to make quick decisions to cope with a rapidly changing environment. Using startups as a research setting, the study in this paper has captured a mechanism called 'calling up' organizational memory. Any action that can trigger remembering the accumulation of knowledge is important to increase the effectiveness of both tacit and explicit knowledge accumulation (Mariano and Cassey, 2016). From the results, we highlight that there are three mechanisms by which founders engage when they use their organizational memory. First, they are able to 'identify' what has been accumulated and embedded as organizational knowledge. In this mechanism, the founders are given a clue from their memory by tracing back what was written in the minutes of the meeting. We also highlight that this is the highest level of organizational memory recall because it is not related to prior knowledge or professional experience. The second mechanism is 'adaptation'. These mechanisms relate knowledge that founders can recall to organizational memory but they cannot definitely identify specific knowledge. Founders tend to discuss with their peers as an 'adaptation' mechanism which is demonstrated as the way they acquire new knowledge. We label this mechanism as the second level of organizational memory recall. The third mechanism is 'correction'. This mechanism illustrates that founders tend to correct the knowledge of their peers and insist on their knowledge. The correction mechanism is so obvious when two founders from different backgrounds defend their opinions in a debate. We highlight that this mechanism is the lowest rate of recall to organizational memory.

With the above explanation we can conclude that there are levels in remembering organizational memory. Starting from the highest degree, namely by remembering directly what has been agreed upon and this applies to members of the organization who have the ability to remember what has been written or embedded in the organization. At this highest level, it can be seen that founders with backgrounds related to agreed values tend to identify accumulated knowledge more quickly. Meanwhile, the further an organization's values are agreed upon with the founder's background, the more likely he is to adjust or even correct what his colleagues propose. This study has an implication to further exploration regarding the relationship between the ability of knowledge recalling and the norm and values of founders. An institutional logic perspective could be the framework for describing how individuals tackle the different norms and values.

References

Adobor, H., Kudonoo, E., & Daneshfar, A. (2019). Knowledge management capability and organizational memory: a study of public sector agencies. International Journal of Public Sector Management.

Argote, L. (2013). Organizational memory. In Organizational learning (pp. 85-113). Springer, Boston, MA.

Bera, P., Burton-Jones, A., & Wand, Y. (2011). Guidelines for designing visual ontologies to support knowledge identification. Mis Quarterly, 883-908.

Brunn, P., Jensen, M., & Skovgaard, J. (2002). e-Marketplaces:: Crafting A Winning Strategy. European Management Journal, 20(3), 286-298.

Chang, S. C., & Lee, M. S. (2008). The linkage between knowledge accumulation capability and organizational innovation. Journal of knowledge management.

Chien, S. Y., & Tsai, C. H. (2012). Dynamic capability, knowledge, learning, and firm performance. Journal of Organizational Change Management.

Costa, F. A. D. (2020). Covert and overt observations in pharmacy practice. In Pharmacy Practice Research Methods (pp. 93-114). Springer, Singapore.

Gonzalez, R. V. D., & Melo, T. M. (2017). Linkage between dynamics capability and knowledge management factors: A structural equation model. Management Decision.

Harvey, J. F. (2012). Managing organizational memory with intergenerational knowledge transfer. Journal of Knowledge Management.

Kawulich, B. B. (2005, May). Participant observation as a data collection method. In Forum qualitative sozialforschung/forum: Qualitative social research (Vol. 6, No. 2).

Kenney, J. L., & Gudergan, S. P. (2006). Knowledge integration in organizations: an empirical assessment. Journal of knowledge Management.

Liao, Z. (2018). Institutional pressure, knowledge acquisition and a firm's environmental innovation. Business Strategy and the Environment, 27(7), 849-857.

Mariano, S., & Casey, A. (2016). The dynamics of organizational routines in a startup: The EREDA model. European Management Review, 13(4), 251-274.

Muskat, B., & Deery, M. (2017). Knowledge transfer and organizational memory: An events perspective. Event Management, 21(4), 431-447.

O'Toole, P. (2011). How Organizations Retain Knowledge. In How Organizations Remember (pp. 131-154). Springer, New York, NY.

Rowlinson, M., Booth, C., Clark, P., Delahaye, A., & Procter, S. (2010). Social remembering and organizational memory. Organization studies, 31(1), 69-87.

Tseng, S. M., & Lee, P. S. (2014). The effect of knowledge management capability and dynamic capability on organizational performance. Journal of Enterprise Information Management.

Wang, E., Klein, G., & Jiang, J. J. (2007). IT support in manufacturing firms for a knowledge management dynamic capability link to performance. International journal of production research, 45(11), 2419-2434.

Comparing knowledge management and wisdom management

MARIA JAKUBIK
Ronin Institute, Social Sciences, 127 Haddon Place, Montclair, NJ 07043, New Jersey, USA
E-mail: maria.jakubik@ronininstitute.org
and
University of Pécs, Faculty of Business and Economics, 80 Rákóczi út, Pécs, HU-7622, Hungary
E-mail: jakubik.maria@pte.hu

To provide solutions to the world's global challenges, there is an urgent demand for wise organizations, wise leadership, wise workers, and most importantly, for wise actions. Since the mid-1990s, Knowledge Management (KM) as a discipline and practice has emerged internationally and it went through several phases of development. Similarly, in the last four decades, we experience a growth in wisdom research, and in intense discourses about Wisdom Management (WM) as a possible venue for dealing with wicked problems. The dilemma is, whether the present phase of KM would be able to address the global problems of the world. Therefore, this paper seeks to answer the question if WM will complement or replace KM. The purpose is to compare KM and WM. The research methodology is an explorative approach of their main characteristics and concepts in a selected range of literature. The findings are presented in a framework to show the similarities and differences of KM and WM. Keywords: Knowledge Management (KM), Wisdom Management (WM), phases of KM, wisdom economy, wise leaders, wisdom workers, global problems of the world.

1. Introduction

The world faces global challenges (e.g., Barnett, 2015; Lufkin, 2017; Maxwell, 2021). Barnett argues that the world faces problems such as "disease, illiteracy and unduly limited education, climate change, dire poverty, lack of capability and basic resource, misunderstandings across communities, excessive use of the earth's resources, energy depletion and so on and so on" (Barnett, 2015, 89). Experts identified fifty grand challenges for the 21st century (Lufkin, 2017) in the following categories: health and humanity, artificial intelligence (AI), cities and global development, energy, and future of the Internet, media and democracy. According to Maxwell, the world faces ten fundamental challenges to be solved: "the climate crisis; the current pandemic; the destruction of the natural world, catastrophic loss of wild life, and mass extinction of species; lethal modern war; the spread of modern armaments; the menace of nuclear weapons; pollution of earth, sea and air; rapid rise in the human population; increasing antibiotic resistance; the degradation of democratic politics" (Maxwell, 2021, Preface). Therefore, there is an urgent demand for wise leadership, wise workers, wise organizations, and most importantly for wise actions to provide solutions to the world's global challenges.

To tackle the major problems of the world, there have already been great advancements in information and communication technology (ICT), artificial intelligence (AI), robotics, engineering, biology, and in other disciplines. This paper focuses on Knowledge Management (KM). The research problem is, if the current phase of KM would be able to help people and organizations to solve current problems, or if KM would need to move to Wisdom Management (WM). The purpose is to explore and compare KM and WM. The research methodology is an explorative approach of the main characteristics and concepts of KM and WM in a selected range of literature. The findings are presented in a framework to show the similarities and differences of KM and WM.

The rest of the paper is organized in three sections. Section two presents the research question, the objectives, the research framework, and the research method. Section three demonstrates the findings by overviewing the evolutionary stages of KM, and by highlighting the emerging discourses about WM. The discussion and conclusions section, first, presents a framework that demonstrates the multidisciplinary character of KM and WM with their main concepts and theories; next, discusses how the research objectives were met; and finally, concludes about the possible future phase of KM.

2. Research Question, Objectives, Framework, and Methods

The research question is: *Will WM complement or replace KM?* The main purpose of this paper is to compare concepts and theories of KM and WM, in order to find out their similarities and differences. For this purpose, a research framework is presented in Fig.1.

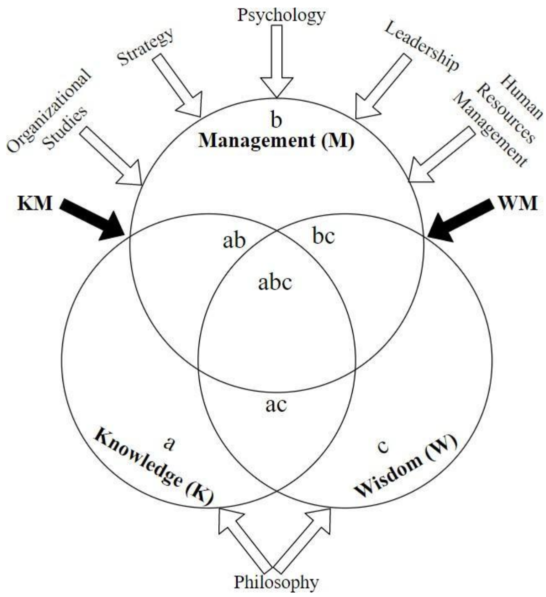


Fig.1. Research framework

The research framework shows the relationships of three concepts: knowledge (K); wisdom (W); and management (M). It also demonstrates the disciplines that are the major foundations and sources of these concepts. This paper focuses on KM and WM.

The research methodology is an explorative approach of the main characteristics and concepts of KM and WM in a selected range of literature. First, the evolutionary stages of KM are presented. The objective is to demonstrate the shifts in focus in each phase of development, and to recognize the characteristics of the current phase of KM. Next, the intense discourses about wisdom economy, wisdom workers, and WM characteristics are explored. The objective is to understand the need for wisdom alongside with knowledge in the next stage of KM to effectively face global challenges ahead.

3. Findings

3.1. Evolution of Knowledge Management

Historical roots of KM could be dated back to late 1980s, when in different forums, conferences, journals, books started to use the term KM. However, internationally KM gained momentum in the mid-1990s (Dalkir, 2011, 15-19). Massingham argues that “There have been several attempts to synthesize the chronology of knowledge management” (Massingham, 2020, 40). He outlines the history of KM (*Ibid.*, 39-43), and presents a timeline of KM history that goes back to 400 BC, to ancient Greek philosophers talking about epistemology and knowledge (*Ibid.*, 42). Similarly, Jashapara (2011, 20-27) takes a historical perspective on KM. He goes back to 3,000 BC, Mesopotamia, 2,300BC Syria, ancient Greece, Rome, and other old civilizations. His point is that archeological findings (e.g., architecture, poems, cave paintings, clay padlets, sacred scriptures, codexes, hand-written books, first printed books) demonstrate writing skills, recording of data, information, and knowledge with the purpose of knowledge sharing.

Other scholars (e.g., Jakubik, 2007; Koenig and Srikantaiah, 2007; Jakubik, 2011; Serenko, 2013; North and Kumta, 2014; Agostini *et al.*, 2020; Serenko, 2021) identified different phases in the development of KM. From the KM theory development perspective, Jakubik (2007 and 2011) identified five distinctive phases as follows:

(i) foundation of the KM theory in the mid-1990s,
(ii) the unified model of dynamic knowledge creation,
(iii) emphasis of the context and the roles of leaders and managers,
(iv) the justification process of organizational knowledge, and
(v) the need for a new theory of the knowledge-based firm, with the focus needing to be on situation, process, action, and change.

From a chronological and main focus points of view, Serenko (2013) defined four generations of KM such as:

(i) prior to mid-1990s: techno-centric view of knowledge processes, knowledge sharing is initiated and driven by management, focus on explicit knowledge;
(ii) mid-1990s to early 2000s: human factors and intellectual capital become important, organizational learning, social and cultural aspects, knowledge sharing processes are initiated and driven by individual employees as their daily practices;
(iii) early 2000s-2013: culture and contextual aspects become important, social learning, collaboration, democratization of knowledge, involvement, managing knowledge as a flow; and
(iv) since 2013: knowledge is seen as a relationship, a shift to the mind economy and intangible knowledge, and to networking organizations.

In a most recent paper, however, Serenko (2021) argues that from research focus and methods perspectives, there could be five phases detected in the KM discipline:

(i) 1996-2001 initiation,
(ii) 2002-2006 early development,
(iii) 2007-2012 rigor and consolidation,
(iv) 2013-2016 methodological advancement, and
(v) 2017-2019 maturity phase.

Serenko (2021) concludes: “A unique attribute of the Methodological Advancement and Maturity phases is a high degree of specialization when over half of all studies are conducted in a unique context of specific topics, publication forums, geographic regions and groups of people” (*Ibid.*).

Knowledge Management has evolved through time in close influence from other disciplines and theories. A good example of the influence of different fields of studies on KM are the classification of Koenig and Srikantaiah (2007). They claim that KM has three stages of development:

(i) Internet, IT, and intellectual capital (IC),
(ii) human relations (HR), Communities of Practice (CoP), organizational culture, organizational learning (OL), learning organizations (LO), social practices, and
(iii) content management and taxonomies.

Similar to Koenig and Srikantaiah (2007), North and Kumta (2014, 37-39) applies the maturity model and argue that KM has four phases of maturity:

(i) data and symbols as knowledge: IT solutions, data processing,
(ii) information as knowledge: information processing, search for meaning,
(iii) process view of KM: knowledge as know-how, organizational knowledge, networking, and
(iv) practice view of KM: knowledge-based management of organizations.

In brief, from the classifications above it is apparent that, even though computer sciences, IT, ICT have dominated the early stages of KM development, the focus of KM research and practice has shifted towards human practices, relationships, networking, and learning. This trend is validated by a bibliometric analysis of the KM literature by Agostini *et al.* (2020). Their objective was to identify the evolution of topics in KM. They studied 85 KM publications in three periods: 1998-2010; 2011-2014; and 2015-2019. Their findings confirm that social and human themes, social networking, innovation, coopetition are the new emerging topics in the KM literature.

3.2. Discourses About Wisdom Management

Wisdom has always been in the center of philosophical discussions. Nevertheless, since the mid-1980s, we experience a revival of wisdom research in several other disciplines too (e.g., in psychology, education, sociology, organizational studies, human resources management (HRM), strategy, leadership, management). Concurring with this paper, authors (e.g., Karami *et al.* 2020; Karami and Parra-Martinez, 2021; Sternberg and Karami, 2021) point out that wisdom becomes more important than ever before due to the global problems in our society. According to Solé (2017, 55-61), wisdom as a higher degree of knowledge, that makes it possible to act wisely. Similar to this paper, he raises vital questions about the future of KM: Is KM still alive? Is KM dead? Is there an evolution towards WM? Surprisingly, in the KM literature, there has been none or very little attention paid on wisdom (Jakubik and Müürsepp, 2021). Therefore, exploring discourses about wisdom and WM could contribute to the current and to the possible future stage of KM.

In the management literature, wisdom and WM are discussed from different angles (e.g., McKenna and Rooney, 2005; McKenna *et al.*, 2009; Rooney *et al.*, 2010; Nonaka and Takeuchi, 2011; Müürsepp, 2013a and b; Banerjee, 2014; Ekmekçi *et al.*, 2014;

Nonaka *et al.*, 2014; Solé, 2017; Stebbins, 2017; Bachmann *et al.*, 2018; Jakubik, 2020a; Müürsepp, 2021). From the HRM perspective, authors (McKenna and Rooney, 2005; Rooney *et al.*, 2010) call for more knowledge, research, and understanding of wisdom itself in WM.

From the management and leadership perspective, there are intense discourses about wise and unwise leadership (McKenna *et al.*, 2009; Ekmekçi *et al.*, 2014), practical wisdom in management (Bachmann *et al.*, 2018; Jakubik, 2020a), and about wise (*phronetic*) leadership (Nonaka and Takeuchi, 2011; Nonaka *et al.*, 2014). This paper argues that we need wise leaders to act wisely and to solve the global challenges of the world. *Who are the wise leaders?* According to McKenna *et al.* (2009, 178-180) wise leaders:

(i) use reason and careful observations,
(ii) allow for non-rational and subjective elements when making decisions,
(iii) value humane and virtuous outcomes,
(iv) have practical actions oriented towards everyday life, including work, and
(v) are articulate, understand the aesthetic dimension of their work and seek the intrinsic personal and social rewards of contributing to a good life.

Concurrently, Nonaka and Takeuchi (2011) argue that wise leaders can:

(i) judge goodness,
(ii) grasp the essence,
(iii) create shared contexts,
(iv) communicate the essence,
(v) exercise political power, and
(vi) foster practical wisdom in others.

Ekmekçi *et al.* (2014, 1202) argue that a wise leader should:

(i) be capable of cognitive complexity and of coping with uncertainty,
(ii) be a deep thinker and rational, has a capacity to understand and question a situation,
(iii) reveal creativity and logic in non-rational process, be capable of creative thinking and move relying on instinct in their judgments,
(iv) have a long-term vision, virtue, and must commit long-term prosperity for humanity, and
(v) have the capacity to reach people by the use of words, impact and acting are also express all these.

Although to have wise leaders is important, we need intelligent workers and wisdom workers to act wisely when facing wicked problems. *Who are the intelligent workers?* Jakubik (2020b, 67) argues that “human intellect is more than knowledge, the intellectual worker is more than a knowledge worker and intellectual work is more than knowledge work”. Jakubik presents the similarities and differences between knowledge work and intellectual work, as well as between knowledge workers and intellectual workers (Jakubik, 2020b, Table 1, 68-69). Intelligence is important as it connects knowledge with wisdom. *Who are the wisdom workers?* According to Pink (2006) a wisdom worker is a creative individual who combines cognitive and emotional skills, is a talented communicator and has the ability to engage others. Stebbins (2017) argues: “The knowledge worker seeks to accumulate more qualifications from the outside. The wisdom worker has a more balanced inward/outward perspective, understanding her (*sic.*) obligations to herself (sic.) and others" (Stebbins, 2017, 4). The author of this paper believes that a wisdom worker is guided by wisdom, human values, morals and virtues in his or her actions.

There are intense discourses about the characteristics of knowledge economy, wisdom economy (Dobson, 2010; Stebbins, 2017) and WM (Müürsepp, 2013b; Banerjee, 2014; Ekmekçi *et al.*, 2014; Nonaka *et al.*, 2014; Müürsepp, 2021). It is essential to understand what WM is when we are moving from the knowledge economy to wisdom economy. Dobson (2010) and Müürsepp (2013b) discuss the characteristics of the knowledge and the wisdom economy. They argue that, the *knowledge economy*:

- focuses on increasing skills and knowledge,
- is innovative,
- wants more,
- demands qualifications,
- is competitive,
- has the goal of hoarding knowledge, and
- it is grasping and selfish.

On the other hand, the *wisdom economy*:

- is ethical,
- considers social values,
- value judgements are attached to knowledge,
- is reflective,
- wants innovations with purpose and considers their consequences,
- understands *'enough'*,
- demands attitude and aptitude,
- is collaborative,
- reinforces sharing of knowledge,
- values community work and relationship-based actions that build self-esteem and skills,
- is gracious, and
- it is socially responsible.

Similarly, Stebbins (2017) argues: "We are moving from the information and knowledge ages into the emerging age of wisdom. This age will lead us to a Wisdom based economy" (Stebbins, 2017, 3-4). Furthermore, he points out that the *wisdom economy*:

- will be more focused on bringing forward the unique gifts of each person,
- understands the concept of '*enough*',
- asks who profits should we gain the world but lose our humanity,
- understands that a person isn't defined by her or his possessions,
- defines prosperity as a state of sufficiency,
- insists on qualities,
- will recruit for attributes as well as for aptitude,
- is both technological and human,
- uses technology as a tool and more interested in how IT is deployed, understanding the need for humanity, or the highest good of all concerned,
- is collaborative, looks to 'we',
- is reflective.

What is WM? Wisdom management is a newly emerging concept. "The concept of wisdom management seems to be gaining momentum" (Müürsepp, 2021, 21). However, it is not even clear what wisdom is. According to Müürsepp, "wisdom helps to connect knowledge with action. However, wisdom is not just know-how. The concept contains more and covers the emotional part of human life, including morality and creativity" (Müürsepp, 2021, 26). Wisdom is definitely a human quality that develops throughout the life. To better understand wisdom as a concept, Sternberg and Karami (2021) offer a 6P framework with the purpose of unifying different aspects of wisdom. Their integrative framework has the following elements: "(a) Purpose of wisdom, (b) environmental/situational Press that produce wisdom, (c) nature of Problems requiring wisdom, (d) cognitive, metacognitive, affective, and conative (motivational) aspects of Persons who are wise, (e) psychological Processes underlying wisdom, and (f) Products of wisdom" (*Ibid.*, 4). They conclude that "wisdom always has been important but today is perhaps at least as important as, or more important than ever because problems that once were local have become global" (Sternberg and Karami, 2021, 15).

Management is a practice-oriented concept. Practical wisdom (*phronesis*) is typical for management practices (Bachmann *et al.*, 2018; Jakubik, 2020a). Wisdom management is a human practice; when in everyday actions knowledge is applied wisely, with ethical and moral considerations, to different problems and in different contexts. McKenna and Rooney (2005), who define wisdom as the highest form of knowing, discuss the role of wisdom in management. They claim that "*wisdom has to be seen* not as a form of knowledge, because doing so renders wisdom (knowledge) as a set of reified notions that are only loosely connected to conduct or practice, rather than *as a set of attributes and skills that allow people to use knowledge effectively and humanely, that is, as a way of being*" (*Ibid.*, 18, emphases added). They, however, express their concerns and they "caution that just as knowledge management is limited by a lack of theoretical completeness in relation to knowledge, it is likely to be the case that wisdom management will be weakened by a lack of knowledge about wisdom. It would be unfortunate and an opportunity lost if an emerging enthusiasm for wisdom in management is met by this fate" (*Ibid.*, 24). The author of this paper shares McKenna and Rooney's (2005) concerns and urges for more research about wisdom, and about the relationship of KM and WM.

4. Discussion and Conclusions

The findings show that there are several disciplines contributed to KM and WM. Authors (e.g., Dalkir, 2011, 8-9; Jashapara, 2011, 10-14; Massingham, 2020, 34-86) argue that KM has a multidisciplinary character. Therefore, it is difficult to provide one single definition of KM. Consequently, KM could be defined from different perspectives. Jashapara (2011, 13) provides examples of KM definitions from integration of information systems (IS) and HRM, HRM process, IS, and from strategy perspectives. Overview of the evolution of KM (*cf.,* section 3.1.) demonstrated that its current stage focuses on human practices, relationships, networking, and learning. Because of this shift from an IS to a HRM practices perspective, this paper adopts that KM is "any process or practice of creating, acquiring, capturing, sharing and using knowledge, wherever it resides, to enhance learning and performance in organization" (Swan *et al.*, 1999, 669).

Similar to KM, WM is multidisciplinary too. The findings (*cf.,* sections 3.1. and 3.2.) reveal that concepts and theories from different disciplines contribute to KM and WM. As a synthesis of these findings, Fig. 2 presents the main disciplines, concepts, and theories that influenced KM and WM. The figure shows the similarities and differences of KM and WM.

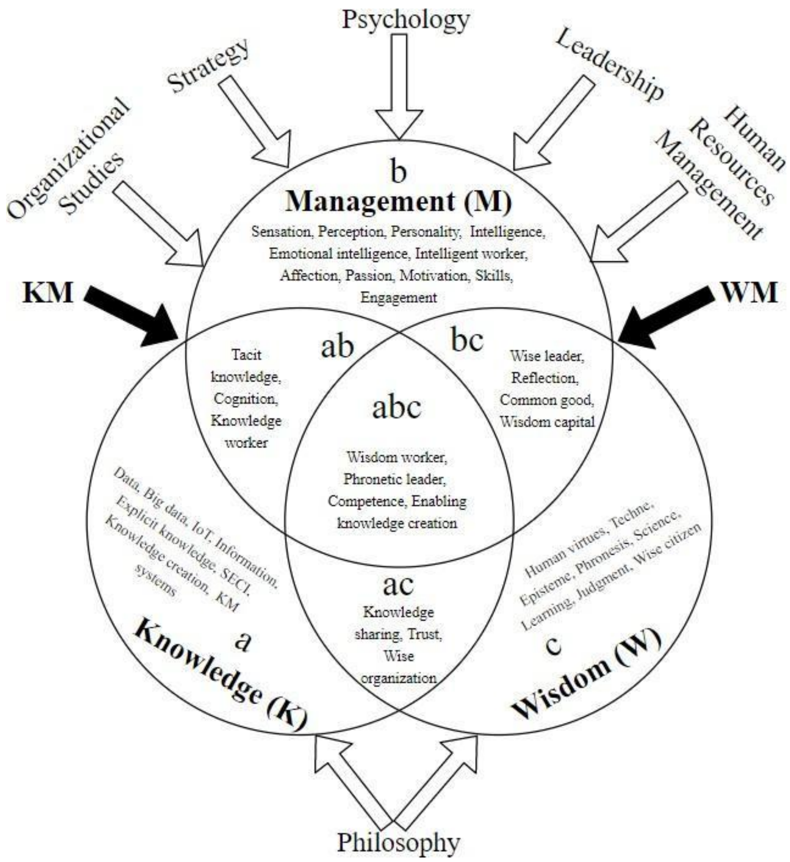


Fig. 2. Comparing Knowledge Management and Wisdom Management

This paper addressed the research problem whether the present phase of KM would be able to address the global problems of the world. The research question was: *Will WM complement or replace KM?* Next, the author of this paper reflects on how the research purpose and objectives were met and she provides a brief conclusion.

The main purpose and objectives of this paper were:

(i) to compare concepts and theories of KM and WM, in order to find out their similarities and differences - This purpose was achieved by: exploring a selected range of literature; finding out KM development phases, its current focus; presenting the discourses about WM; identifying characteristics of key concepts (e.g., wise leadership, intelligent worker, wisdom worker, knowledge economy,

wisdom economy, WM); and by synthesizing the findings in a framework (i.e., Fig. 2).
(ii) to demonstrate the shifts in focus in KM development phases – The objective was achieved by presenting different categorizations (e.g., historical, theoretical, research, inter-organizational) of KM development.
(iii) to recognize the characteristics of the current phase of KM – Findings in section 3.1. showed that the current phase of KM is human practices, relationships, networking, and learning oriented.
(iv) to understand the need for wisdom alongside with knowledge in the next stage of KM to effectively face global challenges ahead – Section 3.2. met this objective by focusing on a necessary shift from knowledge to wisdom economy.

To conclude, the following main points could be added:
(i) The solutions of the global problems of the world (*cf.,* Barnett, 2015; Lufkin, 2017; Maxwell, 2021) would need not only knowledge but wisdom as well. There is an urgent demand for wise leadership, wise workers, wise organizations, and first of all, for wise actions,
(ii) In the current phase of KM research and practice, the focus is not anymore on computer sciences, IT, IS but rather on human practices, relationships, networking, and learning (*cf.,* Jakubik, 2007; Koenig and Srikantaiah, 2007; Jakubik, 2011; Serenko, 2013; North and Kumta, 2014; Agostini *et al.,* 2020; Serenko, 2021),
(iii) Since the mid-1980s wisdom research has intensified in psychology (*cf.,* Karami and Parra-Martinez, 2021; Sternberg and Karami; 2021), education, philosophy (*cf.,* Maxwell, 2021; Müürsepp, 2021), organizational studies, strategy, management (*cf.,* Bachmann *et al.*, 2018; Jakubik, 2020a), leadership, and in HRM (cf., McKenna *et al.*, 2009),
(iv) Need for intelligent workers (*cf.,* Jakubik, 2020b), wisdom workers (*cf.,* Pink, 2006; Stebbins, 2017), and wise leaders (*cf.,* McKenna *et al.*, 2009; Nonaka and Takeuchi, 2011) has increased,
(v) Wisdom is a set of attributes and skills that allow people to use knowledge effectively and humanely, that is, as a way of being (*cf.,* McKenna and Rooney, 2005),
(vi) We are moving from the knowledge economy to the wisdom economy (*cf.,* Dobson, 2010; Müürsepp, 2013b; Stebbins, 2017),
(vii) The multidisciplinary character of KM and WM (*cf.,* Fig. 2) shows that theories and concepts from wide range of disciplines well support the evolutionary character of KM and WM,
(viii) The next phase of KM, should move toward a better understanding of wisdom and WM (cf., Ekmekçi et al., 2014; Jakubik, and Müürsepp, 2021).

This exploratory research is based on a limited number of sources that were selected by the author. Important sources might be unintentionally ignored. Therefore, as future research a more systematic literature review of the current stage of KM and WM could be conducted. Empirical research, about how the current KM practices related to wisdom and practical wisdom in different contexts, would enrich the findings of this paper. Furthermore, there would be a need for more debate about this topic.

Acknowledgements
The author thanks the Ronin Institute for the scholarship that made possible for her to attend the 17th International Conference on Knowledge Management, ICKM 2022, *"Knowledge, Uncertainty and Risks: From individual to global scale"*, organized by the International Council on Knowledge Management at the University of Applied Sciences Potsdam, Germany, June 22-24, 2022.

References

Agostini, L., Nosella, A., Sarala, R., Spender, J.-.-C. and Wegner, D. (2020) "Tracing the evolution of the literature on knowledge management in inter-organizational contexts: a bibliometric analysis", *Journal of Knowledge Management*, 24(2): 463-490, https://doi.org/10.1108/JKM-07-2019-0382.

Bachmann, C., Habisch, A. and Dierksmeier, C. (2018) "Practical Wisdom: Management's No Longer Forgotten Virtue", *J Bus Ethics*, 153: 147-165.

Banerjee, A. (2014) "Knowledge and Wisdom Management", Google Scholar, available at: http://www.delhibusinessreview.org/v_2n2/dbrv2n2p.pdf (accessed 25 Sep 2021).

Barnett, R. (2015) "The coming of the ecological university", In Barnett, R. (2015) *Thinking and Rethinking of University, The selected works of Ronald Barnett*, World Library of Educationalists, London: Routledge Taylor & Francis Group, Chapter 5, pp. 74-90.

Dalkir, K. (2011) *Knowledge Management in Theory and Practice*, Cambridge, Massachusetts: The MIT Press, 2nd edition.

Dobson, J. (2010) "From a Knowledge Economy to a Wisdom Economy", available at: https://www.thersa.org/discover/publications-and-articles/rsa-comment/2010/03/from-a-knowledge-economy-to-a-wisdom-economy (accessed 23 Sep 2021).

Ekmekçi, A.K., Teraman, S.B.S. and Acar, P. (2014) "Wisdom and management: a conceptual study on wisdom management", 10th International Strategic Management, *Elsevier, Procedia – Social and Behavioral Sciences*, 150: 1199-1204, available at: https://www.sciencedirect.com/science/article/pii/S1877042814051842 (accessed 27 Sep 2021).

Jakubik M (2007). "Exploring the knowledge landscape: four emerging views of knowledge", *Journal of Knowledge Management*, 11(6): 6-19, https://doi.org/10.1108/13673270710762675.

Jakubik M (2011) "Becoming to know. Shifting the knowledge creation paradigm", *Journal of Knowledge Management*, 15(3): 374-402, https://doi.org/10.1108/13673271111137394.

Jakubik M. (2020a), "How can practical wisdom manifest itself in five fundamental management and leadership practices", *VILAKSHAN-XIMB Journal of Management, Emerald Insight*, 18(1): 3-25, https://doi.org/10.1108/XJM-08-2020-0078.

Jakubik M. (2020b) "Rise of intellectual workers and intellectual work", *International Journal of Management, Knowledge and Learning*, 9(1): 59-73.

Jakubik, M. and Müürsepp, P. (2021) "What's Next After Knowledge Management? Will Wisdom Management Replace Knowledge Management?", ACIEK Conference, *Innovation, Management, and Governance for Sustainable Growth*, Sorbonne Business School, Paris, France, 14-16 June, 2021, doi: 10.13140/RG.2.2.35764.99201.

Jashapara, A. (2011) *Knowledge Management: An Integrated Approach*, Harlow: Pearson Education Limited, 2nd edition.

Karami, S., Ghahremani, M., Parra-Martinez, F.A. and Gentry, M. (2020), "A Polyhedron Model of Wisdom: A Systematic Review of the Wisdom Studies in Psychology, Management and Leadership, and Education", *Roeper Review*, 42(4): 241-257, https://doi.org/10.1080/02783193.2020.1815263.

Karami, S. and Parra-Martinez, F.A. (2021) "Foolishness of COVID-19: Applying the Polyhedron Model of Wisdom to Understand Behaviors in a Time of Crisis", *Roeper Review*, 43(1): 42-52, https://doi.org/10.1080/02783193.2020.1840467.

Koenig, M.E.D. and Srikantaiah, T.K. (2007) "Three Stages of Knowledge management", Chapter 1, pp. 3-8 in M.E.D. Koenig and T.K. Srikantaiah (Eds) *Knowledge Management Lessons Learned: What works and what doesn't*, Medford: Information Today.

Lufkin, B. (2017) "50 Grand Challenges for the 21st Century". https://www.bbc.com/future/article/20170331-50-grand-challenges-for-the-21st-century (Accessed 20 Sep 2021).

Massingham, P. (2020) *Knowledge management: Theory in Practice*, London: SAGE Publications Ltd.

Maxwell, N. (2021) *The World Crisis – And What to Do About It: A Revolution for Thought and* Action, London: World Scientific, Preface.

McKenna, B. and Rooney, D. (2005) "Wisdom Management: Tension between Theory and Practice in Practice", available at: https://www.researchgate.net/publication/37618441_Wisdom_Management_Tensions_Between_Theory_And_Practice_In_Practice (accessed 28 Sep 2021).

McKenna, B., Rooney, D. and Boal, K.B. (2009) "Wisdom principles as a meta-theoretical basis for evaluating leadership", *The Leadership Quarterly*, 20: 177–190, available at: http://kimboal.ba.ttu.edu/Selected%20writings/wisdom%20principles.pdf (accessed 28 Sep 2021).

Müürsepp, P. (2013a) "The aim of science – Knowledge or wisdom", *Problemos: Filosofijos leidinys*, 84: 72-84.

Müürsepp, P. (2013b) "From Knowledge to Wisdom Economy", Proceedings of The Ninth International Conference on Knowledge-Based Economy & Global Management, 7-8, Nov 3, 2013, Tainan City: Southern Taiwan University of Science and Technology, pp. 41.49.

Müürsepp, P. (2021) "Making Sense of Wisdom Management", *International Journal for Applied Information Management*, 1(2): 63–69, doi: 10.47738/ijaim.v1i2.8.

Nonaka, I. and Takeuchi, H. (2011) "The Big Idea: The Wise Leader", *Harvard Business Review*, May 2011, available at: https://hbr.org/2011/05/the-big-idea-the-wise-leader (accessed 27 Sep 2021).

Nonaka, I., Chia, R., Holt, R. and Peltokorpi, V. (2014) "Wisdom, Management and Organization", *Management Learning*, 45(4): 365-376.

North, K. and Kumta, G. (2014) *Knowledge Management: Value Creation Through Organizational Learning*, Switzerland: Springer International Publishing.

Pink, D.H. (2006) *A Whole New Mind: Why Right-Brainers Will Rule the Future*, New York: Times Bestseller.

Rooney, D, McKenna, B. and Liesch, P. (2010) *Wisdom and Management in the Knowledge Economy*, New York: Routledge, doi: 10.4324/9780203852798.

Serenko, A. (2013) “Meta-analysis of scientometric research of knowledge management: discovering the identity of the discipline”, *Journal of Knowledge Management*, 17(5): 773–812, https://doi.org/10.1108/JKM-05-2013-0166.

Serenko, A. (2021) “A structured literature review of scientometric research of the knowledge management discipline: a 2021 update”, *Journal of Knowledge Management*, Vol. ahead-of-print, No. ahead-of-print, doi: 10.1108/JKM-09-2020-0730.

Solé, J.B. (2017) “Wisdom Management, ECKM 2007-20017 review”, Proceedings of the 18th European Conference on Knowledge Management (ECKM), 7-8 Sep, 2017, Barcelona, Spain, Volume 1: 55-61, available at: https://books.google.fi/books?hl=en&lr=&id=9k88DwAAQBAJ&oi=fnd&pg=PA55&dq=%22wisdom+management%22+source:%22Knowledge+Management%22&ots=slzxV9WtFB&sig=4lXUNA1WaRhG6vEhXz4A4yoFK3M&redir_esc=y#v=onepage&q&f=false (accessed 27 Sep 2021).

Stebbins, G. (2017) “Wisdom Economy”, White Paper, available at: https://www.slideshare.net/GregoryStebbinsEdD/wp-wisdomeconomy-160607 (accessed 25 Sept 2021).

Sternberg, R. and Karami, S. (2021) “What Is Wisdom? A Unified 6P Framework”, *Review of General Psychology*, 25(7): 1–18, doi: 10.1177/1089268020985509.

Swan, J., Scarborough, H. and Preston, J. (1999) “Knowledge management – The next fad to forget people?”, Copenhagen Business School, Copenhagen, Denmark: *Proceedings of the 7th European Conference on Information Systems*, 668-678, available at: https://www.researchgate.net/publication/221408661_Knowledge_Management_-_The_Next_Fad_to_Forget_People, (accessed Sep 25 2021)

Knowledge externalization in the process of building project knowledge management system – a case study of a public organisation

Kamila Brodzińska

PhD Candidate, Doctoral School of Social Science Jagiellonian University in Kraków, Poland

Goal

In project-based organisations, knowledge management is becoming a cornerstone of their operation. Of particular importance in acquiring valuable and useful for employees, as well as for the whole organisation, is one of the processes of its conversion – externalization. It is important that during the implementation of subsequent projects, valuable resources such as knowledge, skills and experience acquired by employees are not lost, but that employees externalize them. The purpose of this article is to seek an answer to the research problem posed, concerning the course and methods of knowledge externalization in project activities and ways of knowledge externalization in the process of building project knowledge management systems in public organisations.

Methodology

The chosen research strategy is a case study. Mixed research (quantitative, qualitative) was conducted among Kraków Museum employees involved in project teams.

Results

Findings show that building a project knowledge management system fosters the externalization of knowledge through, among other things, the individual development of staff, the initiatives they take and the perceived benefits obtained by institutions through change.

Practical implication

Externalization of knowledge makes it possible to unify existing management systems in the institution and thus increase efficiency and promote organisational learning.

Value

The value of the study is the fact of gaining knowledge about the process of building knowledge management systems, through the use of case study, which is gaining importance in the discipline of management and quality.

Key words

project knowledge management, knowledge externalization, tacit knowledge, explicit knowledge, public sector

INTRODUCTION

For almost 30 years, the public sector in Poland has been drawing extensively on management methods used in the business sector. The concept of New Public Management (NPM), which assumes a move away from bureaucratic administration towards more flexible solutions that increase the efficiency of the implementation of public tasks and services, has contributed to their growing popularity. By adopting a strategic orientation, NPM increases the emphasis on standardization, efficiency and competitiveness of tasks *(McLaughlin & Osborne, 2002)*. However, NPM has met with widespread criticism, mainly concerning the negative effects of the marketisation of the public sector. For example, Guy Peters (1999) pointed out that in the long run, attempts to privatize many segments of the public sector may reduce the efficiency of public administration. In response to growing criticism of NPM, the concept of public governance emerged, on the basis of which the idea of good governance was developed. It assumes increasing the capacity to manage the public sector by improving management systems, transparency, development and access to information *(Szumowski, 2017)*. It is based on the idea of increasing the management capacity of the public sector through the improvement of management systems, transparency, development and access to information, thus complementing hierarchically exercised power with participatory and networked solutions. One of the changes in public management practice that resulted from the implementation of the above-mentioned concepts was the attention paid to knowledge management. Knowledge management makes it possible to protect an organisation from the potential loss of acquired knowledge, to increase productivity and thus to react more quickly to new challenges and to prevent potential mistakes *(Durst & Zięba, 2018)*. The usefulness and completeness of knowledge management systems depends on many processes, against the background of which knowledge externalization deserves special attention. Undisturbed transformation of tacit knowledge into accessible knowledge triggers motivation and commitment, leads to knowledge sharing *(Bučková, 2015)*. For these activities to be undertaken in a systematic manner, tools and conditions must be provided to enable this process to take place. A particular knowledge management situation arises when an organisation carries out an increasing number of projects. Projectization, involving the institutionalization of projects in all spheres of social life *(Jacobsson & Jałocha, 2018)*, is a phenomenon that is also occurring in the public sector. Projectized public organisations, are therefore looking for ways to effectively manage the knowledge they acquire through projects. At the same time, there is a research gap – a lackof in-depth studies on the ways and processes of knowledge externalization in building project knowledge management models. This article is an attempt to fill this gap.

The research problem concerns the identification of ways of knowledge externalization in the process of building project knowledge management systems in public organisations.

The chosen research strategy is a case study. The paper is based on mixed research consisting of observation, interviews, surveys and document analysis. It consists of a theoretical part, in which the most important assumptions concerning knowledge management were presented, including project knowledge management, the knowledge spiral model, with particular emphasis on one of its cycles – externalization of knowledge. It also focuses on the professionalization of knowledge management through the creation of knowledge management systems. The empirical part includes a description of the researched institution, results of the conducted research and a summary containing answers to the research questions.

KNOWLEDGE MANAGEMENT

Knowledge as a resource

Human knowledge is a key element that enables the uninterrupted functioning of organisations. Its use allows for increased quality, efficiency and competitiveness, which provides unique value to the organisation *(Rao & Nayak, 2017)*. Knowledge management counteracts the occurrence of factors considered by organisations to be limiting, therefore it has become a separate management concept, the subject of which is the use of resources possessed by the organisation. These resources are considered to be properly functioning solutions for managing information systems, organisational changes and human resources *(Davenport & Prusak, 2004)*.

Recognition of the positive impact of knowledge management in private organisations, also prompts public sector institutions to implement solutions in this field. The need to redefine and improve the management methods of public organisations is related to the different perception of their role in society than before. Higher and higher demands are placed on institutions in connection with their activities, which makes it necessary to search for such solutions that contribute to the improvement of the quality of services while maintaining economic order. The changes taking place in the public sector mean that knowledge is also treated here as a strategic resource for organisations *(Callahan, 2018)*.

There are many models that describe the knowledge management process. These are mainly used in knowledge-based organisations. For example, the model proposed by Ashok Jashapara *(Jashapara, 2011)* model consists of five phases:

- Knowledge discovery
- Knowledge generation
- Dissemination of knowledge
- Exploitation of knowledge
- Retention of necessary knowledge

Another model is the Japanese model of knowledge creation proposed earlier – in 1994 – by Hirotaka Takeuchi and Ikujiro Nonaka, known in the literature as the SECI model, whose name is an acronym derived from the key concepts in the model: socialization, externalization, combination and internalization *(Nonaka, Umemoto, & Senoo, 1998)*. In Poland, it is better known as the knowledge spiral model. The basic observation that contributed to its creation is the repetitiveness of the four cycles of knowledge conversion:

Table 1. Knowledge conversion cycles

Cycles name	**Explanation**
Socialization	transition of tacit (hidden, internal) knowledge to tacit knowledge through dialogue, observation or joint activities.
Externalization	turning tacit knowledge into formal (explicit, accessible) knowledge by articulating the knowledge and giving it a concrete, permanent form, which is the basis for creating new knowledge. This is the sharing of tacit knowledge[1].
Combination	the transformation of formal knowledge into another form of formal knowledge, by combining existing forms of knowledge in a given space, disseminating and systematising them.
Internalization	the conversion of formal knowledge into tacit knowledge, through learning by doing, when employees acquire formalised knowledge through practices, initiatives, training, recognising the knowledge as their own.

Source: Ikujiro Nonaka, Katsuhiro Umemoto and Dai Senoo, From Information Processing to Knowledge Creation: A Paradigm Shift in Business Management, Technology In Society, Vol. 18. No. 2, pp. 205-208

Despite the development of knowledge management and the popularization of other knowledge models, the Japanese model still plays a key role in organisations that undertake activities aimed at acquiring knowledge in a structured manner. It is worth mentioning that it is often the spiral model that is the canvas for the emergence of other models and concepts *(Mittal & Kumar, 2019)*.

In the model referred to above, two levels of knowledge are distinguished: explicit – accessible knowledge, also referred to as formal knowledge, and tacit knowledge – otherwise known as tacit or internal knowledge. Both forms affect the creation and dissemination of knowledge in organisations and are used in the most important knowledge management tasks, as employees have both explicit and tacit knowledge. Explicit knowledge is formal and generally available information, presented orally or in writing – documentation, instructions, diagrams, maps, symbols, textbooks and videos. Its feature is the ease of transfer and registration as well as storage and dissemination *(Toyama & Nonaka, 2015)*. Tacit knowledge, on the other hand, is information and non-verbal experiences, with a high degree of individuality. It has an intuitive dimension and is acquired during the employees' own repeated and long-term experiences. The problem is that it is not verified, because it is not recorded or communicated in any way, it is difficult to formalize *(Seidler-de Alwis & Hartmann, 2008)*.

[1] Following the definitions of W.R.King and G. von Krogh, knowledge transfer is an intentional and unidirectional activity, which consists of communicating knowledge in order to apply it in a concrete manner. Knowledge sharing, on the other hand, is treated as a multidirectional activity that is not only knowledge exchange, but also a change in the level of knowledge of both the recipient and the knowledge sharing entity (King, 2006). Such an understanding is adopted in this article.

Externalization of knowledge

All knowledge conversion cycles are relevant to organisations, but for many years knowledge management theorists have identified the knowledge externalization cycle as a particularly important research area *(O'Meara & Kelliher, 2017)*, *(Tee & Lee, 2013)*, a cornerstone of knowledge creation. Often knowledge conversion at this stage can be complicated due to the problem in articulating tacit knowledge through speech or writing. The process of externalization is therefore considered to be the most creative cycle of conversion, because the tacit knowledge acquired during it, transforming into explicit knowledge, may take many unique forms: metaphors, analogies, concepts, hypotheses or anecdotes *(Naicker, Govender, & Naido, 2014)*.

In the externalization cycle it is important to focus on the cognitive dimension, which enables the transfer of the created knowledge between communities, groups or teams. As a result of a projectized or spontaneous externalization process, people with different backgrounds, social roles or experiences externalize their tacit knowledge. However, in order for this process to be effective and not lead to cognitive errors, it is important to take into account the social, cultural, situational and even historical contexts. The contextualization of imparted knowledge influences its categorization as useful, reproducible under given conditions and taking into account the predispositions of the person who acquired it *(Yasuoka, 2019)*.

There are many ways of converting tacit knowledge into explicit knowledge. They consist in articulating tacit knowledge in a space accessible and assimilable form for third parties. The most obvious way of such externalization is conversation – formal and informal. The process of externalization can therefore be influenced by organising regular meetings *(Brodzińska, Jałocha, & Szostak, 2019)*.

A way to support externalization of knowledge is to initiate joint task performance, which requires constant contact between employees, or to organise a group search for ideas (brainstorming), when during discussions and collective reflections concepts are presented, which then take the form of explicit knowledge. The externalization of knowledge also takes place during trainings, lectures and courses, because often the creation of external knowledge is also a form of externalization. Other forms of externalization include making notes for others involved in the activity, preparing mind maps, photos or instructional videos *(Yasuoka, 2019)*.

Madge *(2012)* points out that when creating knowledge externalization processes in an organisation, care should be taken about the circumstances and conditions in which they are to take place. The person responsible for creating it should consider whether it would be more effective to create a formal or informal setting for a one-off session, as this may determine how much individuals choose to share their knowledge. It is important to create a permanent space where employees can meet informally e.g. during breaks. Michel O'Meara and Felicity Kelliher *(2017)* emphasize the role of the learner's attitude towards acquiring tacit knowledge from employees who hold the knowledge. This is influenced by employee characteristics such as a desire for self-development, proactivity, often boldness or the ability to ask for knowledge support. An important role is also played by the quality of the acquired knowledge, which should be as high as possible. This depends on the skills, self-awareness and willingness of the employee carrying the tacit knowledge to present it.

Knowledge management systems

Recognizing the role of knowledge in the management of organisations necessitates continuous attempts to professionalize knowledge management processes at every stage. Raising the standards of knowledge management is particularly important in two dimensions: human and systemic *(Brzostek & Michnia, 2015)*.

The first one is related to managing human resources in such a way that team members have the broadest possible competence in their field and are ready to share the knowledge they possess. The role that the organisation should play is to guarantee the possibility of acquiring new competences for employees through their continuous education. The organisation should also support non-organisational ways for its employees to acquire knowledge *(Ziółkowski, 2017)*. The systemic aspect is related to the emergence of knowledge management systems that are adapted to the needs and capabilities of the organisation, influence the speed, quality and cost-intensity of activities *(Brzostek & Michnia, 2015)*.

Another element in the professionalization of knowledge management processes is the use and promotion of expert knowledge, which consists in using external advisors in knowledge management projects, who, after analysing the organisation's initial situation, are able to propose the best solutions and practices. It is important that, when conducting a consultancy project, such an advisor has the opportunity to make observations in the organisation and to consult proposals coming from its employees *(Wyrozębski, 2012)*.

The professionalization of knowledge management is also linked to the use of new technological and IT tools. They are particularly important in the case of large or territorially dispersed organisations, because without their existence, employee communication would be definitely hampered. Knowledge-based organisations owe their success to project the flow of information in such a way that the necessary knowledge is transferred quickly and conveniently. It is crucial that during this flow certain information or knowledge is not lost, which employees could have passed on to each other during personal contacts. This is mainly about emotions, attitudes and values. Technological tools for knowledge management used in organisations are characterised by varying degrees of sophistication, which is important because technology cannot become a more important aspect than knowledge itself *(Plebańska, 2016)*.

After gaining some experience with knowledge acquisition, the organisation can take steps to make the knowledge management process take a more formalised form. Very often then, a knowledge management system is implemented that enables all knowledge processes to be carried out in a systematic way *(Dobrowolski & Leśnik, 2016)*. A knowledge management system (KMS) is a set of methods, techniques and tools that are applied within interdependent subsystems such as tacit and explicit knowledge management, knowledge culture and transfer, and organisational solutions. This is done using information technology and a system of metrics *(Brzostek & Michnia, 2015)*.

The extent to which an organisation implements a knowledge management system is determined by factors such as the number of people employed in the organisation, the degree of centralization and decentralization, area of operation, volatility of the environment and financial conditions. Management awareness of the benefits of a KMS is also important.

Project knowledge management

A specific form of organisational knowledge is knowledge derived from projects. Their implementation takes place through successive phases, during which knowledge related to a particular phase is generated. The uniqueness of projects can be a challenge when an organisation takes the action of accumulating knowledge gained during particular processes, in order to use the experience in the future *(Marciszewska, 2016)*. Given the uniqueness of project forms, it is important to consider how to capture and manage knowledge that will be useful in other projects.

In order to guarantee the proper use of project knowledge, it is necessary to take care to collect knowledge at different levels of detail. The type of knowledge depends on the nature of the project – there are activities so similar that certain types of solutions can be applied analogously. However, many activities are undertaken as a reaction to the occurring change, therefore some activities may be reconstructed only on the basis of the analysis of documents or reports (if any are created) *(Sońta-Drączkowska, 2012)*. The accumulated project knowledge gains the value of usefulness, when the best practices are selected from it, which – after a prior generalization – may constitute a base of knowledge or inspiration. When creating knowledge bases it is necessary to remember about the dynamics of processes occurring within projects. A popular way is to collect project knowledge through a document management system – schedules, procedures or memos. They help to reconstruct certain elements occurring during works, allow to know the duration of particular stages or the amount of used resources. Knowledge can also be collected through the use of decision support systems in organisations. They are created from the perspective of the end user, the analysis of the data contained in them allows to consider what results the decision concerning the way of working may have. A form of knowledge management is also increasing the level of project management knowledge among employees. It can take place through courses, training, mentoring *(Wyrozębski, 2014)*. However, it is important that after their completion the acquired information is reflected in the form of training materials or notes.

The timing of the end of the project team's work is very important. One of the key elements of this activity should be to summarise and collect the knowledge resulting from the experience, which the sooner this is done, the less likely the knowledge generated during the work will be dissipated *(Harris, 2016)*.

Due to the limited duration of projects, the dissolution of project teams and the transition of employees to subsequent teams, it is advisable to have a special cell in the organisation that will support and coordinate the process of knowledge management of completed projects. Its task is to optimise project management by implementing recommended methods, techniques and solutions which have contributed to the success of previously undertaken projects *(Kerzner, 2019)*. The separation of such a unit in the organisation is conditioned by the desire to improve efficiency through the use of knowledge retained after the project in the form of post-project knowledge (lessons learned). This is possible thanks to the use of knowledge from other project activities, analysis of success and failure factors and the flow of information about the project at all levels of the organisation *(Jałocha, 2014)*.

To achieve this goal, the Project Management Office (PMO) carries out a number of smaller tasks. According to P. Wyrozębski *(2019)* these tasks include: reporting to senior management related to project status, creating and implementing project management methodologies, disseminating project management from across the organisation, participating in the strategic planning process, identifying, grouping and assigning rank and importance to newly started projects. The PMO is also tasked with helping to allocate resources – preparing the necessary equipment, software and budget *(Jałocha, 2014)*.

The unquestionable benefit of the existence of the PMO can be seen, for example, during: the appointment of a person who holds this function for the first time to a manager's place, in the case of a personnel change in the project team or the need to expand the project team. The introduction of such persons to the team must take place in a quick and effective way, which guarantees providing them with systematically collected knowledge about the project *(Project Management Body of Knowledge, 2013)*.

EXTERNALIZATION OF KNOWLEDGE IN THE KRAKOW MUSEUM – A CASE STUDY

Test methods

The research strategy chosen was a single case study *(K.Yin, 2015)*, the subject of which was the knowledge externalization processes of a selected public organisation. The study was carried out at the Museum of Kraków, located in Poland, in the capital of the Lesser Poland Voivodeship – Kraków. The study focused on issues related to project knowledge management. The case selection was intentional – the museum carries out over 90 projects a year. With such a high involvement in project work, there is a need to improve existing ways of managing project knowledge. The study posed the following research problems:
RQ 1: How is the knowledge externalization process carried out in the project activities of the organisation under study?
RQ 2: What methods are used to improve the process of project knowledge externalization in public organisations?

This research is part of a broader research project, which was carried out over a period of 13 months, within which the processes of project knowledge management at the Museum were examined. The research methods used included a literature review, observations conducted in all departments of the Museum, analysis of found data – the organisation's strategy and documents regulating the activities of the Museum, analysis of internet sources, content on social networking sites and in the media. Documentation of selected projects carried out by the institution was also analysed. Other research methods used included semi-structured and in- depth interviews with project staff and a questionnaire. For the purposes of this study, in order to answer the research questions posed, two sources of data were analysed in detail: in-depth interviews and observation notes. In total, the following were carried out 8 interviews, 23 observations and a questionnaire survey. Information obtained during observations, transcriptions of interviews and notes were documented by keeping a research diary, and then they were systematised in observation and interview protocols created for the purposes of the study.The survey research was conducted through an online questionnaire, the results of which were also collated in the research diary.

Museum of Kraków

The Kraków Museum is one of the most interesting cultural institutions in Malopolska, which in 2019 celebrated 120 years of its foundation. It is known for its valuable and diverse collections, attractive monuments, organisation of unconventional events and creative exhibition activities. The Museum has a very extensive organisational structure consisting of 19 branches, so it is a large, dynamically developing institution.
The Museum's mission and vision illustrate that the institution pays great attention not only to the quality of its services, but also to their professionalism, which is only possible when in the organisation works on eliminating imperfections connected with management processes.

Project management at the Kraków Museum

An important element of management at the Kraków Museum is project management. Exhibition events – both permanent and temporary exhibitions, cyclical events, anniversaries, as well as educational activities – are carried out in the form of projects. The projects include exhibition events – both permanent and temporary exhibitions, cyclical events, anniversaries, as well as educational activities. In accordance with the adopted principles specified in the internal procedures of the Museum, temporary exhibition projects are realised in a 3-year cycle, and educational and popularisation events in a 2-year cycle. Moreover, the Museum carries out investment projects, e.g. connected with the creation of new branches or the renovation of existing ones. Due to the different fields of activity of the Museum and the scale of its programme activities, several dozen projects are carried out at different stages. Annually, the Museum carries out about 90 projects, and due to the potential accession of new places, which become the next branches of the Museum, this tendency is likely to increase. The specific nature of the institution requires the employment of knowledge workers – historians, project managers, exhibition curators. Approximately 30% of the 338 employees of the Museum are involved in projects implemented by the Museum.

The Museum has been carrying out activities to improve its project management system, but the issue of project knowledge management – and therefore project knowledge management – has not been given attention.

Ways of externalizing knowledge during projects

In order to carry out tasks in a project-oriented form, the organisation uses a number of tools that enable the transfer of knowledge about projects in progress.

In the years 2011-2012, the Museum implemented a project management system based on the PMI methodology. Sample project documents were developed, such as a project charter with appendices containing schedules, cost estimates, a promotion plan, a communication plan, and reports. An information flow system was also implemented within the organisation
– between employees and the management and between organisational units involved in project implementation. The next step was to train employees in the basics of project management and the Museum's project procedures. However, the quality of the information contained in the documentation depends on the project management knowledge of the people who completed the documentation, their accuracy and reliability. The information gathered during the survey shows that the quality of individual elements of project documentation varies greatly, mainly due to the multiplicity of documents and time constraints – employees involved in projects often carry out several projects in parallel, and are also busy with other nonproject- related tasks, so they lack the time to systematically complete the documentation.

The organisational structure also includes two organisational units crucial to the project management process. The first is the Exhibitions and Museum Events Production Department, which brings together experienced project managers. The second one is the Management Process Support Section, which performs monitoring functions with respect to the implemented projects. In the context of the subject discussed in this article, it should be emphasised that the Section is a place that collects and transmits information and project materials. Thus, it fulfils selected tasks of the Project Management Office. A form of project knowledge management that takes place within these units is also the evaluation of exhibition projects. Employees can access existing project documentation via the internal intranet. As in the case of the developed project management procedure, the amount and quality of information possessed by the above-mentioned units depends on the willingness of employees to share their experiences and own knowledge resources with others. However, the interviews and questionnaire survey illustrate that many employees are not aware that such activities take place or do not identify these units with the PMO, declaring at the same time that they do not know where they could obtain information on issues related to project management. Sample answer to the question if there is a PMO in the organisation:

It does not exist. It is one of the ideas of the structure to create something like that (...) we have something closer, it is a project office for investments like renovations. There is an administrative and technical team that supervises the work.

Another form of externalizing knowledge is through meetings of both task and project teams. When obtaining information about the regularity and frequency of team meetings it turned out that such meetings are organised, but their regularity. However, their regularity and frequency depend on the project manager's management style. Employees consider this form as one of the most useful, arguing that it is the time when they have a chance not only to learn about the stage of work of other team members, but also to share their own observations. Very often during such meetings, brainstorming or discussions take place, which enables the process of conversion of implicit knowledge into explicit one.

Due to the fact that my project evolved (...) the team was recently constituted and then the programme was developed – in the course of team meetings, and there have been 2 such meetings so far. In the process of brainstorming, discussions, etc. (the programme) came into being, because we are cooperating and, in fact, it is based on the fact that everyone has some experience which they bring and we can draw on that.

A way of externalizing knowledge is also telephone contacts, which are one of the telephone contacts, which are one of the most frequently used forms of communication between employees, because project team members often work in different departments, located in different parts of the city and rarely have the opportunity to meet each other. Employees declare that so far it has not happened that any of the co-workers whom they called asking for support with their own project experience refused to help them:

In 8 years, I have never had anyone refuse me this knowledge when I asked about something that was in their field. If I need some knowledge and I know who has it (...) I write an e-mail, call, go to a meeting and ask for support, help, a book recommendation, etc.

Often the person talking about his/her project experience provides more accurate information than is contained in reports and documents, and also talk about practical aspects of solving a given issue. Project teams highly value this form of knowledge transfer as it saves time during project implementation.

Another way of sharing project knowledge is the use of internal e-mail, through which employees contact each other, requesting documents or asking for help. Employees declare that they are willing to give each other tips and advice, but compared to phone calls, fewer respondents found this form useful, mainly because of the waiting time for information and its accuracy:

If there are issues that concern the whole team or there is information I want to pass on to everyone, I write an email.

Externalization of knowledge also takes place during formal meetings of project team members, usually initiated by project managers or employees of other departments. It is a good opportunity to listen to the voice of other employees, get acquainted with their observations or ideas. During such meetings, problematic issues are solved and thanks to this experience other team members gain knowledge which was unavailable for them. Project managers and people from outside the project teams often not only help employees interpret the knowledge they have revealed, but also initiate its further conversion. Externalization of knowledge takes place through direct conversations as well as notes and reports generated during meetings, which can be seen by other project team members.

A very valuable form of externalization of knowledge in an organisation are informal personal contacts of employees. They take place during breaks, private meetings or chats. Such contacts are made during the implementation of joint undertakings, for example by going out to lunch together. On a broader scale, such opportunities arise during museum celebrations. Such encounters allow for a more intensive exchange of knowledge, which is not classified in any framework or norm.

It is very often the case that when I meet a former colleague who now works in another department of the museum, it turns out in the course of conversation that they had a similar problem in the past and the person says that it was solved in such and such a way (...). Then we as a team can apply this solution at home, taking into account our specific conditions.

Informal personal contacts also allow for the creation of bonds between employees, which favours the process of externalization of knowledge by bridging barriers. This form reduces the feeling of being embarrassed by the lack of knowledge in a given field and employees often prefer to ask for knowledge support in informal situations rather than in a team meeting. Despite the lack of structuring, resulting from the specificity of this contact, it is very effective, and for employees it is an attractive form of knowledge externalization.

There are grassroots meetings organised, and they are probably the best and most creative

– when someone comes up with something, sees something interesting and wants to bring it into the project. Then (...) we meet. They are the best proof that someone is engaged and sometimes, paradoxically, during such casual meetings we talk over tea and something cool comes out.

It is worth noting that employees have different project roles. It is often the case that a person who is a project manager simultaneously performs a substantive in another project. Thus, such a person has different experiences and in each project through contact with colleagues acquires different, valuable knowledge, which also contributes to the externalization of knowledge.

The survey conducted shows that there are forms desired by employees during which the knowledge could be externalized, but they are not realised in the institution in a dimension accessible to all employees. Such forms include trainings and workshops. All participants in the survey agreed that it is important for them that employees share the knowledge acquired during project implementation:

When working on a project, the experience of individual members allows the whole team to operate more efficiently, better estimate the time and budget needed, and avoid mistakes made in other projects and minimise the risk already in the very first stages of a project.

In spite of this, 21 out of 22 employees admitted that in order to perform their tasks in the project they have to search for information on their own, with 15 of them knowing from which source they can obtain knowledge in case of its absence. Employees declare that when they convert tacit knowledge into explicit knowledge, they do it most often during informal meetings with colleagues and during project evaluation. The process of externalization of knowledge at the Kraków Museum is therefore spontaneous. There is no person or organisational unit responsible for its creation and support. Steps are being taken to professionalise the various cells and sections that support the management processes. Some elements of such a system exist, but – as declared by the staff – it needs to be improved and simplified in order to make it more useful.

CONCLUSION

Employees have knowledge gained through their own experience and assimilated knowledge that comes from the environment. In order for knowledge to become available to all employees, it requires other project team members to disclose it in a concrete form. All phases of the knowledge acquisition model described by Ashok Yashapar must therefore occur. In knowledge acquisition, the application of the SECI model plays a key role. During the implementation of projects, all of the described knowledge conversion cycles occur at each of their stages, however, both the realised study and literature analysis show the particular importance of externalizing knowledge. The literature review allows us to conclude that the process of knowledge externalization in public institutions takes place in a spontaneous or organised manner *(Łychmus, 2010)*. It depends on the awareness of the members of the organisation and the existence of units or departments in the organisational structure that would be responsible for creating and supporting the process. Its intensity be subject to the attitudes of employees and the knowledge they have *(Brodzińska, Jałocha, & Szostak, 2019)*.

With reference to the research question (RQ 1: How does the process of knowledge externalization take place in the project activities of the organisation under study?) it can be stated that in the organisation under study, the knowledge externalization process is spontaneous and unplanned. However, knowledge externalization occurs on a regular basis, since – as presented above – there are groups of tasks, during which the externalization process is a natural consequence of teamwork. It is influenced by the possibility of contact – above all personal contact – between employees.

The survey conducted in the studied organisation illustrates that among the following ways of externalizing knowledge during project implementation were used:

Table2 . Ways of externalizing knowledge in the studied institution

Means of externalizing knowledge	**Use in the institution examined**
Staff notes, reports	Regular use
Document management system	Use, during modernisation
Telephone calls	Regular use
Electronic correspondence	Regular use
Formal personal contacts, business talks	Regular use
Informal personal contacts, e.g. during breaks	Regular use
Working in task forces	Depending on the type of project and the decision of the project manager
Project team meetings	Depending on the type of project and the decision of the project manager
Supporting staff initiatives on project activities	Situational use (e.g. during project meetings)
Rewarding knowledge sharing	It is not used
Mentoring	Irregular use (takes place on an informal basis)
Transmission of information via the intranet	Regular use
Training, conferences, workshops	Irregular use

Source: Prepared on the basis of own research

In the institution studied, a very important element of externalization is the variability of the roles of project team members, who, through the necessity of performing different types of duties, acquire new and unique knowledge which they share with their colleagues.

In answering the second research question (RQ 2: What methods are used to improve the process of externalizing project knowledge in public organisations?), these are mainly those methods that influence the building of positive relationships between colleagues, such as meetings, enabling employees to carry out common tasks, and creating a space to exchange experiences *(Szajczyk, 2017)*.

When creating a knowledge externalization process, it is worth paying attention to aspects such as the knowledge life cycle, potential and feasible methods of knowledge extraction, cooperation between actors involved in the process and externalization scenarios. These aspects should be integrated, the omission of any of them may lead to an externalization gap. In order for institutions to apply methods to improve the externalization of knowledge, it is important to carry out an analysis of the stages that occur in the externalization process. In table 3, present your own process proposal.

Table3 . Project stages of knowledge externalization – own proposal

Project stages the externalization process	**Actions taken in the institution under review**	**Good knowledge management practices identified in the literature**
Knowledge life cycle	-Production of documents, reports, notes -Processing of records by the Management Process Support Section -Gathering of available	- Analysing documentation schemes and adapting them to changing conditions -To facilitate information retrieval by introducing a data tagging system
	resources and making them available to interested staff	
Potential and feasible methods of acquiring knowledge	-Formal and informal requests for information -Creating an organisational culture based on trust, knowledge and a desire to improve	-elaborating with other entities that have experience in improving the process
Cooperation between actors involved in the process	-Occasional project and task force meetings	-introduction of regular project and task forces
Externalization scenario	- no such scenario exists in the organisation	-planning the process taking into account organisational conditions -removing externalization barriers resulting from reluctance to share knowledge -introducing a process support system - space for own initiatives, rewarding activity

Source for column 3: Chris Stary, Jeannette Hemmecke, A Framework for the Externalization of Tacit Knowledge Embedding Repertory Grids, Conference: Proceedings of the Fifth European Conference on Organisational Knowledge, Learning, and Capabilities, 2004, Kamila Brodzińska, Beata Jałocha, Agnieszka Szostak, Using Action Research for Improvement of Project, Knowledge Management, Knowledge Management Practices in the Public Sector, IGI Global, 2019, pp.132-139

During the work of project teams when the revealed knowledge is not captured in an appropriate way it will not be externalized but socialized. An improvement activity in this regard is the creation of tools useful for recording and reproducing it. It is also important to raise awareness of the benefits of systematically collecting knowledge which can be reproduced. The improvement of the externalization process is also influenced by the professionalisation of knowledge management processes. It results in enabling employees to raise their competences and stimulating them to share newly acquired, useful knowledge. Another method may be the use of support from experts from outside the organisation, which will make it possible to analyse existing processes, select the most effective activities and implement new ones – not used before, which may prove useful. The work tools used in the institution, including technological tools, also influence the improvement of the externalization process. It is necessary to guarantee all employees access to them and to train them in the use of these tools.

The knowledge management process is integrated into the organisation's management process. Consideration of this by management may be the trigger for a decision to organise this process and implement it during each phase of the organisation's activities. Employees involved in transferring their knowledge holdings to their colleagues often recognise that the transition from tacit to explicit knowledge can support the alignment of the institution's current management systems *(Bučková, 2015)*. This is due, among other things, to the high project maturity of these employees, their experience and the communication style they adopt. However, due to their multiple responsibilities, they are not able to initiate and coordinate the creation of a project knowledge management system. The analysis of the dimensions of knowledge management systems allows us to observe that the direct impact of knowledge externalization on their creation is revealed in the functional and institutional aspects, as they draw directly on the tacit knowledge of employees. Its externalization facilitates the planning and implementation process of knowledge management systems and is helpful in introducing corrections and further improvements *(Brzostek & Michnia, 2015)*. Of the knowledge conversion cycles, it is externalization that is most immersed in the organisational context *(Mahmound, George, Reisel, & Pantsios, 2018)*.

BIBLIOGRAPHY:

Brodzińska, K., Jałocha, B., & Szostak, A. (2019). Using Action Research for Improvement of Project Knowledge Management. In V. Ismyrlis, T. Tarnanidis, & E. Moschidis, Knowledge Management Practices in the Public Sector (pp. 134-135). Hershley: IGI Global.

Brzostek, K., & Michnia, A. (2015). Wdrożenie systemu zarządzania wiedzą w przedsiębiorstwach sektora MSP. Zeszyty naukowe Politechniki Śląskiej: Seria Organizacja i Zarządzanie, pp. 52-53.

Bučková, J. (2015, 12). Knowledge Management in Public Administration Institutions. Procedia Economics and Finance, 34, pp. 390-395. doi:10.1016/S2212-5671(15)01645-7

Callahan, S. (2018). Putting Stories to Work: Discover. In J. Girard, & J. Girard, Knowledge Management Matters.
Words of wisdom form leading practitioners (pp. 65-67). Macon: Sagology.

Davenport, T. H., & Prusak, L. (2004). Working Knowledge: How Organizations Manage What They Know.
ACM: Ubiquity, n.d, pp. 1-2.

Dobrowolski, D., & Leśnik, M. (2016). Zarządzanie wiedzą jako proces. In R. Knosala, Innowacje w zarządzaniu i inżynierii produkcji (p. 95). Opole: Oficyna Wydawnicza Polskiego Towarzystwa Zarządzania Produkcją.

Durst, S., & Zięba, M. (2018). Mapping knowledge risks: towardsa better understanding of knowledge management. Knowledge Management Research & Practice, 17(1), pp. 1-3.

Harris, E. (2016). Post-project reviews. In E. Harris, Strategic Project Risk Appraisal and Management (pp. 111- 112). New York: Routledge.

Jacobsson, M., & Jałocha, B. (2018). A literature review on projectification: Trends, emerging ideas and avenues for future research. The 14th International Research Network on Organizing by Projects (IRNOP) conference. Melbourne.

Jałocha, B. (2014). Zarządzanie portfelem projketów. Koncepcje biznesowe a praktyka organizacji publicznych.
Kraków: Wyd. Uniwersytetu Jagiellońskiego.

Jashapara, A. (2011). Discovering knowledge. In A. Jashapara, Knowledge Management An Integrated Approach (pp. 3-25). New York: Prentice Hall.

K.Yin, R. (2015). Studium przypadku w badanich naukowych. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.

Kerzner, H. (2019). Using the Project Management Maturity Model: Strategic Planning for Project Management.
New Jersey: Willey.

King, W. R. (2006). Knowledge transfer. In D. G. Schwartz, Encyclopedia of Knowledge Management (pp. 538- 544). London: Idea Group Reference.

Łychmus, P. (2010). Treningi kierownicze z perspektywy konwersji wiedzy. In P. Łychmus, Coaching oparty na wiedzy w treningach kierowniczych (pp. 49-59). Warszawa: C.H. Beck.

Madge, O.-L. (2012). Creating a Culture of Learning and Knowledge Sharing in Libraries and Information Service. In H. T. Hou, New Research on Knowledge Management Models and Methods (pp. 247-256). Rijeka: In Tech.

Mahmound, A. B., George, K., Reisel, W. D., & Pantsios, A. L. (2018). Knowledge Management in the Public Sector: The Spilling Effects on Employee Performance Using the SECI Model. Driving Productivity in Uncertain and Challenging Times, At Bristol Business School, University of the West of England, (pp. 8-9).

Marciszewska, A. (2016). Projekty w zarządzaniu instytucjami publicznymi. Wrocław: Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu.

McLaughlin, K., & Osborne, S. P. (2002). The New Public Management in context. In K. McLaughlin, S. Osborne
, & E. Ferlie, New Public Management. Current trends and future prospects (pp. 21-22). New York: Routledge.

Mittal, S., & Kumar, V. (2019). Study of knowledge management models and their relevance in organisations.
International Journal of Knowledge Management Studies, 10(3), p. 328.

Naicker, K., Govender, K. K., & Naido, K. (2014). Conteptualizing knowledge creation, conversion and transfer.
Trends and Development in Management Studies(3), pp. 27-28.

Nonaka, I., Umemoto, K., & Senoo, D. (1998). From Information Processing to Knowledge Creation: A Paradigm Shift in Business Management. Technology In Society, 18(2), pp. 203-208.

O'Meara, M., & Kelliher, F. (2017). Framing the Externalisation Phase of Knowledge Creation in an Organisational Setting. European Conference on Knowledge Management; (p. 770). Kidmore End: European Accounting and Management Review.

Plebańska, M. (2016). Technologiczne narzędzia zarządzania wiedzą a innowacje w przedsiębiorstwach sektora MSP. Zeszyty Naukowe Uczelni Vistula, pp. 105-112.

Project Management Body of Knowledge. (2013). Project Mangement Institute.

Rao, S. S., & Nayak, A. (2017). Enterprise ontology model for tacit knowledge externalization in socio-technical enterprises. Interdisciplinary Journal of Information, Knowledge and Management, 12, p. 100.

Seidler-de Alwis, R., & Hartmann, E. (2008). The use of tacit knowledge within innovative companies: knowledge management in innovative enterprises. Journal of Knowledge Management, 12(1), pp. 133-134.

Sońta-Drączkowska, E. (2012). Sprawne zarządzanie projektami: korzyści dla całej organizacji. In E. Sońta- Drączkowska, Zarządzanie wieloma projektami (pp. 25-26). Warszawa: Polskie Wydawnictwo Ekonomiczne.

Szajczyk, M. (2017). Zarządzanie zespołem projektowym- wyzwania menedżera projektu. Modern Management Review, 22(24), pp. 109-116.

Szumowski, W. (2017). Próba identyfikacji zakresu i zasad koncepcji dobrego rządzenia. Zeszyty Naukowe Uniwersytet Ekonomiczny w Krakowie, pp. 108-110. doi:10.15678/ZNUEK.2017.0962.0207

Tee, M. Y., & Lee, S. S. (2013). Advancing understanding using Nonaka's model of knowledge creation and problem-based learning. International Journal of Computer-Supported Collaborative Learning, p. 313.

Toyama, R., & Nonaka, I. (2015). The Knowledge-creating Theory Revisited: Knowledge Creation as a Synthesizing Process. In J. S. Edwards, The Essentials of Knowledge Management (pp. 95-110). London: Palgrave Macmillan.

Wyrozębski, P. (2012). Badanie praktyk zarządzania wiedzą projektową w polskich organizacjach. In P. Wyrozębski, M. Juchniewicz, & W. Metelski, Wiedza, dojrzałość, ryzyko w zarządzaniu projektami. Wyniki badań (pp. 23-123). Warszawa: Oficyna Wydawnicza SGH.

Wyrozębski, P. (2014). Zarządzanie wiedzą projektową. Warszawa: Difin.

Wyrozębski, P. (2019). Gromadzenie doświadczeń i zarządzanie wiedzą projektową. In P. Wyrozębski, Biuro zarządzania projektami (PMO) (pp. 151-164). Warszawa: Polskie Wydawnictwo Ekonomiczne.

Yasuoka, M. (2019). How To Transfer Tacit Knowledge for Living Lab Practice -Consideration on Tacit Knowledge Representations. Hawaii International Conference on System Sciences 2021. Honolulu : University of Hawai'i at Manoa.

Ziółkowski, S. (2017). Bariery dla transferu wiedzy wewnątrz organizacji. Studia Ekonomiczne. Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach, pp. 252-253.

Are we missing the importance of creativity in knowledge management?

PAUL LAUGHTON
Information Science, University of Pretoria,
Lynwood Road, Pretoria, South Africa
E-mail: paul.laughton@up.ac.za

Creativity is no stranger to knowledge management and has helped drive innovation and break new grounds in the management of knowledge. Organisations in general see creativity as a highly desirable outcome and strive to develop it further. This conceptual paper looks at a number of academic writings to identify the role of creativity in knowledge management. A number of common themes are identified, while highlighting the need for further research into the importance of creativity in knowledge management.

1. Introduction

The absence of creativity in any task may lead to under performance, and inability to improve (Elsbach & Hargadon, 2006). These outcomes can be limiting and restrictive. For organisations to run at their highest potential, creativity needs to be a focus or a core value. Egbu, Botterill and Bates (2001) identified the need for continuous innovation to have a competitive advantage over those that missed opportunities for innovation. This emphasis on creativity as an enabler for innovation should apply when one looks at knowledge management. By design knowledge management is a fairly abstract concept, as it regularly works with explicit resources, requiring some thoughtful creativity and innovation to be effective.

The notion of increasing the use of creativity in knowledge management may cause one to be self-reflective, asking ourselves how we currently apply creativity and where can creativity be further explored in our knowledge management efforts? Gurteen (1998:6) wrote "…. the term 'management' is inappropriate is in its relation to tacit knowledge - the knowledge in people's heads...................... knowledge management - in its creative sense - is more about nurturing than managing". To effectively manage a resource such as knowledge, creativity is required.

As knowledge management becomes increasingly popular and new technological resources are readily available for knowledge managers to use, the role of creativity should not be understated in the effective management of knowledge. This conceptual paper looks at previous works on knowledge management and creativity, as it seeks to better understand the role creativity can play in knowledge management while identifying strategies for increasing the use of creativity in knowledge management. A literature review was conducted to identify academic writings that feature both knowledge management and creativity together. Some of the findings are dealt with in the following section.

2. Creativity and knowledge management in academic literature

It is common a concept that the practice of knowledge management and creativity are very important factors for organisations working in this volatile, unstable, complex and ambiguous (vuca) world, where accelerated change is certain. Creativity and innovation involve the process of creating and applying new knowledge. This being the very epicenter of knowledge management (Gurteen, 1998). Knowledge and creativity are key resources that support those in positions to make decisions that could have significant impact in our current complex conditions (Giampaoli & Ciambotti, 2016).

While facing new challenges and obstacles, people and organisations may have to go beyond their tried and trusted knowledge maps and find a new path, potentially enabling them to find new solutions. Together knowledge and creativity are not only complementary but also synergic in the problem-solving process (Giampaoli & Ciambotti, 2016). The foundational elements of knowledge management (KM) (socialization, internalization, combination, & externalization) impact on an organisations ability to be creative (Rahimi, Arbabisarjou, Allameh & Aghababaei, 2011). While other humane aspects of knowledge management such as intuition may also contribute to the creative process (Akinci & Sadler-Smith, 2012). According to Gurteen (1998) “Knowledge Management needs to fundamentally focus on creativity and innovation”.

The source of creativity is the ability to create something. Creating, from a psychological perspective, refers to the production of something unique from other things. This encompasses the reduction or enhancement of a phenomenon, along with the transformation and combination of a phenomenon with other phenomena, objects, or things. Creativity also refers to the creation of something new (a product, a solution, a work of art, etc.) that has some kind of value (Pir Khaaefi, 1999).

Creativity is a desired behaviour in an organization, and highly valued (Baer & Oldham, 2006), as it brings into existence something which did not exist before, either as a product, a process or a thought (Rahimi, Arbabisarjou, Allameh & Aghababaei, 2011). According to Moorhead and Griffin (1989) creativity is the process of creating original perspectives and imagination on a given situation. Along with these definitions, creativity is also popularly seen as the ability to apply knowledge to solve problems.

Often, we do not need new information to form new ideas, but rather focus and think about the information and knowledge that we already have in abundance. Aligning with this, new thoughts and ideas come from a kind of thought-play of the mind (making it fun) which may involve concepts such as daydreaming of what could or might be (Gurteen, 1998). This aligns with Saulais and Ermine (2012) writings where they describe creativity “as an evolutionary process of an existing knowledge capital”.

People were born to be creative and the need to create is a common driving force in human beings (Smilkstein, 2004). Gurteen (1998) questions this notion and asks “why do we not see more creativity in our world, why are only a minority seen to be considered to be ‘creative and innovative’?” He answers this question with the limitation and impact on our perspectives and ways to think (world view).

Giampaoli and Ciambotti (2016) empirical research results confirm knowledge sharing increases organisational creativity, while having a direct and positive impact on an organisation’s ability to solve problems. These findings should encourage organisations to rely on their knowledge and creativity to achieve what others have not done before, key ingredients in the recipe to success. It is widely accepted that “knowledge is a key component of all forms of creativity” within the confines of modern innovation management (Chapman & Magnusson, 2006).

For creativity to exists, knowledge must not just be shared, but also be used and recombined (a transfer and processing of knowledge is essential). In some instances, the distinctions between knowledge creation and usage seem to be obscure, especially when facing complex systems such as creativity (Rahimi, Arbabisarjou, Allameh & Aghababaei, 2011).

Creativity is an outcome of the combination of existing knowledge and new knowledge (Kogut & Zander, 1992). The recombination of new and existing knowledge depends on the creation of knowledge and its use. More recently, the importance of knowledge creation and usage has been substantially supported for the creation of value and establishing them as crucial factors in the creativity process. According to Gurteen (1998) “creativity and innovation are at the cutting edge of knowledge management”.

Establishing knowledge creation and knowledge usage in the meta-model of knowledge management for creativity requires the following considerations (Rahimi, Arbabisarjou, Allameh & Aghababaei, 2011):

- The creation and usage of knowledge are two activities so closely related, one cannot be present without the other.
- The creation and usage of knowledge in the creativity process often cause a spiral effect, reciprocally functioning as cause and effect.
- The natures of the two activities are unique for creativity. For knowledge creation, the creativity of human beings, tacit and specialization of knowledge are keys for the novelty of creativity; for knowledge use, the harmony amongst users and the shared understanding, codifiability and diversity of knowledge are important for the success of creativity.

Findings from Giampaoli & Ciambotti (2016) confirm that elements such as work design, organisational culture and organisational structure that promote increased
creativity have the ability to enhance an organisation's ability to solve problems. Their findings also suggest that to successfully exploit opportunities and achieve what no one else has done before, people need to rely on both their knowledge and creativity (Giampaoli & Ciambotti, 2016).

The working environment has a significant impact on creativity, even if employees are very creative they will not be able to express their ability in a restrictive organizational structure or culture (Sawyer, 2006). Similarly, Liebowitz (1999:39) identified the importance of appealing incentives and a healthy culture for the encouragement of knowledge sharing and innovation. According to Vicari (1988) tolerant culture means creative culture. Scholars have begun to discover the importance of a firm's capacity to
put creativity into practice and not creativity itself (Klein & Sorra; 1996; Weinzimmer et al, 2011; Giampaoli & Ciambotti, 2016). Aligning with the concept of the organisation's culture, Gurteen (1998) identifies the biggest hinderance to creativity is the paradigm of an organisation. This concept of a paradigm relates to the way things are perceived, communicated and the way the world in interpreted. This concept of paradigm is often referred to as a worldview or mindset. A key factor to understand about a paradigm is
that it works at the subconscious level, "it's a bit like thinking the whole world is coloured red not knowing we are wearing rose-tinted contact lenses".

3. Contextual analysis of creativity and KM in academic literature

Creativity is not a new concept to knowledge management, and ultimately a desirable and natural outcome if knowledge is nurtured within your organisation. Based on the work of Gurteen (1998); Giampaoli & Ciambotti (2016); Rahimi, Arbabisarjou, Allameh & Aghababaei (2011); Chapman & Magnusson (2006) and Saulais & Ermine (2012) the following observations were drawn:

Value of creativity and knowledge management: all the authors had a high regard for creativity and viewed creativity has a desired outcome of knowledge management in an organisation. Creativity seems to be a critical feature for thriving in a competitive environment.

The relationship between knowledge management and creativity: knowledge was considered to be a pre-requisite for creativity to occur. Focus was placed on how creativity stems from existing knowledge and information. New ideas and creativity stem from things that currently occur in society, often it is the application of existing knowledge in a new context or scenario that generates creativity.

The role organisation culture plays in the support of creativity: the culture of an organisation plays a significant role in the nurturing of creativity. Organisation cultures that allow people a safe space to try something new with little judgement are best suited for creativity. Cultures that promote play and fun are more likely to generate creativity.

Creativity as a focal point within knowledge management: only one of the authors (Gurteen, 1998) focuses on the role of creativity within knowledge management. This finding seemed significant and lay the foundational motivation for challenging knowledge managers and researchers to put more emphasis on creativity when managing knowledge. Creativity was viewed by other authors as a complementary outcome when knowledge is freely shared and nurtured. Little is published on the use of creativity within knowledge management or how KM can incorporate more creativity to achieve it's desired objectives.

4. Inhibitors of creativity

Knowledge managers looking to encourage creativity should be aware of common inhibitors of creativity. According to Gurteen (1998), there are a number of common inhibitors of creativity that organisations and knowledge managers should be mindful of:

- Enforcing someone to learn, trying to make them more creative or getting them to do things in supposed disciplined ways, will often backfire and have the opposite effect.
- Fear is one of the more common blocks on creativity. Fear of the unknown, the what-ifs, fear of failure.
- New ideas by nature are extremely vulnerable, and easily destroyed if not thoughtfully nurtured. One should view new ideas as new born babies that require thoughtful attention and input to ensure they survive and become sustainable, giving them sufficient opportunity to develop.
- Due to the nature of the knowledge economy, information overload has become an inhibitor of creativity, with so much information available we can spend too much time trying to figure it out, possibly getting a little lost along the way and losing the initial inspiration.
- Judgement is yet another significant inhibitor of creativity. Often, one is quick to judge others which leads to a form of self-judgement, potentially holding one in a kind of 'creative straight jacket'. To be effective the organisation culture needs to discourage judgement.
- These items together form our paradigm (or world view) which often limits us or confines our thinking. The real limitation of creativity is the stranglehold that our limiting paradigms and mental models exert over us. We are prisoners of our own past - our upbringing, our education and past business experiences. These restrict and constrain us in our ability to think and act creatively.

Knowledge managers should look to reduce or restrict these inhibitors in an attempt to increase creativity. However other strategies can be implemented to enhance or improve the level of creativity, which are discussed in the following section.

5. Strategies for increasing creativity in knowledge management

There are a number of strategies that knowledge managers can potentially implement to enhance or and improve the level of creativity in knowledge management. The following are discussed in no order of significance.

Creative confidence is the ability to create a safe space for team members working with knowledge management to try new things, adapt current processes and practices with little to no pressure (Sakai-Miller, 2021). This creative confidence can enhance the practice dreaming, which allows people in your team to think without limitations (either technological or resources). When we think past our limitations and dream we can often come up with some very creative ideas, allowing people to entertain ideas that were thought to be previous out of reach (Abazov, 2021).

Learning from failure, requires an organisation to have a spirit of inquiry and a culture of openness and trust. Learning from failure requires thoughtful analysis of and discussion of potentially unwanted outcomes, often resulting as a by-product of experimentation.
Knowledge managers should look to foster this culture and focus on the potentially positive learnings that come from failure (Cannon & Edmondson, 2005).

Incorporate the concept of play at work, people who experience more fun and joy in their work environment are more likely to be comfortable with expressing their ideas and solutions to a particular problem (Ford, McLaughlin & Newstrom, 2003). Knowledge managers could further explore this concept of play in an attempt to unlock increasing levels of self-expression, potentially leading to further creativity and innovation.

The 50-solutions mentality, is a useful approach to challenges. This perspective encourages teams to stretch themselves to identify a significant number of solutions, while opening the door to viewing a challenge from a number of perspectives (Sakai- Miller, 2021). The increasing number of possibilities may be useful in identifying new ways to achieving a desired outcome, which can be of use to knowledge managers.

Set aside time for creative thinking, a positive way to encourages creative thinking is to set aside time for collaborative creative thinking. Often people are too busy or not focussed on creative thinking, but when we ensure we set time aside the barriers to creative thinking are significantly reduced. Regular team interaction focusing on creativity can positively impact the creativity of the team (Abazov, 2021).

6. Conclusion

As the practice of knowledge management continues to evolve and adapt to incorporate changes in our environments and changes in technology, the role of creativity should not be understated. Creativity is highly desirable, but like anything of value it takes special input to make it flourish. The management of knowledge in any organisation is a demanding task that requires creativity and innovation for it to successful, supporting Gurteen (1998) view that “creativity and innovation are at the cutting edge of knowledge management”.

For knowledge management to reach new heights greater emphasis and focus should be placed on the role of creativity. The very notion of managing a resource that cannot be seen (tacit by nature) should evoke a number of unconventional and unique approaches. This may cause one to further explore the possibility of being more creative in our practice.

For creativity to abound in knowledge management we need to seriously reflect on our world view and limiting beliefs which keep us from going further outside of the box. One needs to also closely manage the culture in which teams are operating in to ensure freedom of ideas and expression is a reality, limiting the judgement. There is so much potential in knowledge management, it is up to us to take it to the next level, with the help of creativity and possibility.

References

Abazov, R. (2021). *5 Ways to improve your creative thinking*. Available online: https://www.topuniversities.com/blog/5-ways-improve-your-creative-thinking (Accessed 28 April 2022).

Akinci, C. & Sadler-Smith, E. (2012). Intuition in management research: a historical review. *International Journal of Management Reviews*, 14: 104–122.

Cannon, M. D. & Edmondson, A. C. (2005). Failing to learn and learning to fail (intelligently): How great organizations put failure to work to innovate and improve. *Long range planning*, 38(3): 299-319.

Chapman, R. & Magnusson, M. (2006). Continuous innovation, performance and KM. *Knowledge and Process Management*, 13(3): 129-31.

Egbu, C., Botterill, K. & Bates, M. (2001). *The influence of knowledge management and intellectual capital on organisational innovations*. Proceedings of the 17th Annual conference of the Association of Researchers in Construction Management, 5-7 September 2001, Salford, UK.

Elsbach, K. D. & Hargadon, A. B. (2006). Enhancing Creativity Through "Mindless" Work: A Framework of Workday Design. *Organization Science*, 17 (4): 470-483.

Ford, R. C., McLaughlin, F. S. & Newstrom, J. W. (2003). Questions and Answers about Fun at Work. *Human Resource Planning*, 26(4): 18-33.

Giampaoli, D. & Ciambotti, M. (2016). *The Vital Role of Knowledge Management and Creativity for Performance*. 17th European Conference on Knowledge Management[7], Belfast.

Gurteen, D. (1998). Knowledge, creativity and innovation. *Journal of Knowledge Management*, 2(1): 5-13.

Klein, K. & Sorra, J. (1996). The challenge of innovation implementation. *Academy of Management Review*, 21(4): 1055-1080.

Kogut, B. & Zander, U. (1992). Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization Science*, 3: 383-397.

Liebowitz, J. (1999). Key Ingredients to the success of an organization's knowledge management strategy. *Knowledge and Process Management*, 6 (1): 37–40.

Moorhed, G. & Griffin, R. W. (1989). *Organizational behavior* (2nd Ed). Boston: Houghton Mifflin Company.

Rahimi, H., Arbabisarjou, A., Allameh, S. M. & Aghababaei, R. (2011). Relationship between Knowledge Management Process and Creativity among Faculty Members in the University. *Interdisciplinary Journal of Information, Knowledge, and Management*, 6.

Sakai-Miller, S. (2021). *4 Strategies for Boosting Creativity*. Available online: https://www.iste.org/explore/classroom/4-strategies-boosting-creativity (Accessed 25 April 2022).

Saulais, P. & Ermine, J. L. (2012). Creativity and knowledge management. Vine: the *Journal of Information and Knowledge Management Systems*, 42(2/3).

Sawyer, R.K. (2006). *Explaining creativity: the science of human innovation*. Oxford University Press, New York, NY.

Smilkstein, R. (2004). Born to be creative and critical thinkers. Proceedings of Imaginative Education Research Group Conference, Vancouver, BC Retrieved Dec (Vol. 23, p. 2007).

Sohail, M. & Saeed. M. (2018). Relationship between knowledge management and creativity among teachers of public and private sector universities at Lahore. *Bulletin of Education and Research*, 40(2): 91-104.

Vicari, S. (1998). *La creatività di impresa*. Tra caso e necessità, ETAS, Milano, IT.

Weinzimmer, L.G., Michel, E.J. & Franczak, J. (2011). Creativity and firm-level performance: the mediating effects of action orientation. *Journal of Managerial Issues*, 23(1): 62–82.

Track – Big Data

Prediction of concrete bridge deck condition ratting based on climate data in addition to bridge data: five states as a case study

Fariba Fard

Department of Information Science, University of North Texas, 1155 Union Circle #311068, Denton, TX 76203-5017, USA
E-mail: Fariba.Fard@unt.edu

Keywords: bridge deck. condition rating prediction. random forest. Gradient Boosting Machine (GBM).

Evaluating the impact of learning from climate data, in addition to bridge data, on the performance of concrete deck condition rating prediction is critical for identifying the right data needed to enhance bridge maintenance decision making. Few studies have considered such an evaluation and utilized a small size of samples that prevent revealing the knowledge hidden within the big size of data. Although, such evaluation over big data seems quite necessary, class imbalance problem makes it challenging. To alleviate such a problem, five states, including Alabama, Iowa, New York, Pennsylvania, and South Carolina, were selected as the case study. Not only are the states located in three different climatically consistent regions defined by the National Ocean and Atmospheric Administration (NOAA), but also their concrete deck conditions ratings are somewhat balanced. To conduct the evaluation, this research developed the bridge data set pertaining to 56,288 bridges across the afore-mentioned states through employing the GIS technology. The bridge data set contains bridge data derived from National Bridge Inventory (NBI), and climate data derived from Parameter-elevation Relationships on Independent Slopes Model (PRISM) climate maps and NOAA. Then, two machine learning algorithms, including random forest and GBM, were trained
- with and without climate data - and their prediction performances were compared. The results indicated that: (1) random forest outperforms GBM with an accuracy of 63.3%, and (2) the change in the prediction performance after further learning from climate data was marginal since the accuracy reached to 64.9%.

INTRODUCTION

Bridges are critical nodes in the U.S. transportation system, and their performances affect the operating capacity of the system in terms of safety, efficiency, and economy (Hooks & Frangopol, 2013). The availability of data in various sources provides an opportunity for developing machine learning models aimed at predicting bridge deterioration that is fundamental for well-informed decision in bridge maintenance and rehabilitation. To do so, a framework is required to build a conceptual foundation on collecting data from different and heterogeneous sources which ultimately leads to developing accurate models to predict bridge deck condition rating.

BIG BRIDGE DATA ANALYTICS FRAMEWORK

Liu and El-Gohary (2016) established a big bridge data analytics framework for advancing bridge deterioration prediction and interpretation. Three main types of data are processed, integrated, and analyzed in the big bridge data analytics framework. These data include 1) NBI, 2) traffic, and 3) climate data.

RELATED WORKS AND KNOWLEDGE GAP

Majority of the existing studies that developed predictive models, have used only bridge data and did not consider weather data or sufficiently evaluate their impact on the prerformance of the models. For instance, Huang (2010) developed an artificial neural network (ANN) model that learned only form bridge data and not from climate data. On the other hand, Huang, Mao, and Lee (2010) employed a data mining technique to explore factors leading to deterioration of bridges. The authors infer that not only bridge characteristics, but also weather-related factors influence deterioration. Chang, Maguire, and Sun (2017) combined least absolute shrinkage and selection operator (LASSO) penalized regression with the cross validation to select the influential predictor variables on deterioration of Wyoming bridges. The results indicate that climate data are not influential. In another study conducted by Kim and Yoon (2010), a multiple regression model was developed on bridges in North Dakota. The year built, followed by the volume of traffic, and structural characteristics were identified as the most contributing predictor variables to deterioration. While climate data, such as precipitation, snow fall, and temperature were not identified among those variables.

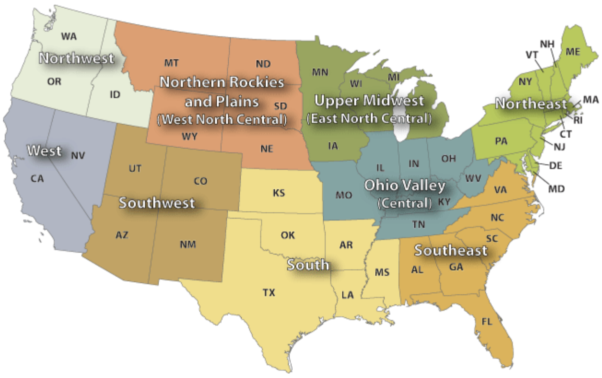
SIGNIFICANCE OF THE STUDY

The objective of this study is to incorporate a big size of bridge data to quantify the impact of climate data on the performance of the predictive models.

CASE STUDY

Concrete bridge decks located in five states, including Alabama, Iowa, New York, Pennsylvania, and South Carolina, were selected as the case study. Concrete bridge decks are those that their type of deck system or item 107, in the NBI data, is equivalent to “Concrete Cast-in-Place” or “Concrete Precast Panels” (Santamaria Ariza, Zambon, S. Sousa, Campos e Matos, & Strauss, 2020). Such conditions were applied to the Big Bridge Data (BBD) collected by Fard (2022) that contains bridge data derived from NBI, and climate data derived from PRISM climate maps and NOAA. By applying that condition, 56,288 bridges across the afore-mentioned states were retained for conducting the experiment. Note that these five states are in three different climatically consistent regions defined by the NOAA, as shown in Figure 1. According to Figure 1, Pennsylvania and New York are in the Northeast, Iowa is in the upper Midwest, while South Carolina and Alabama are in the Southeast region.

Figure 1 Nine climatically consistent regions within the U.S. “Reprinted from National Centers for Environmental Information”.



PREDICTOR AND RESPONSE VARIABLES

Total of 25 predictor variables, including 17 numerical and 8 categorical, were introduced as predictor variables. Table 1 lists the 27 predictor variables.

Table 1 List of 25 predictor variables.

No.	Predictor variable	Type	No.	Predictor variable	Type	No.	Predictor variable	Type
1	Age	Numeric	10	Precipitation	Numeric	18	Highway_District	Categorical
2	ADT	Numeric	11	Temp_Min	Numeric	19	Design_Load	Categorical
3	ADTT	Numeric	12	Temp_Max	Numeric	20	Reconstructed	Categorical
4	Lanes_On	Numeric	13	Days_MinLT_32F	Numeric	21	Main_Material	Categorical
5	Number_Spans_Main	Numeric	14	Days_MaxGT_70F	Numeric	22	Main_Design	Categorical
6	Length_Max_Span	Numeric	15	Avg_Wind_Speed	Numeric	23	Spans_Material	Categorical
7	Curb_Width	Numeric	16	Peak_Wind_Speed	Numeric	24	Spans_Design	Categorical
8	Deck_Area	Numeric	17	Precipitation	Numeric	25	Deck_Geometry	Categorical
9	Operating_Rating	Numeric						

Basically, condition ratings of bridge decks vary between 0 and 9. Table 2 represents descriptions of the bridge deck in various conditions ratings, which were taken from the U.S Department of Transportation (FHWA 1995).

Table 2 Description of deck condition ratings.

Condition rating	Description	Condition rating	Description
9	Excellent condition	4	Poor condition
8	Very good condition	3	Serious condition
7	Good condition	2	Critical condition
6	Satisfactory condition	1	Imminent failure condition
5	Fair condition	0	Failed condition

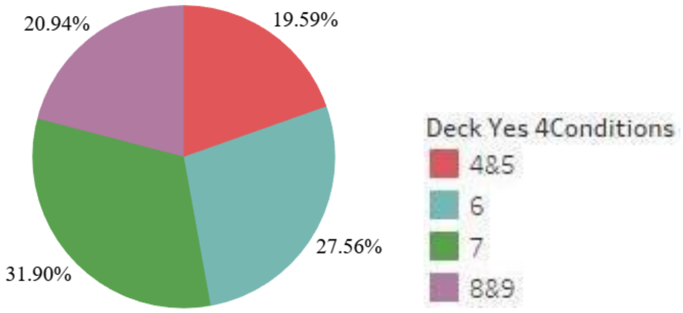
Note that in the five states, there was no concrete bridge deck with condition rating of 3, and a negligible proportion (0.79%) of the concrete bridge decks had condition ratings of 0,1, and 2. In addition, bridges with condition ratings of 3 or less have reached serious conditions that need reconstruction, not inspection. Therefore, concrete bridge decks with condition ratings of 0, 1, and 2 were removed from further analysis. In addition, concrete bridge decks with condition ratings of 4 and 5 as well as 8 and 9 were combined and two new condition ratings, termed as 4&5 and 8&9, were created. Therefore, the response variable is a multiclass condition rating of concrete decks, with four different categories including 4&5, 6, 7, and 8&9, as listed in Table 3.

Table 3 Multiclass response variable.

Response variable	Values
Multiclass	4&5, 6, 7, 8&9

Combing samples from such nearby condition ratings has been used by the existing studies to create a balanced dataset (Zhu & Wang, 2021). As shown in Figure 2, the response variable is somehow balanced.

Figure 2 Balanced response variable with four classes.



EXPERIMENT SETUP

The total number of 56,288 concrete bridge decks across the five states were split into 80% and 20% of the training and test data sets, respectively. The training data set was used to develop two classification models, including random forest and GBM, with and without climate data, using h2o package (Cook, 2016) in R programming language. The models intend to classify concrete decks, into 4 condition ratings, including 4&5, 6, 7, and 8&9. For classification tasks, $ntree = 500$ and $mtry = 5$ were introduced for both classification algorithms, where $ntree$ and $mtry$ denote the number of trees used to build the model and the number of variables randomly selected at each node, respectively.

RESULTS AND DISCUSSION

Then, the prediction performance of the developed models was evaluated using the test data. According to Table 4, random forest and GBM achieved accuracy of 64.9%, and 63.4% by contributing climate data, and 63.3 and 60.9 without contributing the climate data, respectively.

Table 4 Multiclass response variable.

Machine learning model	Test data	
	With climate data	Without climate data
Random forest	64.9	63.3
GBM	63.4	60.9

CONCLUSION

The results indicate that random forest outperforms GBM with an accuracy of 64.9% in predicting condition ratings of bridge decks in five states. However, climate data marginally could increase the accuracy by only 1.6%. Further research seems necessary to develop other machine learning models and compare their performances with the results obtained in this study. In addition, incorporating of bridge historical data is recommended that may lead to a better predictive model.

REFERENCES

Breiman, L. 2001. "Random forests." *Machine Learning* 45 (1): 5-32. https://doi.org/10.1023/A:1010933404324.

Chang, M., M. Maguire, and Y. Sun. 2017. "Framework for mitigating human bias in selection of explanatory variables for bridge deterioration modeling." *Journal of Infrastructure Systems* 23 (3): 04017002. doi:https://doi.org/10.1061/(ASCE)IS.1943-555X.0000352.

Cook, D. 2016. *Practical machine learning with H2O: powerful, scalable techniques for deep learning and AI.* California, United States: O'Reilly Media, Inc.

Fard, Fariba. 2022. *FaribaFard/BigBridgeProject_V1.0: First Release.* Dallas, Texas, April 25. https://github.com/FaribaFard/BigBridgeProject_V1.0.

Huang, R. Y., I. S. Mao, and H. K. Lee. 2010. "Exploring the deterioration factors of RC bridge decks: a rough set approach." *Computer-Aided Civil and Infrastructure Engineering* 25 (7): 517-529. doi:https://onlinelibrary.wiley.com/doi/10.1111/j.1467-8667.2010.00665.x.

Huang, Y. 2010. "Artificial neural network model of bridge deterioration." *J. Perform. Constructed Facil* 24 (6): 597–602.

Kim, Y. J., and D. K. Yoon. 2010. "Identifying critical sources of bridge deterioration in cold regions through the constructed bridges in North Dakota." *Journal of Bridge Engineering* 15 (5): 542-552. https://doi.org/10.1061/(ASCE)BE.1943-5592.0000087.

Liu, K, and N. El_Gohary. 2016. "Semantic modeling of bridge deterioration knowledge for supporting big bridge data analytics." *In Proc. 2016 ASCE Construction Research Congress (CRC).* Reston, VA: ASCE. 930-939. https://doi.org/10.1061/9780784479827.094.

Santamaria Ariza, M., I. Zambon, H. S. Sousa, J. A. Campos e Matos, and A. Strauss. 2020. "Comparison of forecasting models to predict concrete bridge decks performance." *Structural Concrete* 21 (4): 1240-1253.

USDOT. 2021. *Federal highway administration national bridge inventory.* 03 09. Accessed June 1, 2021. https://www.fhwa.dot.gov/bridge/nbi/ascii.cfm.

Zhu, J., and Y. Wang. 2021. "Feature Selection and Deep Learning for Deterioration Prediction of the Bridges." *Journal of Performance of Constructed Facilities* 35 (6). doi:https://doi.org/10.1061/(ASCE)CF.1943-5509.0001653.

Stock2Vec: An Embedding to Improve Predictive Models for Companies

Ziruo Yi[a], Ting Xiao[a,b], Ijeoma Kaz-Onyeakazi[b], Cheran Ratnam[b], Theophilus Medeiros[a], Phillip Nelson[c] and Yuvraj Baweja[c]

[a]Department of Computer Science and Engineering, University of North Texas, US

[b]Department of Information Science, University of North Texas, US

[c]Texas Academy of Math and Science, US

ziruoyi@my.unt.edu, ting.xiao@unt.edu, ijeomakaz-onyeakazi@my.unt.edu, cheran.ratnam@unt.edu, theophilusmedeiros@my.unt.edu, phillipnelson@my.unt.edu, yuvrajbaweja@my.unt.edu

ABSTRACT

Building predictive models for companies often relies on inference using historical data of companies in the same industry sector. However, companies are similar across a variety of dimensions that should be leveraged in relevant prediction problems. This is particularly true for large, complex organizations which may not be well defined by a single industry and have no clear peers. To enable prediction using company information across a variety of dimensions, we create an embedding of company stocks, Stock2Vec, which can be easily added to any prediction model that applies to companies with associated stock prices. We describe the process of creating this rich vector representation from stock price fluctuations and characterize what the dimensions represent. We then conduct comprehensive experiments to evaluate this embedding in applied machine learning problems in various business contexts. Our experiment results demonstrate that the four features in the Stock2Vec embedding can readily augment existing cross-company models and enhance cross-company predictions.

AUTHOR KEYWORDS

Stock2Vec; predictive models; machine learning; natural language processing.

1. INTRODUCTION

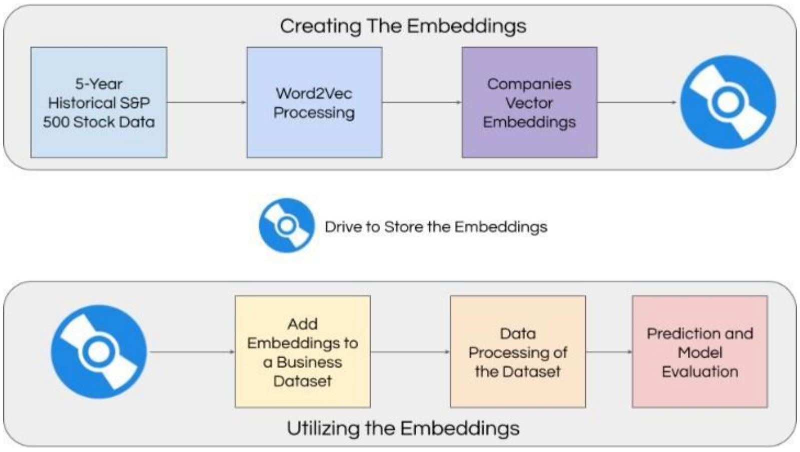
Traditionally, predictive models are used to solve prediction problems of companies based on inference leveraging historical data of companies in the same industry. For instance, sales forecasting models can relate company sales to industry sales, and they also relate company variables to industry variables to solve short-, medium- and long-term prediction problems (Sharp et al., 1982). However, many companies, especially those that are large and complex, are growing businesses across different industries so they do not simply belong to one specific industry. Moreover, some innovative companies do not have peers yet, so no existing industry can accurately define them. Therefore, only considering data from companies in the same industry is not enough and will lead to inaccurate predictions of companies. Besides industry, companies are similar to each other and share characteristics in other dimensions such as the geographic location and the company size. This raises a question: can we learn cross-company information in different dimensions to reduce the limitation of existing predictive models?

Companies are complex organizations and cannot be adequately represented by text-based embeddings due to three reasons: First, information about companies may not be fully available in a corpus. Second, information about companies may be outdated. Third, information about a company may need higher-order reasoning than what

a text-based embedding can provide. Compared with texts, stock prices, which are the valuation of companies, are more informative. For instance, changes to the structure and function of a company can ultimately affect its stock price. Additionally, a stock reflects the cumulative prediction of many external observers to a company's performance. In essence, the fluctuations of stock prices are sensitive to a variety of company characteristics.
Moreover, the factors that impact one company can also impact a similar company in many dimensions such as geography and industry. Therefore, creating a non-textual embedding based on stock price changes across companies over time is a good way to represent a company.

In this paper, we introduce Stock2Vec, an inexpensive but efficient embedding of company stocks to learn cross-company inference. Figure 1 illustrates the two-stage pipeline of our work. In the first stage, we use ordered daily price changes in S&P 500 stock over a 5-year span and the Word2Vec algorithm to create an embedding for each stock, then we determine the most useful four features using machine learning models and obtain an optimized 4-dimensional embedding as our Stock2Vec embedding. In the second stage, we evaluate the Stock2Vec embedding by adding it to predictive models and working on two company characteristic prediction tasks: predicting the company's environmental impact and the company size with and without the Stock2Vec embedding. Experiments show that the addition of the Stock2Vec embedding enables a significant performance boost over using the original models only.

Figure 1
Overview of the two-stage pipeline of creating and utilizing the Stock2Vec embedding.



Our contributions are summarized as follows: (i) We introduce Stock2Vec, an inexpensive but powerful embedding to represent companies. (ii) We present a detailed analysis of the Stock2Vec embedding. (iii) We achieve better performance on two company characteristic prediction tasks with the addition of Stock2Vec embedding to predictive models. We also make a discussion on the experiment results.

2. LITERATURE REVIEW

2.1. Stock Predictions and Machine Learning

Stock prediction is important in data-driven decision making and deriving strategies. Stock forecasting has been focused on informative external data sources such as: accounting performance of companies, macroeconomic effects, etc. (Wang et al., 2020). Machine learning methods can predict trends (Behpour et al., 2021), analyze large- scale financial datasets and outperform traditional statistical time series models. Therefore, machine learning methods have been used along with historical stock prices in traditional technical analysis. Wu et al. (2019) propose that machine learning employs two major approaches in forecasting equity returns: The first method is Artificial Neural Network (ANN), which is used as a black box for standard factors correlated with equity returns. The second approach is forecasting the price time-series, in which price, trading volume, and other statistics are used to represent trading activities as a time-series. The researchers also identify three weaknesses in existing models used by computer scientists and financial professionals: undefined source of alpha, lack of heterogeneous sets used in models, and ineffectiveness in leveraging cross-sectional effects, which result in forecasting problems in the training pairs. Shen et al. (2012) postulate that features, which are used as inputs to algorithms, are mostly derived from industry-specific data. Since these approaches ignore pertinent information that may be available in other entities, the researchers break through the industry-centric barriers by building a trading model to utilize features that are not bound by the market.

Reinforcement learning (RL), which is generally used to solve the control problem in the power system (Yu et al., 2019), can also be applied to stock predictions. Shin et al. (2019) proposes a deep multimodal reinforcement learning policy combining convolutional neural network (CNN) and long short-term memory (LSTM) model to predict stock prices: various charts are generated from stock trading data and go through the CNN layer and the LSTM layer to become features. Then they use reinforcement learning to define agents' policy neural network structure, reward and action, and provide buying, selling and holding probabilities as final output.

2.2. Word Embedding and Word2Vec Algorithm

Knowledge representation is critical to learning inference and there are many ways to draw meaning from information. For instance, exploring similarities through comparison is used to measure similar states and circumstances that lead to predictive outcomes. One method to draw meaning through associative relationships between entities is embedding. Embeddings are low-dimensional translations of high-dimensional vectors, where vectors with stronger relationships are placed closer together. Word embeddings cluster similar words together and represent words in multi-dimensional vectors (Rizkallah et al., 2020). Thus, word embeddings provide highly informative context about the relationship of words and transform seemingly arbitrary combinations of letters into representations that are useful for reasoning and generalization. Mikolov et al. (2013) find that word embeddings retain 70% of their original meaning and help models perform well in natural language processing tasks. In natural language processing, there are many applications for word embeddings, for example, being used in recommendation systems, named entity recognition, and sentiment analysis.

Researchers in Google introduced Word2Vec, a popular embedding that carries correlated features based on statistical relationships of words and yields a vector representation of words in sentences of large corpuses. The Word2Vec algorithm is a fusion of two learning models: Continuous Bag of Words (CBOW) and Skip-gram (Mikolov et al., 2013). The Word2Vec embedding not only allows reasoning based on similarities but also preserves relative relationships among mathematical features. For instance, word embeddings can follow reasoning similarly to conceptual understandings such as "woman" + ("king" - "man") = "queen" and "swimming" + ("walked" – "walking") = "swam". The success of Word2Vec has led to a series of similar strategies that successfully create data representations in different forms and contexts. For example, Node2Vec (Grover and Leskovec, 2016) creates vector representations of nodes in a graph based on the context of connections within the graph. Doc2Vec (Lau and Baldwin, 2016) creates a vector representation of entire documents instead of words. And deep learning models such as XLNet are used to capture more complex and context-dependent relationships (Peters et al., 2018; Yang et al., 2019).

2.3. Stock2Vec Works

The Word2Vec algorithm has been used as a sector embedding for predictions related to company stock attributes, and in some cases named as Stock2Vec. Stock2Vec in Wang et al. (2020)'s study learns the intrinsic relationships among stocks to make better predictions, in which more similar stocks are categorized closer to deploy interactions in the embedding. Lien Minh et al. (2018) also build a Stock2Vec embedding to predict stock movement. First, they extract and preprocess financial news from a financial news dataset, then label news and stock prices from the news as either negative or positive. Using labeled news and Harvard IV-4 dictionary, the Stock2Vec embedding is created and then trained by a Bidirectional gated recurrent unit (BGRU) on daily S&P 500 Index stock prices from Yahoo Finance. Finally, the trained Stock2Vec embedding is used to predict the effect of financial news on stock prices. The researchers also find that Stock2Vec is more effective than other embedding methods such as Glove and Word2Vec since it takes the sentiment value of words into consideration. Similarly, Lu et al. (2021) also train Stock2Vec on stock news and sentiment dictionaries. But they include political news and stock forum speech in sentiment analysis, and the Stock2Vec embedding is trained on CSI 300 stock data. They first use the Bayesian model for sentiment classification of the stock forum speech, then use Bidirectional Long Short-Term Memory (Bi- LSTM) to extract trading data and investor sentiment index-related features, and finally use Contextual Long Short- Term Memory (CLSTM) to integrate and process the political news within context. The stock trend is classified as positive with the probability of the stock going up and as negative if in the reverse case.

Compared with previous works, our work has two main differences: First, the Stock2Vec embeddings in existing works are derived with text data from news and sentiment dictionaries while our Stock2Vec embedding is generated on numeric data from daily price change of S&P 500 Index stock. This is premised on the fact that organizations within the same industry share similar growth or decline in Corporate Social Performance (CSP), which is asserted by Short et al. (2016). Second, previous works focus on predicting stock trends using company characteristics, but none of them leverage Word2Vec algorithm to build an embedding that utilizes stock prices to predict company characteristics. Dorfleitner et al. (2015) find that Environment, Social, Governance (ESG) scores play a major role in the decision-making process of managers and investors who are mindful of social responsibilities among other things. Crespi and Migliavacca (2020) underscore this by stating that the tendency of linear growth in the ESG score is enhanced by their size and profitability. Zumente and Lāce (2021) posit that ESG score has a correlation with the trading volume of company stock. Therefore, we use ESG rating and company size as our company characteristic prediction tasks, and we show that our Stock2Vec embedding can effectively predict company characteristics.

3. STOCK2VEC EMBEDDING

In this section, we first introduce the two datasets used and how to combine and preprocess them to the appropriate format for our work, then we describe the detailed steps to generate our Stock2Vec embedding.

3.1. Dataset and Preprocessing

The dataset used in our work is a combination of two datasets. The first dataset is the S&P 500 stock data[1], which is about historical stock prices for all companies in the S&P 500 index from 2013 to 2018. This dataset contains daily trading information including trading date, opening price, highest price, lowest price, closing price, the number of shares traded, and company stock name. The other dataset is the S&P 500 companies with Financial Information[2], from which we can get S&P 500 companies' names, sectors, and symbols.

Our data preprocessing step is done using Python. We first combine two datasets according to the companies' stock names and then remove irrelevant information such as the number of shares traded. After the cleanup, the preprocessed dataset contains the daily price change of the S&P 500 stocks, corresponding sectors, and dates. Each of the 505 stocks is represented by 1826 data points (365*5+1) containing the daily stock price change for each day from 2013 to 2018.

3.2. Creating the Stock2Vec Embedding

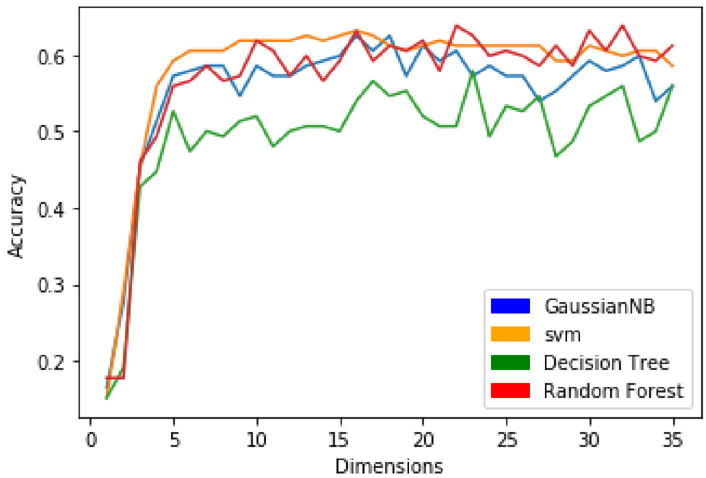
We first order the stocks in the dataset based on their daily price change for each day. The purpose of this step is to create a 'sentence', which is a sequence of stock fluctuation information for each day. The daily stock price change is calculated as the difference between the closing and opening price for one day, and then divided by the opening price.

[1] https://www.kaggle.com/camnugent/sandp500
[2] https://datahub.io/core/s-and-p-500-companies-financials#resource-s-and-p-500-companies-financials_zip

The 1826 generated stock sentences are then input to the Word2Vec algorithm which results in a high- dimensional Word2Vec embedding. One problem is that what each feature represents is unclear, because feature values are arbitrary numbers that describe relationships between objects. As a result, the optimal number of features required to represent a company is unclear. To decide an appropriate dimensionality for our Stock2Vec embedding, one efficient way is to observe the utilization of the embedding in prediction problems and identify the minimum number of dimensions necessary to achieve a high accuracy. Specifically, we define an industry sector prediction task based on the daily stock price changes, then split the data into a 70% training set and a 30% testing set respectively. Next, we use the Word2Vec embedding on four classifiers: Gaussian Naive Bayes (Rish et al., 2001), Support Vector Machines (SVM) (Zhang, 2012), Decision Tree (Safavian and Landgrebe, 1991), and the Random Forest (Breiman, 2001). If one classifier gets a higher accuracy with the Word2Vec embedding at a given dimensionality, it stands to reason that the Word2Vec embedding with that dimensionality can better represent the original data. Afterwards, we train the Word2Vec embedding on the training set and then classify each company to its corresponding sector with the remaining test set.

Figure 2 demonstrates the accuracy of each classifier's prediction based on the Word2Vec embedding with varying dimensions. The accuracy of four classifiers begin to plateau at roughly 4 dimensions, which means that afterwards increasing dimensions of embedding only increases the prediction accuracy by a negligible amount.
Thus, we decide the optimal dimensionality of the embedding is four and named the new four-dimensional embedding Stock2Vec. Obviously, Stock2Vec is a lower dimensional embedding for a simpler but more meaningful representation of each company.

Figure 2
Sector prediction accuracy of four classifiers based on the Word2Vec embedding with varying dimensions. It can be noted that the accuracy of all classifiers plateau when the embedding has more than four dimensions.



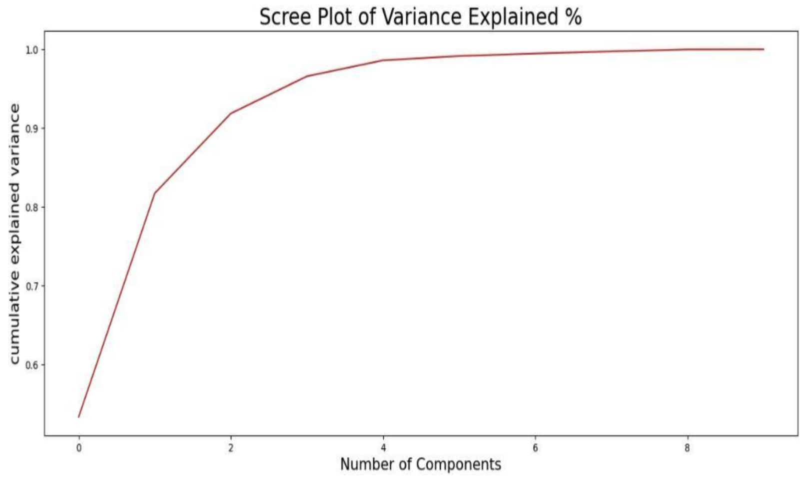
4. STOCK2VEC EMBEDDING ANALYSIS

We analyze the nature of our Stock2Vec embedding from two angles. We first check whether the Stock2Vec embedding is optimal, then we explore whether the Stock2Vec embedding leads to prediction mistakes.

4.1. Checking the Stock2Vec Embedding

To further confirm that our four-dimensional Stock2Vec embedding is optimal, we apply Principal Component Analysis (PCA) (Shlens, 2014) to create a graphical representation of the Word2Vec embedding and visualize it on a cartesian coordinate plane. PCA is an algorithm utilized to compress data into a lower-dimensional space but is less adept at retaining relationships between other objects. If the Word2Vec embedding gets a larger explained variance at a given dimensionality, that means it represents a higher percentage of information of the original data. Figure 3 below shows that when the dimensionality of embedding is four, the embedding captures over 95% of the variance in the dataset, which means that the four-dimensional embedding maintains around 95% information of the original data. Moreover, the Word2Vec embeddings with less than four dimensions maintain much less original information while embeddings with more than four dimensions do not maintain obviously more original information. Although embeddings with more than four dimensions are a little bit more informative, higher- dimensional embedding requires more computing resources, which is very costly. Therefore, four is confirmed as the most appropriate dimensionality and is applied to our Stock2Vec embedding.

Figure 3
Principal Component Analysis of the Wor2Vec embedding in varying dimensions and their corresponding explained variances. It shows how well the Stock2Vec embedding can capture the information in the original data.



4.2. Characterizing the Stock2Vec Embedding

As shown in Figure 2, the SVM model provides a more stable metric of the accuracy out of the four classifiers. Thus, we use the SVM model to discuss the company sector prediction results. The SVM classifier can consistently attain around 60% accuracy when predicting sectors utilizing our Stock2Vec embedding, which indicates that about 60% of the information of a given company could be represented by the Stock2Vec embedding. There are mistakes in the prediction results and the SVM classifier determines which sectors are commonly confused and therefore are more closely related. To analyze the prediction mistakes, we create a confusion matrix to visualize these confusions.

Figure 4 shows the confusion matrix produced by the SVM classifier, in which the commonly mistaken sectors can be clearly identified. The sectors along the vertical axis of the matrix indicate the companies' actual sectors while the sectors along the horizontal axis indicate the sectors predicted by the SVM classifier. Therefore, accurate predictions lie along the diagonal and confusions lie outside the diagonal. Sectors such as Health Care and Information Technology are often misidentified, but these sectors have overlap in their services. This fact suggests that most of the incorrect predictions are due to the similarities between sectors, which means that these prediction mistakes are not caused by our Stock2Vec embedding.

This conclusion is further supported by Figure 5, which is the PCA plot showing the clusters formed by the companies' sectors. Sectors such as Utilities and Energy have few errors in the confusion matrix, and the clusters formed by them on the plot are far from clusters formed by other sectors. This fact suggests that Utilities and Energy do not share many traits with other sectors. In contrast, the clusters formed by highly confused sectors such as Industrials, show a wider distribution that overlaps with various clusters formed by other sectors. Therefore, our Stock2Vec embedding makes it easy to visualize and interpret the relationships between companies and their corresponding sectors.

Figure 4
Confusion matrix of the SVM classifier predicting companies' sectors using the Stock2Vec embedding. Redder areas indicate a greater amount of confusion while bluer areas indicate little to no confusion.

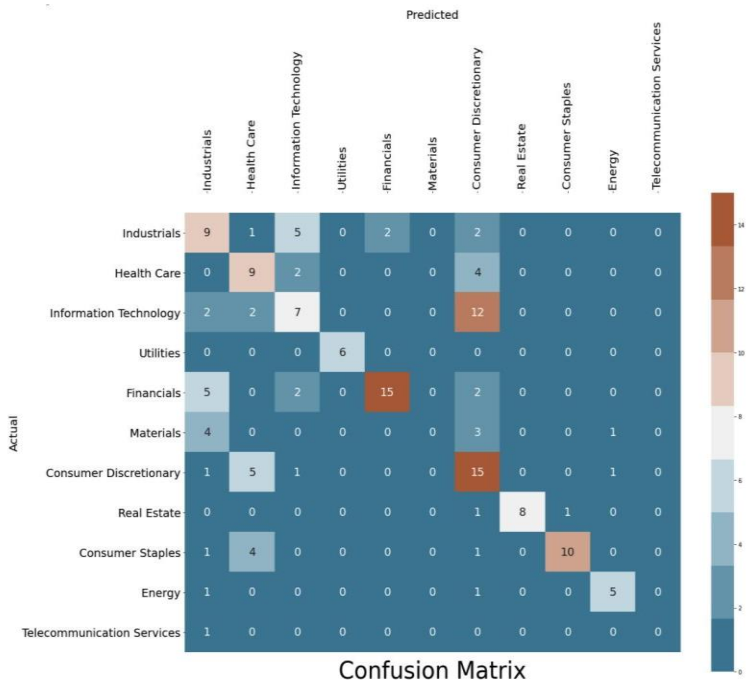
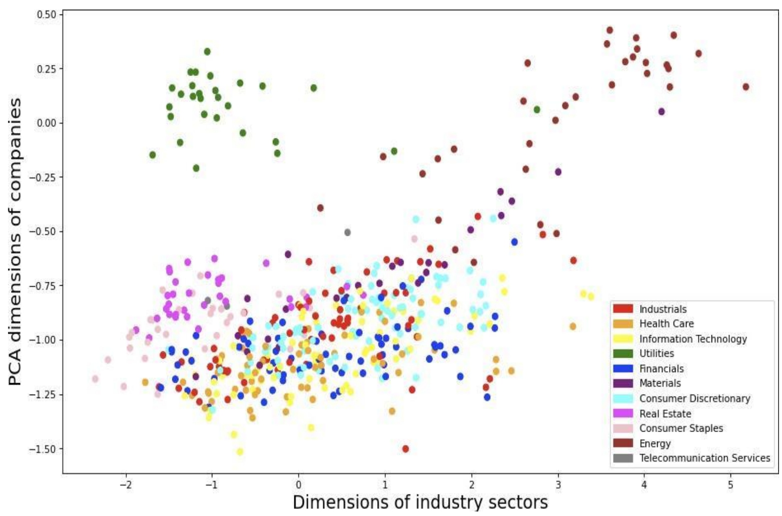


Figure 5
PCA plot of colored-coded companies and their corresponding sectors. The corresponding relationships are indicated down right.



5. APPLICATIONS

To evaluate the viability of the Stock2Vec embedding, we design two prediction tasks of company characteristics: predicting the environmental impact of companies and predicting the size of companies. We also compare the results of using and not using the Stock2Vec embedding in these two tasks.

5.1. Data

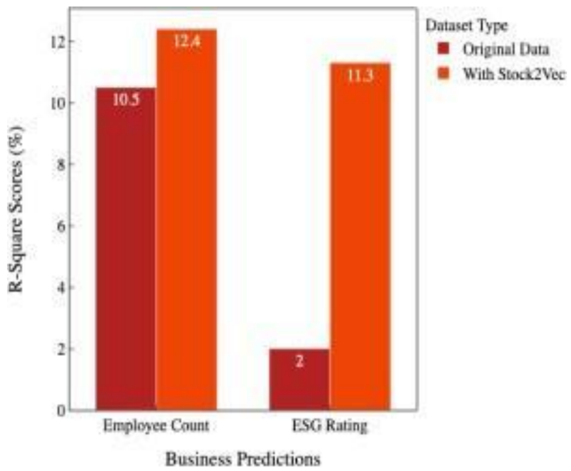
For the prediction task of companies' environmental impact, we gather the Environmental, Social, and Governance (ESG) ratings' data from Sustainalytics web portal[3] and built an ESG ratings' dataset. For the prediction task of companies' size, we obtain an employee-count dataset from Liberated Stock Trader website[4]. We split our datasets into a 70% training set and a 30% testing set.

5.2. Experiments and Results

For the prediction task of companies' environmental impact, we first perform a linear regression model using the ESG ratings dataset. Then we add the four features in the Stock2Vec embedding to the ESG ratings dataset from four dimensions (Dim1, Dim2, Dim3, and Dim4), and run the linear regression model again on the enlarged dataset. The same steps are repeated on the prediction task of companies' size. And for both tasks, we use the Ordinary Least Squares (OLS) estimator, which is a linear least squares method to estimate the unknown parameters in a linear regression model, to fit the model to the datasets.

As seen in Figure 6, when the four additional features are not added to the two datasets, the linear regression model achieves R2 values of 10.5% for the prediction of employee count, and 2.0% for the prediction of ESG rating. After adding our Stock2Vec embedding, the R2 value of the prediction of employee count increases from 10.5% to 12.4%, noting a 1.9% improvement. Similarly, the R2 value for the prediction of ESG rating rises from 2.0% to 11.3%, which is a 9.3% boost. Therefore, the linear regression model receives a drastic improvement in accuracy on the two prediction tasks with the help of our Stock2Vec embedding.

Figure 6
R2 values achieved by the regression models for the ESG rating prediction and the employee count prediction. For both tasks, the R2 values increase when the Stock2Vec embedding is included.



To confirm the importance of the Sotck2Vec embedding, we need to interpret the model and determine which features are most beneficial for the prediction tasks. Feature importance is the most useful interpretation tool to identify important features, so we use the Random Forest Importance (Strobl, 2008), which is a reliable and efficient technique and can be applied to any model, to extrapolate the significance of each feature in the OLS model for the ESG Ratings prediction task and the employee-count prediction task.

3 https://www.sustainalytics.com/esg-data
4 https://www.liberatedstocktrader.com/sp-500-companies-list-by-number-of-employees/

As shown in Figure 7, the importance score indicates the significance of each feature in the prediction making process, and the four features in our Stock2Vec embedding (Dim1-4 represent four dimensions that the four features add to) contribute the most to the ESG Ratings OLS model. Similarly in Figure 8, the Stock2Vec features also bear high importance compared with other features used in the employee-count prediction. The feature importance scores confirm that the Stock2Vec embedding plays a major role in raising the R2 values of the prediction tasks.

Figure 7
Random Forest Importance score of each feature used in the ESG Ratings OLS model.

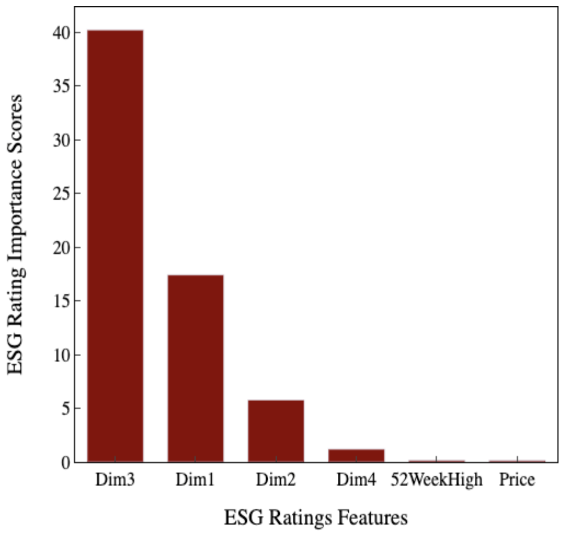
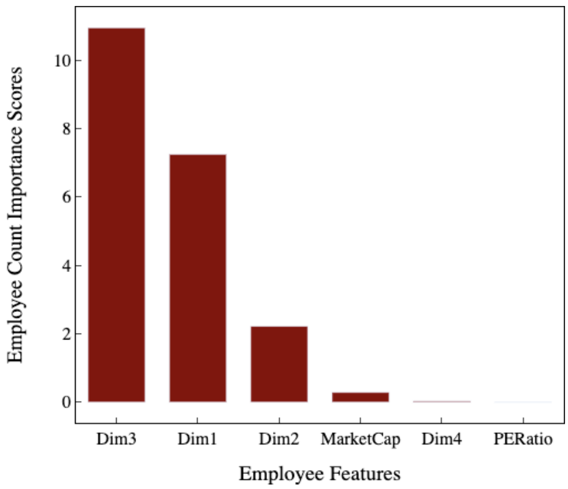


Figure 8
Random Forest Importance score of each feature used in the employee-count OLS model.



5.3. Discussion

5.3.1. Employee-Count OLS Model

We use the residuals plot to analyze the variance of error in the employee-count OLS regressor. As shown in Figure 9, the residual points are randomly dispersed around the horizontal axis, which indicates that the model is appropriate for the data.

Figure 9
Residuals for the employee-count OLS model

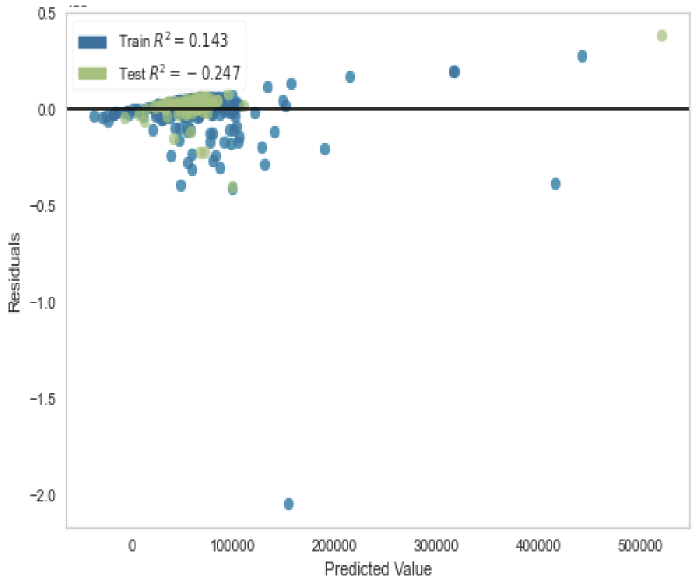


Figure 10 shows that the R-squared, which measures the accuracy of the model, is 10.5% for the employee- count model without adding the four features in the Stock2Vec embedding. And after adding the Stock2Vec features, as shown in Figure 11, the R-squared is 12.4% for the model. Additionally, the p-values of the four Stock2Vec features X3, X4, X5, X6 are lower than the p-value of the variable X2, which is 0.919. X2 is the MarketCap feature from the original employee-count dataset. Hence, the addition of the four features in the Stock2Vec embedding helps improve the model.

Figure 10
Prediction summary of the employee-count OLS model without the Stock2Vec embedding.

Dep. Variable:	NumberOfEmployees	R-squared:	0.105
Model:	OLS	Adj. R-squared:	0.100
Method:	Least Squares	F-statistic:	21.21
Date:	Tue, 02 Feb 2021	Prob (F-statistic):	1.94e-09
Time:	16:15:24	Log-Likelihood:	-4846.9
No. Observations:	366	AIC:	9700.
Df Residuals:	363	BIC:	9712.
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	4.272e+04	8254.598	5.175	0.000	2.65e+04	5.9e+04
x1	245.4177	37.685	6.512	0.000	171.309	319.527
x2	0.7379	64.785	0.011	0.991	-126.664	128.139

Omnibus:	629.039	Durbin-Watson:	0.581
Prob(Omnibus):	0.000	Jarque-Bera (JB):	303098.303
Skew:	9.783	Prob(JB):	0.00
Kurtosis:	142.615	Cond. No.	238.

Figure 11

Prediction summary of the employee-count OLS model with the Stock2Vec embedding.

```
==============================================================================
Dep. Variable:      NumberOfEmployees   R-squared:                       0.124
Model:                            OLS   Adj. R-squared:                  0.109
Method:                 Least Squares   F-statistic:                     8.447
Date:                Thu, 04 Feb 2021   Prob (F-statistic):           1.38e-08
Time:                        07:14:52   Log-Likelihood:                -4843.0
No. Observations:                 366   AIC:                             9700.
Df Residuals:                     359   BIC:                             9727.
Df Model:                           6
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const       2.943e+05   2.67e+05      1.101      0.271   -2.31e+05     8.2e+05
x1           237.0795     38.374      6.178      0.000     161.613     312.546
x2            -6.5753     64.656     -0.102      0.919    -133.728     120.577
x3         -1.631e+05   1.65e+05     -0.988      0.324   -4.88e+05    1.62e+05
x4          8.814e+04   9.55e+04      0.923      0.357   -9.96e+04    2.76e+05
x5          1.822e+05   1.94e+05      0.938      0.349      -2e+05    5.64e+05
x6          4043.7607   1.89e+04      0.214      0.831   -3.32e+04    4.13e+04
==============================================================================
Omnibus:                      633.299   Durbin-Watson:                   0.598
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           315535.382
Skew:                           9.908   Prob(JB):                         0.00
Kurtosis:                     145.472   Cond. No.                     1.11e+04
==============================================================================
```

5.3.2. ESG Ratings OLS Model

Figure 12 shows the residuals plot of the ESG ratings OLS model, which is used to analyze the variance of error in the ESG ratings OLS regressor. We find that the residual points are randomly dispersed around the horizontal axis of the predicted value, which confirms that it is a well fitted model.

Figure 12

Residuals for the ESG ratings OLS model

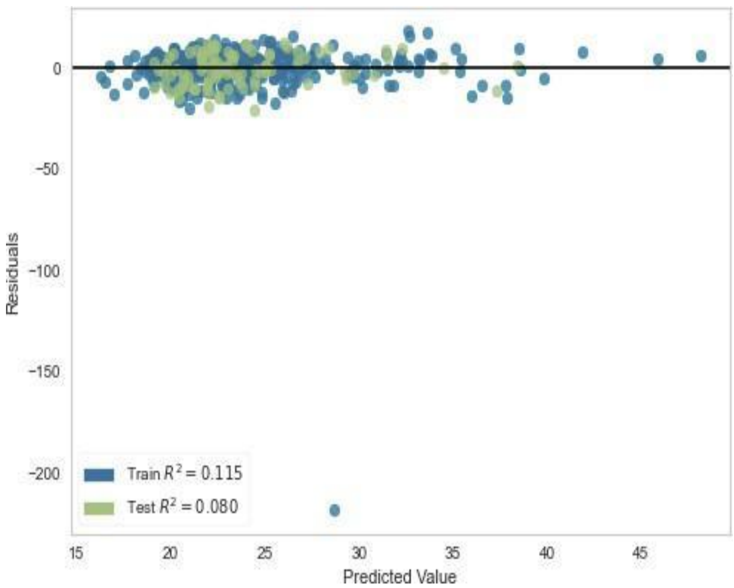


Figure 13 and Figure 14 show that the R-squared of the ESG ratings model performs better when we add the four features in the Stock2Vec embedding. Initially in Figure 13, the accuracy score is only 2.0%. After adding the four features, the accuracy immediately boosts to 11.3 % in Figure 14. Consequently, the four Stock2Vec features represented by X8, X9, X10 and X11 are statistically significant with p-value scores of 0.003, 0.035, 0.008, and 0.001 respectively. All the p-values of the four features are less than the p-value of X2, which is 0.980. This indicates that the four features have a significant impact on the ESG ratings prediction. In conclusion, the addition of the Stock2Vec embedding significantly improves the model.

Figure 13
Prediction summary of the ESG ratings OLS model without the Stock2Vec embedding.

Dep. Variable:	ESGRating	R-squared:	0.020
Model:	OLS	Adj. R-squared:	0.005
Method:	Least Squares	F-statistic:	1.317
Date:	Tue, 02 Feb 2021	Prob (F-statistic):	0.240
Time:	13:52:46	Log-Likelihood:	-1833.9
No. Observations:	461	AIC:	3684.
Df Residuals:	453	BIC:	3717.
Df Model:	7		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	24.1842	1.387	17.435	0.000	21.458	26.910
x1	-0.0175	0.047	-0.375	0.708	-0.109	0.074
x2	-0.0036	0.015	-0.235	0.815	-0.033	0.026
x3	-0.0717	0.456	-0.157	0.875	-0.968	0.824
x4	-0.2598	0.142	-1.835	0.067	-0.538	0.018
x5	-0.0241	0.045	-0.538	0.591	-0.112	0.064
x6	0.0601	0.044	1.377	0.169	-0.026	0.146
x7	1.449e-11	7.73e-12	1.875	0.061	-6.97e-13	2.97e-11

Omnibus:	821.756	Durbin-Watson:	1.969
Prob(Omnibus):	0.000	Jarque-Bera (JB):	638692.856
Skew:	10.850	Prob(JB):	0.00
Kurtosis:	184.052	Cond. No.	2.36e+11

Figure 14
Prediction summary of the ESG ratings OLS model with the Stock2Vec embedding.

Dep. Variable:	ESGRating	R-squared:	0.113
Model:	OLS	Adj. R-squared:	0.091
Method:	Least Squares	F-statistic:	5.193
Date:	Thu, 04 Feb 2021	Prob (F-statistic):	1.04e-07
Time:	08:05:21	Log-Likelihood:	-1810.9
No. Observations:	461	AIC:	3646.
Df Residuals:	449	BIC:	3695.
Df Model:	11		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-24.3845	18.063	-1.350	0.178	-59.884	11.115
x1	-0.0184	0.046	-0.396	0.692	-0.109	0.073
x2	-0.0004	0.015	-0.025	0.980	-0.029	0.028
x3	-0.7662	0.509	-1.506	0.133	-1.766	0.233
x4	-0.0406	0.140	-0.289	0.772	-0.317	0.235
x5	-0.0095	0.047	-0.204	0.839	-0.101	0.082
x6	0.0373	0.044	0.842	0.400	-0.050	0.124
x7	1.684e-11	7.56e-12	2.228	0.026	1.99e-12	3.17e-11
x8	32.8607	10.930	3.007	0.003	11.381	54.340
x9	-13.7810	6.504	-2.119	0.035	-26.563	-0.999
x10	-34.5832	12.999	-2.660	0.008	-60.130	-9.037
x11	-4.4503	1.368	-3.253	0.001	-7.139	-1.762

Omnibus:	870.489	Durbin-Watson:	2.001
Prob(Omnibus):	0.000	Jarque-Bera (JB):	881909.869
Skew:	12.192	Prob(JB):	0.00
Kurtosis:	215.881	Cond. No.	4.44e+12

6. CONCLUSION

In this paper, we present Stock2Vec, an efficient embedding to represent companies. We use two datasets and two company characteristic prediction tasks to analyze and evaluate the capability of the Stock2Vec embedding. The experiment results demonstrate that the Stock2Vec embedding can characterize companies using only their stock information, and that company characteristic prediction results can be improved by adding the four Stock2Vec features to predictive models. Moreover, the construction of the Stock2Vec embedding is fast and can be shared readily, which is less expensive than other existing methods. Our Stock2Vec embedding breaks the limitation of existing predictive models and provides insights to future studies of business such as stock market analysis and predictive models generation. In the future, we plan to include more advanced embedding techniques such as BERT to improve the prediction performance.

REFERENCES

Breiman, L. Random Forests. Machine Learning 45, 5–32 (2001). https://doi.org/10.1023/A:1010933404324

Crespi, F., & Migliavacca, M. (2020). The Determinants of ESG Rating in the Financial Industry: The Same Old Story or a Different Tale? Sustainability: Science Practice and Policy, 12(16), 6398.

Dorfleitner, G., Halbritter, G., & Nguyen, M. (2015). Measuring the level and risk of corporate responsibility – An empirical comparison of different ESG rating approaches. In Journal of Asset Management (Vol. 16, Issue 7, pp. 450–466). https://doi.org/10.1057/jam.2015.31

Grover, A., & Leskovec, J. (2016). node2vec: Scalable Feature Learning for Networks. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 855–864.

J.A. Sharp, D.H.R. Price (1982). Industry modelling, Omega, Volume 10, Issue 3, Pages 237-247, ISSN 0305-0483 Lau, J. H., & Baldwin, T. (2016). An Empirical Evaluation of doc2vec with Practical Insights into Document

Embedding Generation. In Proceedings of the 1st Workshop on Representation Learning for NLP. https://doi.org/10.18653/v1/w16-1609

Lien Minh, D., Sadeghi-Niaraki, A., Huy, H. D., Min, K., & Moon, H. (2018). Deep Learning Approach for Short- Term Stock Trends Prediction Based on Two-Stream Gated Recurrent Unit Network. IEEE Access, 6, 55392–55404.

Lu, R., & Lu, M. (2021). Stock Trend Prediction Algorithm Based on Deep Recurrent Neural Network. Wireless Communications and Mobile Computing, 2021.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. In arXiv [cs.CL]. arXiv. http://arxiv.org/abs/1301.3781

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems (pp. 3111-3119).

Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized word representations. Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers).

Rish, I., & Others. (2001). An empirical study of the naive Bayes classifier. IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence, 3, 41–46.

Rizkallah, S., Atiya, A. F., & Shaheen, S. (2020). A Polarity Capturing Sphere for Word to Vector Representation.

NATO Advanced Science Institutes Series E: Applied Sciences, 10(12), 4386.

Safavian, S. R., & Landgrebe, D. (1991). A survey of decision tree classifier methodology. IEEE Transactions on Systems, Man, and Cybernetics, 21(3), 660–674.

Behpour, S., Mohammadi, M., Albert, M. V., Alam, Z. S., Wang, L., & Xiao, T. (2021). Automatic trend detection: Time-biased document clustering. Knowledge-Based Systems, 220, 106907.

Shen, S., Jiang, H., & Zhang, T. (2012). Stock market forecasting using machine learning algorithms. Department of Electrical Engineering, Stanford University, Stanford, CA, 1-5.

Shin, H. G., Ra, I., & Choi, Y. H. (2019, October). A deep multimodal reinforcement learning system combined with CNN and LSTM for stock trading. In 2019 International conference on information and communication technology convergence (ICTC) (pp. 7-11). IEEE.

Shlens, J. (2014). A tutorial on principal component analysis. arXiv preprint arXiv:1404.1100.

Short, J. C., McKenny, A. F., Ketchen, D. J., Snow, C. C., & Hult, G. T. M. (2016). An Empirical Examination of Firm, Industry, and Temporal Effects on Corporate Social Performance. Business & Society, 55(8), 1122– 1156.

Strobl, C., & Zeileis, A. (2008). Danger: High power! – exploring the statistical properties of a test for random forest variable importance

Wang, X., Wang, Y., Weng, B., & Vinel, A. (2020). Stock2Vec: A Hybrid Deep Learning Framework for Stock Market Prediction with Representation Learning and Temporal Convolutional Network. In arXiv [q- fin.ST]. arXiv. http://arxiv.org/abs/2010.01197

Wu, Q., Brinton, C. G., Zhang, Z., Pizzoferrato, A., Liu, Z., & Cucuringu, M. (2019). Equity2Vec: End-to-end deep learning framework for cross-sectional asset pricing. In arXiv [cs.LG]. arXiv. http://arxiv.org/abs/1909.04497

Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. R., & Le, Q. V. (2019). Xlnet: Generalized autoregressive pretraining for language understanding. Advances in neural information processing systems, 32.

Yu, Y., Wang, S., Du, Y., Viswanathan, V., Su, R., Ramakrishna, S., ... & Gupta, A. K. (2019, March). Application of off-policy integral reinforcement learning for H∞, input constrained control of permanent magnet synchronous machine. In 2019 IEEE Applied Power Electronics Conference and Exposition (APEC) (pp. 2570-2576). IEEE.

Zhang, Y. (2012). Support Vector Machine Classification Algorithm and Its Application. Information Computing and Applications, 179–186.

Zumente, I., & Lāce, N. (2021). ESG Rating—Necessity for the Investor or the Company? Sustainability: Science Practice and Policy, 13(16), 8940.

Application of Big Data Analytics in Precision Medicine: Lesson for Ethiopia

MISGANAW TADESSE WOLDEMARIAM*
Information Science, Jimma University, Jimma
Jimma, Jimma Zone, Ethiopia
E-mail: misganawt2012@gmail..com

DANIEL ALEMNEH†
University of NT, Denton
Denton, Texas 76201, USA
†E-Mail: Daniel.Alemneh@unt.edu

Precision medicine is an emerging approach for disease treatment and prevention that considers individual variability in genes, environment, and lifestyle for each person. Big data analytics (BDA) using cutting-edge technologies helps to design models that can diagnose, treat and predict diseases. In Ethiopia, healthcare service delivery faces many challenges specifically in relation to prescribing the right medicine to the right patient at the right time. Thus, patients face challenges ranging from staying on treatment plans longer, and then leaving treatment, and finally dying of complications. Therefore, the aim of this paper is to explore the trends, challenges, and opportunities of applying BDA in precision medicine globally and take lessons for Ethiopia through a systematic literature review of 19 peer reviewed articles from five databases. The findings indicated that cancer in general, epilepsy, and systemic diseases altogether are areas currently getting big attention. The challenges are attributed to the nature of health data, failure in collaboration for data sharing, ethical and legal issues, interoperability of systems, poor knowledge skills and culture, and poor infrastructure. Development of modern technologies, experimental technologies and methods, cloud computing, Internet of Things, social networks and Ethiopia's government initiative to promote private technological firms could be an opportunity to use BDA for precision medicine in Ethiopia.

Key words: precision medicine, big data, big data analytics, Ethiopia

1. Introduction

1.1. Background

Precision medicine is not a one-size-fits-all approach, but a new area of disease treatment and prevention in which patients receive personalized medicine (Schaefer et al., 2019). It aims to combine comprehensive data (big data) collected over time about human genetics, environment and lifestyle to improve understanding and discovery of disease, support drug development, and provide appropriate therapies (Huang et al., 2016).

According to the Precision Medicine Initiative, precision medicine is an emerging approach for disease treatment and prevention that considers individual variability in genes, environment, and lifestyle for each person. This approach will help doctors and researchers predict more accurately which treatment and prevention strategies for a particular disease will work in which groups of people (United States National Library of Medicine, 2020).

Although precision medicine may appear to be a radical and innovative concept, it was first described around 2000 years ago in "The Yellow Emperor's Canon of Internal Medicine" (Roda et al., 2017). Precision medicine stands in contrary to the "one size fits all" concept, which expects researchers to develop disease treatment and prevention strategies for an "average" patient without much consideration for inter-individual variances (Roda et al., 2017). Unfortunately, when treating according to standardized therapeutics protocols, a given drug may be ineffective or inappropriate for a high proportion of patients, ranging from 40% to 70% or higher (Roda et al., 2017).

Medicines contributes to the improvement of quality of life and life expectancy by relieving symptoms, delaying disease progression, and curing diseases (Kefale et al., 2020). However, no drug is entirely harmless and can be associated with emergency department visits, hospitalization, in-patient, and outpatient care complications (Kefale et al., 2020). Medication Related Problems (MRPs) are unwanted effects that actually or potentially interfere with health outcomes. They are significant causes of patient morbidity, mortality, economic loss, and contribute to overall pressure on the healthcare system (Kefale et al., 2020). MRPs include medication errors, adverse drug events, and adverse drug reactions (ADRs). The findings of Kefale et al. (2020) indicated that, almost half of the study participants had indication-related MRPs, while effectiveness and safety-related MRPs occurred among one in four patients. In addition to this, different socioeconomic, disease-related, medication-related, and healthcare-related variables contribute to the development of MRPs and ADRs.

Ethiopian government has developed health sector development plan (HSDP-I to HSDP-IV) and health sector transformational plan (HSTP) recently in order to mitigate challenges observed in healthcare service quality and access (Federal Ministry of Health [FMOH], 2015). Moreover, government and non-government supported initiatives are playing a significant role in adopting ehealth services like, mobile health, telemedicine, Electronic Health Record (EHR), Electronic Health Management Information System (eHMIS) etc., to excel service delivery which in turn generates a potential amount of electronic health record (FMOH, 2014). This shows a possible need for a new approach which can leverage the existing trends in a vast collection of data and get knowledge for quality clinical decision making and prescribing the right medicine for the right patient at the right time. Therefore, this issue can be solved by employing Big Data Analytics (BDA) on abundant health data across the country to reach for the anticipated better healthcare service delivery.

BDA is often a complex process of examining big data for information such as hidden patterns, correlations, market trends, and customer preferences that can help organizations make informed business decisions (Chai et al., 2021). Akal et al. (2019) defined BDA as the process of capturing and storing huge volumes of data which have different formats and are generated in high velocity. It also refers to the process of analyzing big data for the purpose of decision making, strategic planning, and policy formulation (Akal et al., 2019). BDA in healthcare involves analyzing large amounts of electronic data related to the health and well-being of patients. This data is so diverse that it is difficult to measure with conventional software or hardware (Galetsi et al., 2020). Therefore, such huge data can be easily analyzed and used for decision making purpose through applying BDA (Chai et al., 2021).

1.2. Problem Statement

In Ethiopia, BDA concepts and their implementation are either not known or left as insignificant in both government and non-government data intensive work environments. A case study conducted in four big data generating industries in Ethiopia namely, Ethiopian Telecommunication Corporation, Agricultural Transformation Agency, Payment systems, and Ethiopian educational networks indicated that, they are not using BDA for decision making due to lack of BDA awareness, data integration challenges, lack of skilled expertise in the areas, etc. (Akal et al., 2019).

Regarding precision medicine in Ethiopia, there is no research work done indicating the application of BDA in Ethiopian health sector in reaching for precision medicine or other healthcare service. According to WHO report, Ethiopia is among the countries which do not have a policy for the implementation of BDA and also BDA is not yet implemented in the health sector (WHO, 2016).

Developing countries like Ethiopia face challenges in providing quality medical diagnosis and treatment to patients. Gutema et al. (2018) and Gebretekle et al. (2018) noted that inability of medical professionals in prescribing precise medication for their patient is an indication of some of the challenges observed on a daily basis. The outcomes due to such medical errors could be fatal or sometimes result in an elongated medical attention which is uneconomical and may lead the patient to dropout from the treatment plan altogether (Endalemaw et al., 2020).

The aim of this study is to investigate the current trends, challenges and opportunities of applying BDA in precision medicine globally and take lessons for Ethiopia through literature review of peer reviewed articles. This review is organized around the following research questions:

RQ1: What are the current trends in excelling the quality of medication using big data analytics?

RQ2: What challenges are observed in employing big data analytics for precision medicine?

RQ3: What opportunities are available for the successful implementation of big data analytics for precision medicine in Ethiopia?

2. Methodology

The literature review is organized with the main goal of excavating application of BDA in developed and developing countries focusing on its implementation, challenges and opportunities in order to identify concepts, tools, policies, and methods significant for the proper implementation of BDA in Ethiopia. For this particular study, systematic literature review (SLR) was employed to search for peer-reviewed journal articles from reputable sources. As Xiao & Watson (2017) presented, systematic literature review can be conducted within the framework of the following eight steps. These are (1) formulating the research problem; (2) developing and validating the review protocol; (3) searching the literature; (4) screening for inclusion; (5) assessing quality; (6) extracting data; (7) analyzing and synthesizing data; and (8) reporting the findings. The steps are pictorially presented in Figure 1.

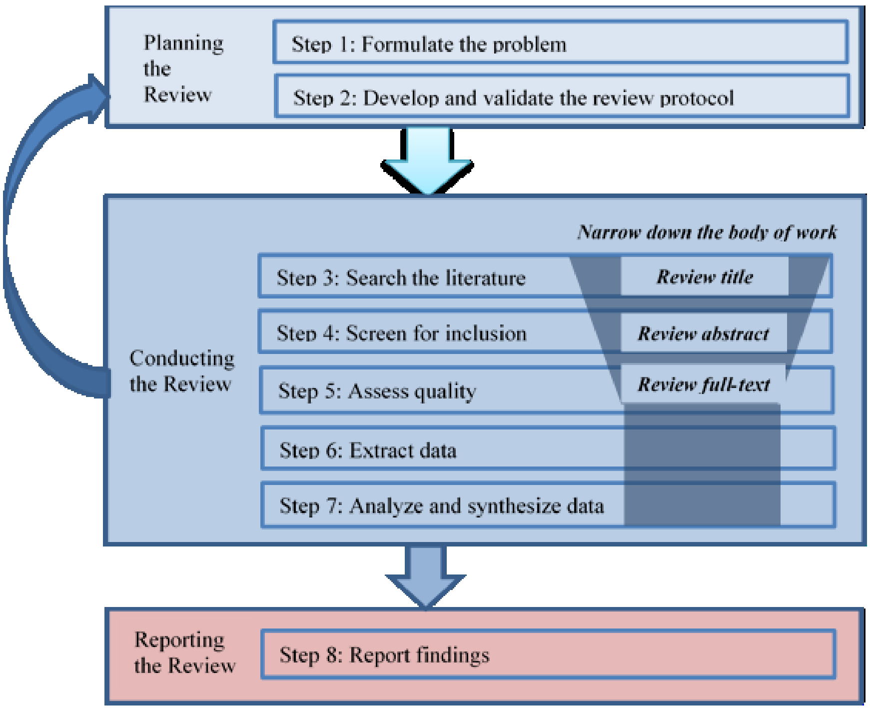


Fig. 1. Process of Systematic Literature Review

2.1. Planning the Review

Step 1: Formulate the Problem

The problem is well articulated and presented in (section 1.2.), which clearly presented the challenges in prescribing medicines in developing countries like Ethiopia, and this can be solved through the emerging technologies like BDA.

Step 2: Develop and Validate the review protocol

In this phase a review protocol is developed to systematically execute the review process. Thus, the review protocol was developed based the trends followed by different scholars to develop review protocol (Haghi Kashani et al., 2021; Mengist et al., 2020). Accordingly, we have prepared a protocol consisting of, search question, inclusion and exclusion criteria, database to be searched, search strategy, data extraction and analysis, declaration of interest and time frame.

Regarding search questions, as articulated under the section problem statement, the research is governed by three research questions with the overall goal of finding literatures dictating about implementation of BDA in precision medicine. Specifically, the questions addressed, implementation of BDA in precision medicine in the developed and developing world, identifying challenges and opportunities of BDA in precision medicine, and exploring opportunities to implement BDA for precision medicine in Ethiopia. In terms of inclusion and exclusion criteria, as presented in Haghi Kashani et al. (2021), criteria were set to pick significant articles for the identified problem. Table 1 presents the inclusion and exclusion criteria used for this literature review.

Table 1. Inclusion and Exclusion Criteria

Criteria	
Inclusion	1. Research article that presents challenges, opportunities, and practices of BDA in precision medicine 2. Research articles published from 2015 to 2022 3. JCR indexed journal articles 4. Peer-reviewed journal articles
Exclusion	1. Research articles that do not explicitly discuss about BDA and precision medicine 2. Books, book chapters, conference papers, symposiums, and non- English scripts 3. Short articles (less than six papers) 4. Commentaries or review articles

In relation to the databases to be searched, five electronic databases namely, Emerald, IEEE Xplore, PubMed, Science Direct, and Taylor and Francis were used based on their popularity in the area of interest.

In terms of the search strategy, keywords representing research questions were used to search articles which met the inclusion criteria. Whereas, data extraction was done through intensive reading of the articles and recording the findings in a sheet prepared for this particular purpose, analysis was done by putting ideas into already identified four themes.

The authors of this article declare that they have no conflict of interests.

2.2. Conducting the Review

Step 3: Search the Literature

Literature search is conducted using key words, "*big data analytics", "precision medicine"* and *"big data analytics and precision medicine"* on five identified electronic databases based on article's title. Titles were used to support the searching process and avoids the extra time to be wasted in reading the entire content. The search results were exported to an excel sheet for further understanding and support screening as shown in Table 2.

Step 4: Screen for Inclusion

In this SLR, article screening is done by manually reading the abstract and looking for association in terms of the research questions. Thus, articles matched with the inclusion criteria and having proximity to the research topic are selected to undergo the SLR. Accordingly, based on the proximity to the research topic and criteria for choosing each paper, 19 published articles which correlated with the criteria were selected to undergo an in-depth review. The summary of selected articles is presented in Table 2.

Table 1. Inclusion and Exclusion Criteria

Database	Search Result	Selected Articles
Emerald	23	-
IEEE Xplore	7	3
PubMed	47	13
Science Direct	38	3
Taylor and Francis	314	-

Step 5: Assess Quality

The authors have critically reviewed each of the articles in the final list to confirm quality. Accordingly, all of the 19 articles have strong relationship with the research topic and have good quality.

Step 6: Extract Data

The data from each of the articles was extracted through a rigorous review and organized in a form of annotated bibliography. Thus, a bibliography for all of the 19 articles was prepared to support the subsequent analysis and synthesis.

Step 7: Analyze and Synthesize Data

The data organized in annotated bibliography was analyzed based on the main goal of the SLR. After getting the concept behind each of the selected articles and considering the research questions, data was synthesized to build up sound and significant answers for each question and overall achieve the aim of the literature review.

2.3. Reporting the Review

Step 8: Report Findings

The findings of this systematic literature review are presented in the discussions and conclusion sections of this paper. Further, the authors made recommendations for all the concerned bodies for the proper implementation of BDA in precision medicine based on the findings of the paper.

3. Literature Review

The literature review consists of four sections, namely: overview of precision medicine, trends of BDA in precision medicine, challenges and opportunities of applying BDA in precision medicine. Further, discussion, conclusion and recommendations are included based on the findings of the SLR.

3.1. Overview of Precision Medicine

Precision medicine in its broad sense is not a new approach, where medical doctors have been delivering the usual therapies and preventative care that will best suit the particular patient considering all relevant individual-oriented factors (Schaefer et al., 2019). Though, medical diagnosis and treatment capabilities are growing, the need for a better care and treatment is always a question for humankind. Precision medicine promotes transformation in healthcare, medical decisions, treatments, practices and products customized to a subgroup of patients based on understanding of individual genes, environment and lifestyle (Naithani et al., 2021).

Precision medicine which was announced by the former president of the United States, Barack Obama in 2015, has been identified as a promising advancement to impact automated medical decision making (Ghitza, 2015). It aims to establish a new research framework accelerating the innovation of the right treatment, for the right patient, at the right time, (Sadat Mosavi & Filipe Santos, 2020). The core of precision medicine is the analysis, identification and verification of biomarkers for a large sample population and specific disease types with the help of cutting-edge technologies such as genome and proteome determination, so as to accurately find the causes of diseases and therapeutic targets, (Vargas & Harris, 2016). The identification of biomarkers lays a necessary foundation for precise medicine.

Naithani et al. (2021), also noted that precision medicine has a wide range of applications in both diagnostic and therapeutic fields which include: development of new omics signature-based biomarkers; development of pharmacotherapy based on pharmaco-omics; precision medicine in oncology; precision medicine in chronic disorders; precision medicine in emergency care; and precision medicine in infectious diseases.

Precision medicine is challenged by its nature of data-intensiveness; hence it requires large volumes of data which needs to be collected and analyzed which is not only costly but also needs labor and technical knowhow. In addition to this, data anonymization, security, ethical challenges, noise, and turnaround time of data analysis are critical issues to be considered during implementation of precision medicine, (Naithani et al., 2021). The ultimate goal of precision medicine is to provide rational pharmacotherapy, i.e. to prescribe the right drug to the right patient in the right dose at the right time avoiding harm to the patient (Naithani et al., 2021).

3.2. Trends of BDA in Precision Medicine

The concept of big data refers to sets of data with a scale and complexity that enforces the use of dedicated analytical and statistical approaches, though exact and universally accepted definition does not exist (Alonso et al., 2017). Whereas, in the case of biomedicine, big data include large-volume and highly diverse biological, genetic, clinical, environmental and lifestyle information collected from single individuals as well as large cohorts in relation to their disease and/or wellness status at one or several time points (Auffray et al., 2016). As Wu et al. (2017) noted, big data in healthcare constitutes, large omics data collected through next generation sequencing technology and mass spectrometry, and electronic health record (EHR) which is collected through patient-doctor interaction in healthcare institutions.

Big data in healthcare and medicine refers to various large and complex data, which are difficult to analyze and manage with traditional software or hardware. BDA covers integration of heterogeneous data, data quality control, analysis, modeling, interpretation and validation, (Ristevski & Chen, 2018). Application of BDA generates comprehensive knowledge discovered from the available huge amount of data which provides comprehensive benefits to the patients, clinicians, and health policy makers (Ristevski & Chen, 2018).

Ristevski and Chen (2018) noted that, applications of BDA can enhance the patient-based service to detect spreading diseases earlier, generate new insights into disease mechanisms, monitor the quality of the medical and healthcare institutions as well as provide better treatment methods. BDA can be used for treating different types of diseases from infectious to complications like cancer. As Gameiro et al. (2018) noted, cancer and psychiatry have the lead in using precision medicine to provide tailored medicine for patients based on their unique characteristics.

According to Shah and Masters (2020), lung cancer is the leading cause of cancer death in men and women in the United States and more than 80% of these patients have non-small cell lung cancer. The authors mentioned that, in recent days due to advance in technology, immunotherapy has made possible to treat patients even those whose cancer stage reached stage IV.

In a study conducted by Prabhakaran et al. (2020), there is a clear finding indicating a critical progress within the advancement of precision medicine for colorectal carcinoma, impacting areas of screening, treatment, and potentially prevention. Advancement in molecular techniques have made it possible for far better selection of patients for treatments and it is vital that mutational investigation is performed at the time of diagnosis to guide treatment. Striano and Minassian (2020) noted that, due to advances in next generation sequencing, the previously known medications for epilepsy which are anti-seizure drugs are being replaced by personalized therapeutic approaches (precision medicine). Hence, the genetic tests could forward directions for the genetic causes of several epileptic conditions which have been unveiled and remarkably improved our knowledge on the pathogenesis of epileptogenesis.

Pinker et al. (2018) noted that, radiogenomics can be used to identify breast tumor status using genetic and radiomic data by correlating with individual’s life style and environmental variables to aid the treatment and prevention of breast cancer. In a study conducted by Song et al. (2020), precision medicine is used to treat diseases of motor system, circulatory system, respiratory system, urinary system, endocrine system, nervous system, and reproductive system as well as cancer. Moreover, precision medicine sets a conducive environment for efficient drug development and reduces the cost of development.

3.3. Challenges of Applying BDA in Precision Medicine

The health data by itself creates challenges for the success of precision medicine. Some of these challenges hindering BDA are collecting large amount of data, heterogeneity of data sources, noise of experimental omics data, variety of experimental techniques, environmental conditions, and biological nature (Ristevski & Chen, 2018). The issues indicated above should be carefully analyzed before combining such heterogeneous data and before employing machine learning methods.

These inadequacies might lead to the lack of quality of some of the data points, such as missing values or outliers. In spite of these downsides of the omics data (i.e. genomics, epigenomics, microbiomics, lipidomics, proteomics, transcriptomics etc.), EHRs data are exceptionally impacted by the staff who entered the patient’s information, which can lead to entering missing values, inaccurate data as a result of mistakes, misconception or wrong understanding of the original data (Wu et al., 2017).

Pastorino et al. (2019) indicted that, challenges in BDA are altogether credited to failure in collaboration for data sharing, ethical and legal issues, data heterogeneity, data security, analytical flows in analyzing data and lack of suitable infrastructure for data storage and computation. One of the biggest challenges to carry out precision medicine experiments is providing adequate protection for private health data while sharing data among stakeholders (hospitals, research organizations, pharmaceutical companies, insurance companies, and ownership of cloud servers) (Sethu et al., 2020). Precision medicine is also challenged by heterogeneous data integration and joint processing to support new knowledge discovery. In significance to precision medicine advances, genomic analysis and molecular profiling are a computationally intensive task (Panayides et al., 2019).

3.4. Opportunities of Applying BDA in Precision Medicine

Rapid development of the emerging information technologies, experimental technologies and methods, cloud computing, Internet of Things, social networks supply large volume of generated data which is growing tremendously in numerous research fields, (Ristevski & Chen, 2018). Innovation of new platforms, tools and methodologies for storing and structuring data has made collecting data from EHRs, social media, patient summaries, genomic and pharmaceutical data, clinical trials, telemedicine, mobile apps, sensors and information on well-being, behavior and socio-economic indicators to be easier. Healthcare professionals can benefit from such Big Data immensely (Pastorino et al., 2019).

Although some of the hardest challenges for computing systems are focused on extreme data analytics and data-intensive simulations such as streaming data analysis and virtual patient design; machine learning on high dimensional data represents a prevalent concern. As a result, advanced machine learning methods such as deep learning and platforms for cognitive computing represent the future toolbox for data-driven analysis of biomedical big data (Cirillo & Valencia, 2019).

According to Song et al. (2020), in the future, with the decrease of gene sequencing cost, the application of metagenomic sequencing and the progress of molecular pathology epidemiology will boost the development of precision medicine. In cognizant to the above, investment over new technology and human capital are essential for the successful implementation of BDA in precision medicine, (Pastorino et al., 2019).

4. Discussion

In our study, the SLR showed insights on the current status of precision medicine, the challenges faced by the developed countries while implementing BDA and associated opportunities that would be exploited to further grow a rich BDA in the area of precision medicine. Many scholars presented, precision medicine is being utilized in the areas of diagnosis, treatment and prediction of disease outcomes, (Naithani et al., 2021; Schaefer et al., 2019; Vargas & Harris, 2016). Nowadays, precision medicine is used in treating diseases ranging from infectious to chronic diseases, cancer, mental disease etc. (Gameiro et al., 2018). Developed nations are taking advantage of Big Data to solve problems which were impossible to manage using human cognitive skill. Developing countries are also looking forward to exploit big data analytics to alter health related issues.

Many of the challenges in implementing precision medicine are associated with the data itself. Hence, big data comprises data which is characterized by: volume, value, velocity, variety, veracity, and variability. Data quality is the first challenge which might lead to making wrong clinical decisions if not taken with great care (Wu et al., 2017). Integrity of data is another challenge whenever sharing and storing data (Pastorino et al., 2019).

Infrastructure which is used for capturing, storing, analyzing and sharing big data demands a big budget and seems challenging for developing or under developed countries. Initiatives in developed countries are building platforms for storing, processing and sharing data. They are also providing a software solution to manage data in distributed manner. This trend can also be used by developing countries like Ethiopia to exploit BDA so as to leverage its benefits.

Human capital is very essential for the successful implementation of BDA in precision medicine. It all begins from equipping physicians with the knowledge of recording the correct patient data in the right format into the right database and extends to applying appropriate analysis techniques to generate knowledge by the data scientists. Therefore, trainings for every stakeholder like, physicians, data scientists, IT professionals, etc. to fill the knowledge and skill gap could potentially affect the effective implementation of BDA in developing countries.

BDA needs a policy to guide its implementation and to provide the necessary support expected from each stakeholder. In the case of Ethiopia, Big Data analytics has no policy for its implementation and administration (WHO, 2016). Moreover, advancement in computing, high band-width data transmission, cloud computing, sophisticated machine learning methods and the like will be potential enablers for the growth of BDA in developed and developing countries (Ristevski & Chen, 2018).

In Ethiopia, the government's withdrawal of a law which has been disallowing private telecommunication firms to invest and operate has brought opportunities to attract global telecom firms to invest and build their own network infrastructure which could enhance the existing Ethio-Telecom network facilities. This would potentially help the health sector to build a national data repository on a cloud-based storage; systems capable of collecting and storing data across the country; accessing global databases for different types of data for research and other purposes; and promote data and knowledge sharing among institutions and many more. Further, the fast-growing health data and global push factor to use BDA would be the opportunities for implementing big data analytics specially to support the attempt to provide tailored clinical services (precision medicine) to the people of Ethiopia. The infancy or total absence of BDA in the health sector of Ethiopia would be both a challenge and an opportunity. The challenge is, it requires a lot of budget to setup the infrastructure (hardware and software) needed for the successful implementation of BDA. Whereas, all the challenges learned from countries which already built and using BDA based systems would be an opportunity to design and develop such kind of systems at relatively lower cost and with better understanding of the problem domain. Further, policies, strategies and methods can be easily adopted from the pioneers and even through collaborations technical support, knowledge, skill transfer and resource support could be obtained.

5. Conclusion

Precision medicine is an emerging area for disease treatment and prevention whereby a patient receives personalized medicine in contrast to a one size fits all approach. It aims to combine big data collected over time about an individual's genetics, environment, and lifestyle, to advance disease understanding and interception, aid drug discovery, and ensure delivery of appropriate therapies. The core of precision medicine is the analysis, identification and verification of biomarkers for a large sample population and specific disease types with the help of cutting-edge technologies such as genome and proteome determination, so as to accurately find the causes of diseases and therapeutic targets. The identification of biomarkers lays a necessary foundation for precision medicine.

BDA can enhance patient-based services, to detect spreading diseases earlier, generate new insights into disease mechanisms, monitor the quality of the medical and healthcare institutions as well as provide better treatment methods. These days, BDA is being used for treating different types of diseases ranging from infectious to complicated like, cancer and psychiatry. Many scholars indicated that BDA is being used for treating lung cancer, colorectal cancer, epilepsy, breast cancer, and in general systemic diseases of motor, circulatory, respiratory, urinary, endocrine, nervous, and reproductive systems.

The major challenge in applying BDA in precision medicine is attributed to healthcare data. Some of the challenges of healthcare data are collecting large amount of data; heterogeneity of data sources; noise; variety of experimental techniques; and environmental conditions and biological nature of healthcare data. Further, failure in collaboration for data sharing, ethical and legal issues, data protection, analytical flows in analyzing data and lack of appropriate infrastructure for data storage and computation poses a challenge when applying BDA for precision medicine.

Rapid development of the emerging information technologies, experimental technologies and methods, cloud computing, Internet of Things, social networks supply large volume of generated data which is growing tremendously in numerous research fields. Innovation of new platforms, tools and methodologies for storing and structuring data has made collecting data from EHRs, social media, patient summaries, genomic and pharmaceutical data, clinical trials, telemedicine, mobile apps, sensors and information on well-being, behavior, and socio-economic indicators is becoming easier. Advanced machine learning methods such as deep learning and platforms for cognitive computing represent the future toolbox for data-driven analysis of biomedical big data. In the future, with the decrease of gene sequencing costs, the application of metagenomic sequencing and the progress of molecular pathology epidemiology will boost the development of precision medicine.

In Ethiopia, the government's withdrawal of a law which has been disallowing private telecommunication firms to invest and operate has brought opportunities to attract global telecom firms to invest and build their own network infrastructures which could enhance the existing Ethio-Telecom network facilities. Further, the fast-growing health data and global push factor to use BDA would be the opportunities for implementing BDA in Ethiopia specifically to support the attempt to provide tailored clinical services (precision medicine).13

6. Recommendations

Based on the major findings identified from the literature review the following recommendations are forwarded for concerned bodies.

- The government of Ethiopia should develop a policy framework to govern big data in healthcare institutions with respect to using it for building automated clinical decision support systems using BDA.
- Infrastructure is key for the successful implementation of BDA in any sector. Therefore, government should give due attention to equip and maintain hardware and software resources for health institutions.
- The Federal Ministry of Health should collaborate with countries having better experience in using health data for decision support purposes.
- The Federal Ministry of Health should prepare standard guidelines and frameworks to collect, store and share health data across health institutions for research and decision support purposes.
- Doctors, nurses, and data clerks should record and store health data correctly thinking that data is significant for future data-driven activities.
- Researchers and data scientists should give due attention to work on healthcare big data to solve public health problems in terms of disease prevention, treatment and prediction.
- Government and private healthcare institutions and research centers should plan and implement big data analytics strategies to enhance service delivery which brings quality patient care and treatment outcomes.

References

Akal, T. D., Beshah, T., Sackmann, S., & Negash, S. (2019). Challenges of identifying and utilizing big data analytics in a resource-constrained environment: In the case of Ethiopia. *IFIP Advances in Information and Communication Technology*, 234–254. https://doi.org/10.1007/978-3-030-20671-0_17

Alalawneh, A. A., & Alkhatib, S. F. (2020). The barriers to big data adoption in developing economies. *The Electronic Journal of Information Systems in Developing Countries*, *87*(1). https://doi.org/10.1002/isd2.12151

Alonso, S. G., de la Torre Díez, I., Rodrigues, J. J., Hamrioui, S., & López-Coronado, M. (2017). A systematic review of techniques and sources of big data in the healthcare sector. *Journal of Medical Systems*, *41*(11). https://doi.org/10.1007/s10916-017-0832-2

Auffray, C., Balling, R., Barroso, I., Bencze, L., Benson, M., Bergeron, J., Bernal-Delgado, E., Blomberg, N., Bock, C., Conesa, A., Del Signore, S., Delogne, C., Devilee, P., Di Meglio, A., Eijkemans, M., Flicek, P., Graf, N., Grimm, V., Guchelaar, H.-J., … Zanetti, G. (2016). Making sense of big data in health research: Towards an EU action plan. *Genome Medicine*, *8*(1). https://doi.org/10.1186/s13073-016-0323-y

Chai, W., Labbe, M., & Stedman, C. (2021, December 14). *What is Big Data Analytics and why is it important?* Search Business Analytics. Retrieved April 26, 2022, from https://www.techtarget.com/searchbusinessanalytics/definition/big-data-analytics

Cirillo, D., & Valencia, A. (2019). Big Data Analytics for Personalized Medicine. *Current Opinion in Biotechnology*, *58*, 161–167. https://doi.org/10.1016/j.copbio.2019.03.004

Endalamaw, A., Dessie, G., Biresaw, H., Belachew, A., Amare, D., Workineh, Y., Birhanu, M., Belay, G. M., Wondim, A., Wubneh, C. A., Birhanu, Y., Asmare, Y., Wagnew, F., Minuye, B., Kefale, D., Nigussie, Z., & Habtewold, T. D. (2020). Medication errors in Ethiopia: Systematic review and meta-analysis. https://doi.org/10.21203/rs.3.rs-35808/v1

FMOH. (2014). *Ethiopian National eHealth Strategic.* Ethiopian National e-Health Strategy. Retrieved Feb 20, 2022, from https://www.who.int/goe/policies/ethiopian_ehealth_strategy2014.pdf?ua=1

FMOH. (2015). Ethiopia Health Sector Transformation Plan (2015/16-2019/20). Retrieved Feb 20, 2022, from https://www.globalfinancingfacility.org/ethiopia-health-sector-transformation-plan-201516-201920

Galetsi, P., Katsaliaki, K., & Kumar, S. (2020). Big Data Analytics in health sector: Theoretical Framework, techniques and prospects.International *Journal of Information Management, 50*, 206–216. https://doi.org/10.1016/j.ijinfomgt.2019.05.003

Gameiro, G. R., Sinkunas, V., Liguori, G., & Auler-Júnior, J. O. (2018). Precision medicine: Changing the way we think about Healthcare. Clinics, 73. https://doi.org/10.6061/clinics/2017/e723

Gebretekle, G. B., Haile Mariam, D., Abebe, W., Amogne, W., Tenna, A., Fenta, T. G., Libman, M., Yansouni, C. P., & Semret, M. (2018). Opportunities and barriers to implementing antibiotic stewardship in low and middle-income countries: Lessons from a mixed-methods study in a tertiary care hospital in Ethiopia. *PLOS ONE, 13*(12). https://doi.org/10.1371/journal.pone.0208447

Ghitza, U. E. (2015). A commentary on "A new initiative on precision Medicine". *Frontiers in Psychiatry, 6*. https://doi.org/10.3389/fpsyt.2015.00088

Gutema, G., Håkonsen, H., Engidawork, E., & Toverud, E.-L. (2018). Multiple challenges of antibiotic use in a large hospital in Ethiopia – a ward-specific study showing high rates of hospital-acquired infections and ineffective prophylaxis. *BMC Health Services Research, 18*(1). https://doi.org/10.1186/s12913-018-3107-9

Haghi Kashani, M., Madanipour, M., Nikravan, M., Asghari, P., & Mahdipour, E. (2021). A systematic review of IOT in Healthcare: Applications, techniques, and Trends. *Journal of Network and Computer Applications, 192*, 103164. https://doi.org/10.1016/j.jnca.2021.103164

Huang, B. E., Mulyasasmita, W., & Rajagopal, G. (2016). The path from Big Data to Precision Medicine. *Expert Review of Precision Medicine and Drug Development, 1*(2), 129–143. https://doi.org/10.1080/23808993.2016.1157686

Kefale, B., Degu, A., & Tegegne, G. T. (2020). Medication-related problems and adverse drug reactions in Ethiopia: A systematic review. *Pharmacology Research & Perspectives, 8*(5). https://doi.org/10.1002/prp2.64115

Mengist, W., Soromessa, T., & Legese, G. (2020). Method for conducting systematic literature review and meta-analysis for Environmental Science Research. *MethodsX*, *7*, 100777. https://doi.org/10.1016/j.mex.2019.100777

Naithani, N., Atal, A. T., Tilak, T. V. S. V. G. K., Vasudevan, B., Misra, P., & Sinha, S. (2021). Precision medicine: Uses and challenges. *Medical Journal Armed Forces India*, *77*(3), 258–265. https://doi.org/10.1016/j.mjafi.2021.06.020

Panayides, A. S., Pattichis, M. S., Leandrou, S., Pitris, C., Constantinidou, A., & Pattichis, C. S. (2019). Radiogenomics for precision medicine with a big data analytics perspective. *IEEE Journal of Biomedical and Health Informatics*, *23*(5), 2063–2079. https://doi.org/10.1109/jbhi.2018.2879381

Pastorino, R., De Vito, C., Migliara, G., Glocker, K., Binenbaum, I., Ricciardi, W., & Boccia, S. (2019). Benefits and challenges of Big Data in Healthcare: An overview of the European initiatives. *European Journal of Public Health*, *29*(Supplement_3), 23–27. https://doi.org/10.1093/eurpub/ckz168

Pinker, K., Chin, J., Melsaether, A. N., Morris, E. A., & Moy, L. (2018). Precision Medicine and radiogenomics in breast cancer: New approaches toward diagnosis and treatment. *Radiology*, *287*(3), 732–747. https://doi.org/10.1148/radiol.2018172171

Prabhakaran, S., Leong, J., Petrelli, N. J., & Khatri, V. P. (2020). Precision medicine in colorectal surgery. *Surgical Oncology Clinics of North America, 29*(1), 23–34. https://doi.org/10.1016/j.soc.2019.09.001

Ristevski, B., & Chen, M. (2018). Big Data Analytics in medicine and Healthcare. *Journal of Integrative Bioinformatics, 15*(3). https://doi.org/10.1515/jib-2017-0030

Roda, A., Michelini, E., Caliceti, C., Guardigli, M., Mirasoli, M., & Simoni, P. (2017). Advanced bioanalytics for precision medicine. *Analytical and Bioanalytical Chemistry, 410*(3), 669–677. https://doi.org/10.1007/s00216-017-0660-8

Sadat Mosavi, N., & Filipe Santos, M. (2020). How prescriptive analytics influences decision making in Precision Medicine. *Procedia Computer Science, 177*, 528–533. https://doi.org/10.1016/j.procs.2020.10.073

Schaefer, G. O., Tai, E. S., & Sun, S. (2019). Precision Medicine and Big Data. *Asian Bioethics Review, 11*(3), 275–288. https://doi.org/10.1007/s41649-019-00094-2

Sethu, S. G., Nair, R. S., & Sadath, L. (2020). Big Data in precision medicine and its legal implications. 2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS). https://doi.org/10.1109/iciis51140.2020.9342723

U.S. National Library of Medicine. (2020, September 17). Precision medicine: Medlineplus genetics. *MedlinePlus*. Retrieved Feb 17, 2022, from https://medlineplus.gov/genetics/understanding/precisionmedicine

Vargas, A. J., & Harris, C. C. (2016). Biomarker development in the Precision Medicine Era: Lung Cancer as a case study. *Nature Reviews Cancer, 16*(8), 525–537. https://doi.org/10.1038/nrc.2016.56

WHO. (2016). Who | Ethiopia: Data and statistics. Country Context Indicators. Retrieved Feb 10, 2022, from https://www.who.int/gho/countries/eth/data/en/

Wu, P. Y., Cheng, C. W., Kaddi, C. D., Venugopalan, J., & Hoffman, R. (2017). –OMIC and electronic health record Big Data Analytics for Precision Medicine. *IEEE Transactions on Biomedical Engineering, 64*(2), 263–273. https://doi.org/10.1109/tbme.2016.2573285

Xiao, Y., & Watson, M. (2017). Guidance on conducting a systematic literature review. *Journal of Planning Education and Research, 39*(1), 93–112. https://doi.org/10.1177/0739456x17723971

Track – Information Engineering

Using Data Visualization Tools to Mitigate the Influx of Information in Organizations

Tereza Raquel Merlo
University of North Texas, Adjunct Professor,
tereza.merlo@unt.edu,
0000-0002-2042-5415

SUMMARY

Considerable research has been conducted on the topic of information overload using different approaches, from marketing and customer demand to information technologies and sciences, and even among mental health professionals. In business the critical question is how does information overload impact processes, operations, and profitability, and how can data visualization help to solve issues with data management and consumption in organizations. The ability to quickly and effectively process information and make decisions equates to organizational survival in a dynamic, knowledge-based economy where all segments of society are heavily affected by information technologies and systems and data management industries. The growing number of systems apparatuses challenges both individuals and organizations, resulting in reports of fatigue and experiences that compromise successful performance.

The objective of this literature review is to discuss how data visualization tools help address information overload and optimize decision making and the business intelligence process in organizations. It concludes that data visualization, indeed, is critical in helping individuals capture, manage, organize, visualize, and present understandable data, but that decision making is affected by cognitive factors that interfere with data processing and interpretation in decision makers.

Keywords: Data Visualization, Information Overload, Business Intelligence.

INTRODUCTION

The development of information systems tools and data management applications made available via multiple resources through technological communication apparatuses impacts not only on the amount of data available to users across functions, segments, and industries, but, most critically, it affects the way individuals collect, process, and use information and share knowledge (Bawden, 2009). Information availability and the demand for access is increasingly becoming a topic of concern and study by researchers and practitioners that are focused on data and information access and management, as well as business processes improvement professionals and decision-makers. Some of the concerns concentrate on the amount of data available compared to the amount of data that can possibly be processed and absorbed by humans. The other aspect of this phenomenon is more objective in terms of profitability and productivity, suggesting that managers and CEOs should focus on the quality of data and user satisfaction, even mental health, as information overload becomes a counterproductive factor in organizations relying heavily on information access, use, and sharing.

This paper sets out to analyze information production, access, use, and overload behind the profound technological advances that have been transforming and redesigning the ways in which data is managed, accessed, and applied to organizational goals, strategies, processes, and product innovation. This descriptive analytical work presents a literature review analysis and highlights the research work of Merlo and Al-Hawamdeh (in press, 2022), which focused on a survey of over 100 respondents regarding information overload challenges and organizational tools designed to manage, share, and evaluate information. The study discussed the results from individuals' perceptions and experiences with data overload and data visualization tools applied in their organizations and their impact on productivity, identifying if, and which, data visualization systems are available for them. It also discussed the respondents' experience with information overload related to work, analyzing how the results impact productivity. The level of engagement in information sharing/overload and visualization practices were the focus of observation in the study.

Based on an intricate perspective on information overload and data visualization, this study will look at how data visualization tools can help organizations prevent information loss, anxiety, underuse of information systems, issues with poor or lack of information access, negative impacts on productivity and business processes improvement

CONCEPTUALIZING THE INFORMATION OVERLOAD PHENOMENON

Advances in Information and Communication Technologies (ICTs) over the past decade have resulted in massive data creation and processing, which has led to reduced usability and complicated the access and management of intricate datasets and information that are available via multiple applications at any time and in various formats. From handheld devices to cloud computing infrastructure, data is available everywhere, both on premises and in the cloud, based on intelligent machines and artificial intelligence programs.

As data and information access proliferate, the problems with information quality and knowledge creation intensifies, resulting from the relatively new paradigm of technology expansion and the origins of knowledge production and information access and sharing responsibilities. Reviewing information overload and its impact on organizations and decision-making, Edmunds and Morris (2000) indicate that there are two approaches to be considered: boost in technology and intelligent agents, with the latter referring to the emerging set of solutions in Artificial Intelligence (AI) and Machine Learning (ML) and how they help businesses.

Information Overload is the result of a combination of factors that impacts information processing capacity, causing real problems which are defined as "pathologies of information" (Eppler and Mengis, 2004). Blair (2010) defends that the issues with information overload are rooted in medieval principles of information processing, arguing that information issues result from the inevitable process of creativity in information processing and knowledge creation, where the position of knowledge is the result of the "frame of reference" associated with the Renaissance period. Hiltz and Turoff (1985) advocate for structured communication systems as a strategy to avoid information overload. The authors claim that the use of computer-mediated communication systems (CMCS) and telecommunication networks are valuable in processing the information and communication flow, stressing that it presents as many solutions as potential problems. While the costs of automatization are lowered, behavioral problems with excessive automatization could become an underlying issue.

Technological apparatuses increasing data production, processing, and consumption only accentuates information overload and the issues related to it. In business, individuals are exposed to multiple information systems applications with a demand for access and response to data that is unprecedented. From desktop computers to handheld devices and the ability to connect at anytime from anywhere both on-premises or in the cloud, individuals are exposed to a large amount of information and the sense of being overwhelmed that can be counterproductive and lead to inefficiency. In the study of Merlo and Hawamdeh (in press, 2022), the impact of information overload on productivity and how data visualization tools play a role in the maximization of data and information consumption was investigated, with the authors defending that excessive exposure to data and information negatively impacts employees' consumption of information and their ability to innovate and collaborate in teams and actively engage in normal work-related tasks. The authors argue that data visualization is critical in minimizing the unfavorable effects of data overload and maximizing the use of knowledge management systems and the understanding of data by individuals in an organization, in particular those in leadership positions who rely on data to guide business strategies and operations. It is also suggested that data visualization can reduce information overload and maximize the effective use of information systems, data analytics and visualizations tools by individuals in organizations.

THEORETICAL FRAMEWORK

A growing rate of data generation imposes the need to reevaluate data and information consumption and management requirements. Data acquisition, processing, storage, sharing, and use are some of the critical aspects of data management and governance that guide knowledge processing in organizations. Efficiency in modern computer systems interfaces allows us to explore large and complex datasets that can make individuals feel exhausted and disengaged. Merlo and Al-Hawamdeh (in press, 2022) investigated a few critical questions that guided the work in this paper, which aims to contribute to the expansion of studies focusing on the question: How does data visualization help organizations dealing with information overload? The authors presented a series of arguments from respondents and some of the findings that inspired the development of a conceptual framework for information overload and data visualization impacting decision-making in organizations. Empirical analysis was utilized as a base model and main theoretical framework for this analysis.

Throughout the research, analysis of certain dimensions of information overload and data visualization are clearly intertwined, while other aspects are perceived as complementary. Gaps in information overload and access are identified and the conceptual framework presented signals areas for further research in the field, resulting in information mismanagement and anxiety that is damning to a successful data management process and a fluid business process, and creates poor business outcomes and decreased productivity. From the perspective of practitioners, the findings of the literature review can be considered valuable as a conceptual framework where the independent variable is Information Overload and dependent variable is the Use of data analytics tools, supporting that data visualization tools are critical in managing issues with information overload and data management in organizations. For example, practitioners could first identify issues with excessive exposure to data and information via multiple information systems in organizations, as documented in the study of Merlo and Al-Hawamdeh (in press, 2022), where participants expressed a high level of stress about the demand of organizational access to information as compared to personal information access. Participants also expressed anxiety with the amount, quality, and periodicity in which they are required to deal with a large amount of information in organizations.

A second key aspect of this study aimed to demonstrate how data visualization tools assist users in collecting, organizing, representing, and consuming information, particularly large sets of information that are now being partially processed by artificial intelligence systems in many medium and large size corporations.

Reflecting on how these results will help organizations and individuals avoid approaching data management and visualization superficially by instead prioritizing the foundation of information systems in data quality, effective use, and security, the presented conceptual framework provides a base to explore the issues with information overload in organizations and the misuse or under- use of data visualization tools, techniques, and systems for building a business process roadmap that will consider complex human factors, information systems availability, and processes guidelines in order to increase productivity and profitability.

The investigation of Merlo and Al-Hawamdeh (in press, 2022), resulted from the survey research collecting data from 100 individuals in high tech and educational industries regarding information overload and data visualization, which reported that 45% respondents spend between 30 and 60 percent of their work time accessing information online and that:

> information overload and information fatigue on productivity and employees is well documented, we seek here to understand the level of awareness and knowledge of the organization's employees with data analytics tools and data visualization tools and technologies. Based on the theories presented and discussed in this study, and the analysis of data collected from participants, it is appropriate to conclude that data visualization tools can positively decrease the effect of information overload by mitigating the absorption of extraneous information and providing meaning for better data organization and presentation, and that individuals are struggling with the large set of unstructured and, considered, unvaluable, information in the workplace. (in press, 2022)

Merlo and Hawamdeh (in press, 2022) state that survey findings corroborate the initial hypothesis, that information overload is a growing and pervasive process that negatively impacts workers and business outcomes and profitability. Conclusions also signaled that due to information fatigue, individuals (a total of 69.23% of respondents) stated that the exposure to such a large amount of information at work causes them stress and anxiety, a phenomenon that is described by Buchanan and Kock (2001) as Information Fatigue Syndrome, resulting from issues of information overload and its adverse effect on individuals and the decision-making process. Future research should provide an empirical study investigating the impact of information overload and access on individual performance in the workplace, highlighting the role of leadership and decision makers.

An investigation conducted by Davis and Ganeshan (2009) sought to examine the causes leading to information overload by developing a causal model of information overload, building from research on cognitive psychology, marketing, and information systems and incorporating attributes related to aversion to loss and the tendency to hoard information and information input attributes such as redundancy and the volume of information processed and how it affects decision performance. Davis and Ganeshan (2009) build a model for decision performance that investigated decision accuracy in relation to the amount of information used in the decision making. The results of the study indicated that there is a "concern that information will become unavailable in the future leads to individuals holding on to more information than they would without the threat" and that holding of information leads to "a general hoarding type behavior when it comes to information that does not necessarily improve the quality of decisions and very often leads to delayed and poorer quality decisions in the real world." (p. 10). The authors concluded that the more anxious that individuals feel about acquiring and holding on to information (aversion to loss or information hoarding) the lower the level of satisfaction with decisions based on this information, while people with continuous access to information were more satisfied with the quality of their decision making. Therefore, the conclusions indicated that confidence in decision making is not directly correlated with the volume of information processed.

Data Visualization Tools facilitating data and knowledge processing in Organizations: Business Intelligence

Business Intelligence (BI) is defined as systems that combine data gathering, data storage, and knowledge management. It uses artificial intelligence for business, helping companies to make better business decisions by providing access to analyzed data (Negash and Gray, 2008). Based on that concept, it is supported that BI plays a critical role in supporting decision-making systems in organizations. Power (2002) elaborates that BI provides the framework for data-driven systems (DSS), emphasizing the processing of large volumes of both structured and unstructured data, allowing data availability in the time and format that is appropriate to users. Data Analytics and Business Intelligence reports indicate that data visualization software quality is ranked based on completeness and ability to execute, according to the Kronz et al. (2022) annual report, where Tableau is indicated as one of the top leaders in both categories for many years until the most recent Gartner report, dated March 2022, showing that Microsoft now leads the way for solutions in business intelligence (Gartner, 2022a).

The Analytics Business Intelligence (ABI) platforms "enable less technical users, including businesspeople, to model, analyze, explore, share and manage data, and collaborate and share findings, enabled by IT and augmented by artificial intelligence (AI)" (Kronz et al., para. 2). The volume of and variety of data in the current market is driving vendors of ABI platforms to focus on cloud deployment companies looking to address scalability and performance needs. The platform functionality includes twelve critical capabilities: Security, Governance, Cloud-enabled analytics, Data source connectivity; Data preparation, Catalog; Automated insights, Data Visualization, Natural language query, Data storytelling, Natural language generation, and Reporting.

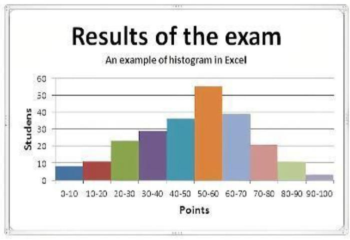
According to Sarveswar (2021):

> Advanced Analytics refers to as data mining, forecasting and predictive analytics, which takes an advantage of statistical analysis techniques to predict certain measures on facts. Corporate Performance Management is a general form of category that usually provides a container for several pieces to get into so that the whole tells a story. For suppose Take an example of balanced scorecard that displays portlets for financial metrics combined with organizational learning. (p. 87).

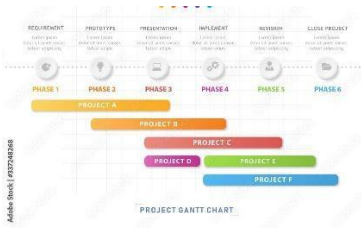
Data visualization tools present a software solution for data visualization, allowing the input of data in a format that is easily manipulated. According to Gartner (2022b) (the world's largest information technology research, developing rigorous research methodology to ensure independent and objective insights from advisory experts and clients), "Many platforms are adding capabilities for users to easily compose low-code or no-code automation workflows and applications. This blend of capabilities is helping to expand the vision for analytics beyond simply delivering datasets and presenting dashboards to delivering enriched contextualized insights, refocusing attention on decision-making processes, and ultimately taking actions that will deliver business value." (Kronz et al., 2022, para. 4).

Sarveswar (2021) supports that Business Intelligence provides superior tools that permit employees to "easily convert their business knowledge through analytical intelligence to solve many business problems and issues, like Internet delivered marketing campaigns, increase response rates from direct mail and much more." (p. 90). There are a multitude of formats and applications for data visualization, with the most common including theta visualization formats in **Figure 1** below:

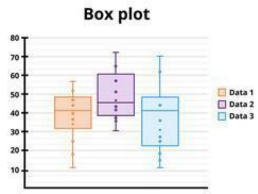
- **Histograms**



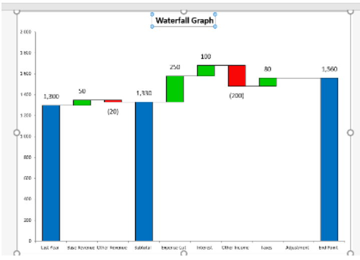
- **Gantt charts**



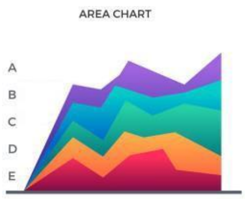
- **Box-and-whisker plots**



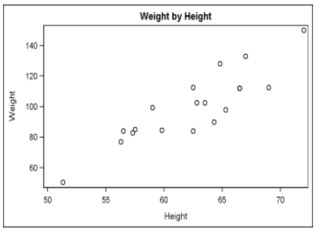
- **Waterfall charts**



- **Area charts**



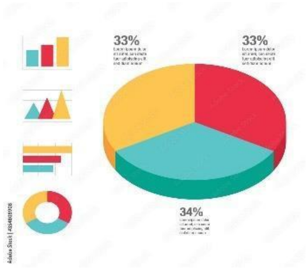
- **Scatter plots**



- **Infographics**

- **Maps**

- **Pie charts**



- **Bar charts**

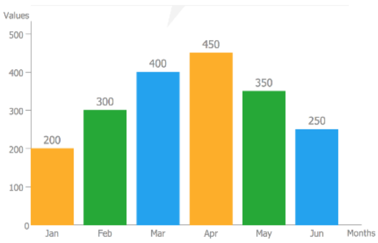


Figure 1: Most Common Data Visualization Formats. Source: Images from Stock Adobe

Data Visualization tools adopted by modern organizations: Tableau and Microsoft (Power BI)

Power (2002) states that "managers and their support staff need to consider what information and analyses are actually needed to support their management and business activities" (p. 6). The author defends that most managers are not interested in large amounts of data or information, but rather in a summarized version that is represented in charts or graphs, sometimes with a few tables. Whether needing to access information periodically or daily, most managers prefer that information is available on demand.

Data visualization will allow managers to better design strategies based on a simplified understanding of the data related to business processes and operations. Data visualization will enhance data processing and organization, in a visual format that helps individuals in the organization to absorb, handle, and manage complex information and data sets (Kopáčková and Škrobáčková, 2006).

Tableau

One of the most widely used data visualization tools, Tableau was recognized as the leader in visualization and Business Intelligence in the world, according to the Gartner report (2022). With the capability to convert raw data into a meaningful and structured format, it supports the creation of a simple chart with creative and interactive visualizations. Akhtar et al. (2020) state that: "tableau, a tool use for complex visualization and simplification of complex data. It was designed to help the user to create visuals and graphics without the help of any programmer or any prior knowledge of programming." (p. 28).

Tableau is reported to present a user-friendly interface with the capability to: integrate multiple sources of information in different formats, import data, output visualization, and allow users to present data in an interactive manner. Tableau offers desktop, web-hosted, and server versions, and since its acquisition by Salesforce was completed in 2019 (Tableau, 2019, para 1), it also offers integration with Customer Relationship Management (CRM). Another option for Tableau users is Tableau Public, a free, limited access version with basic data visualization features.

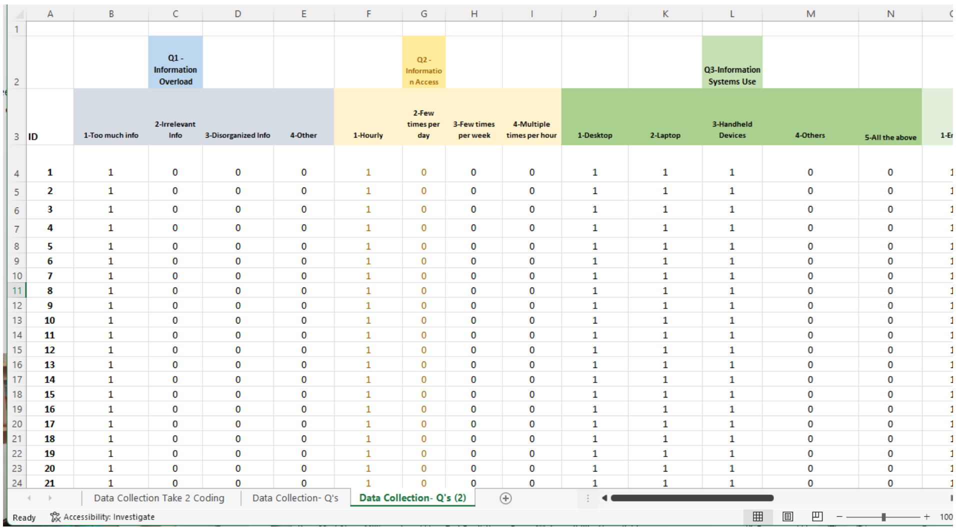
Microsoft – Power BI

Microsoft (2022) Power BI relies on business intelligence clouds that create and share interactive data visualizations across global datacenters (para. 1). It committed to the national cloud compliance regulations in its goal to provide an advanced business intelligence solution to help manage datacenters around the globe. According to the Gartner report (2022), Microsoft was named the leader in BI, taking the lead over Tableau by providing reliable performance that meets regulatory requirements for privacy, extending data loss prevention and governance, and providing efficient organizational data security by using the Power BI to support data analytics reporting and visualization, and the infrastructure based on AI to transform big data in a competitive and fast manner, connecting queries and data models for reporting in dashboards, allowing users to get access to real-time analytics and an automate workflows that allows a faster decision-making. Research indicates that Power BI is increasingly acquiring popularity due to its perceived advantages, such as intuitive and easy to use features for a much lower cost. (Gartner, 2022).

Kokina, Pachamanova and Corbett (2017) investigated the role of data visualization and analytics in the performance management of a small e-commerce organization. The authors explain how the utilization of Tableau and Microsoft Excel helped organize data and draw analytics, allowing visualization that was able to help identify business problems, market penetration and expansion, and address the professional development of competencies in predictive analytics. The authors discussed metrics and key performance indicators (KPIs) and the benefits of data visualization and analytics to drive business strategy and increase profitability.

Below is an example of how data visualization helps with data understanding. Figure 2 shows data in an Excel spreadsheet versus data cleaned and organized in Power BI dashboard in Figure 3:

Figure 2: Data Organized in Spreadsheet – Excel

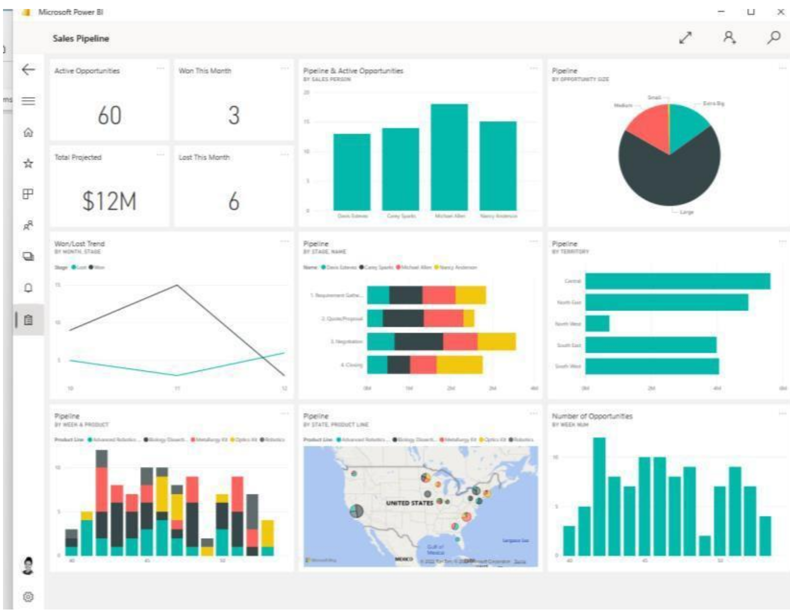


	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1														
2			Q1 - Information Overload				Q2 - Information Access					Q3-Information Systems Use		
3	ID	1-Too much info	2-Irrelevant Info	3-Disorganized Info	4-Other	1-Hourly	2-Few times per day	3-Few times per week	4-Multiple times per hour	1-Desktop	2-Laptop	3-Handheld Devices	4-Others	5-All the above
4	1	1	0	0	0	1	0	0	0	1	1	1	0	0
5	2	1	0	0	0	1	0	0	0	1	1	1	0	0
6	3	1	0	0	0	1	0	0	0	1	1	1	0	0
7	4	1	0	0	0	1	0	0	0	1	1	1	0	0
8	5	1	0	0	0	1	0	0	0	1	1	1	0	0
9	6	1	0	0	0	1	0	0	0	1	1	1	0	0
10	7	1	0	0	0	1	0	0	0	1	1	1	0	0
11	8	1	0	0	0	1	0	0	0	1	1	1	0	0
12	9	1	0	0	0	1	0	0	0	1	1	1	0	0
13	10	1	0	0	0	1	0	0	0	1	1	1	0	0
14	11	1	0	0	0	1	0	0	0	1	1	1	0	0
15	12	1	0	0	0	1	0	0	0	1	1	1	0	0
16	13	1	0	0	0	1	0	0	0	1	1	1	0	0
17	14	1	0	0	0	1	0	0	0	1	1	1	0	0
18	15	1	0	0	0	1	0	0	0	1	1	1	0	0
19	16	1	0	0	0	1	0	0	0	1	1	1	0	0
20	17	1	0	0	0	1	0	0	0	1	1	1	0	0
21	18	1	0	0	0	1	0	0	0	1	1	1	0	0
22	19	1	0	0	0	1	0	0	0	1	1	1	0	0
23	20	1	0	0	0	1	0	0	0	1	1	1	0	0
24	21	1	0	0	0	1	0	0	0	1	1	1	0	0

Data Collection Take 2 Coding | Data Collection- Q's | Data Collection- Q's (2)

Ready | Accessibility: Investigate

Figure 3: Data Visualization of Sales metrics Dashboard on **Power BI**



The International Data Corporation (IDC) Global DataSphere, a global market intelligence firm, reported that: “More than 59 zettabytes (ZB) of data will be created, captured, copied, and consumed in the world this year." (Businesswire, 2020, para. 2). The IDC claims that by 2024 there will be a growth rate of 26% in data consumption, despite of the negative impact of COVID- 19 in new data creation. The key findings in data creation, per reports from the Global DataSphere forecast include the following:

a) The amount of data created over the next three years will be larger than the amount of data created over the past 30 years
b) Productivity/embedded data is the fastest growing category in data creation (40.3% CAGR for 2019-2024 forecast period)
c) Sensors are being embedded into anything and everything, affecting the data contextualization process and increasing the amount of metadata.
d) Data privacy and regulatory initiatives continue to intersect with video surveillance data, creating a conflict between security, privacy, and individuals' rights. (Businesswire, 2020, para. 5- 9).

Zhang and Melguizo (2022) discuss the Data 4.0 data-driven economy in the World Economic Forum stating that: "Institutions as well as the private sector play a huge role in the transition to a truly data-oriented culture." (para. 2). The authors discuss how the number of networked devices is projected to reach 125 billion by 2030, and how the explosion in data creation and consumption impacts the private sector and the organizational decision-making process. Zhang and Melguizo (2022) highlight the Forrester survey, with 4,036 high-level executives that demonstrated how 70% of data decision makers are gathering data faster than they can analyze and use, with more than 67% constantly needing more data (para. 6).

In predicting how data can help address the world's biggest challenge, King (2022) defends that "data can help us tackle our largest societal challenges, including climate change, inequality, global health and economic resilience" (para. 1), arguing that business leaders have the responsibility to use the limitless data power available to resolve the data challenges of 2022. King (2022) stresses that the value of data is well established in the private sector and that successful businesses have to capture this value through efficient product design and a global marketing data strategy, worth an estimated 52 billion dollars in 2021. (para. 3).

As a result of the virtual meeting where leaders were invited to share perspectives on the power of data to business, the following slogans were shared:

1- **"Leverage predictive analytics"** by Igor Tulchinsky, Founder, Chairman and CEO, WorldQuant - focusing on Big Data and IoT; "Developing a global talent pool", stressing the power of AI and Machine Learning for faster business solutions, in particular healthcare with the COVID-19 emergency response, where enabling talents to contribute significantly is critical to harness opportunities.
2- **"Enable data to flow across borders"** by Dr. Norihiro Suzuki, Vice President and Executive Officer, Chief Technology Officer, General Manager of the Research & Development Group and General Manager of the Corporate Venturing Office, Hitachi, Ltd. – advocating for the use of knowledge to accelerate solutions, stressing the importance of data flows and trusting relationships with stakeholders.
3- **"Inclusive and responsible solutions",** Crystal Rugege, Managing Director, The Centre for the Fourth Industrial Revolution (C4IR) Rwanda - where the importance of data and technology in problem solving was accentuated over the past two years because of the COVID-19 pandemic, which amplified the need for agile data governance that allows a framework for quick adaptability and response to an ecosystem that is constantly evolving and being affected by socio-economic developments, technological advances, privacy rights, and more (para. 26).
4- **"Change the fate of the ocean**", by Kimberly Mathisen, CEO, The Centre for the Fourth Industrial Revolution (C41R) Ocean - highlighting the Industry 4.0 and how new technologies allow limitless management and measurement of data, empowering collaboration.

The importance of data and its possibilities based on the explosion of technological systems such as AI, ML, and BI tools to business is undeniable. The optimization of those technologies to drive competitive advantage, collaboration, innovation, and profitability is centered around a data-based decision-making process, which must rely on a data-driven economy and the capability to quickly respond to an ecosystem of factors. In a review of empirical studies investigating decision making and the use of visual representations graphics, Padilla, Creem-Regehr, Hegarty and Stefanucci (2018) concluded that the effective use of data visualization tools is subject to a series of psychological factors that affect decision making. The authors present a model of visualization decision making that emphasizes memory and involves a variety of processes that influence the message and ultimately behavior. Padilla et al. (2018) argues that visual-spatial factors and working memory are examples affecting decision making and that knowledge-driven processes affect or override the effects of visualization methods, making the job of data and business analysts and visualization designers more complex as systems are designed, built, and implemented. As professionals work towards simplifying and representing data, factors of cognition interfere with data consumption patterns and decisions must be recognized and understood.

FINAL REMARKS AND RECOMMENDATION FOR FUTURE WORK

This study indicates a clear correlation between information overload and data visualization, indicating the impact of those variables on business performance. The focus of this paper is to urge the research community to recognize that there are benefits to using data visualization tools to minimize or eliminate the identified counterproductive and negative impacts of IO in organizations. Previous research by Merlo and Al-Hawamdeh (in press, 2022) was critical in the development of this study, as its findings resulted from a survey focusing on the impact of IO and the use of data visualization on productivity, providing relevant data on users' perceptions and experiences.

A number of studies focusing on data analytics have been published in recent years, with a few discussing data visualization and its impact on business processes, strategies, and operations, and, most importantly, how data visualization is a powerful tool in creating BI and a competitive advantage by enabling a fast and efficient response to customers' needs and demands and market changes in the midst of a fast-paced knowledge-based society. Research on data visualization remains limited, with little literature and few comprehensive frameworks available to help guide decisions based on visualization tools.

Businesses as social systems are constantly having to make decisions, and that task becomes complicated as factors influencing business processes and operations are changing rapidly, being impacted by big data, artificial intelligence, and multiple machine learning solutions in a competitive global market. This study reports the cases of Tableau and Microsoft as examples of Business Intelligence solutions, competing for innovation and leadership in the industry, with Microsoft Power BI taking the lead in performance and popularity in the market over Tableau over the past two years (Gartner, 2022).

When discussing the science behind management decisions, Herbert (1960) argued that humans experience three phases in the process of making a decision: Intelligence, Design, and Choice. Although data visualization and analytics systems are critical in processing complex volumes of datasets, organizations must recognize that human factors and the approach to data consumption- behavior and leadership styles can impact the success of data visualization and information systems management applications that are still susceptible to data interpretation and decisions on how the data is managed, information is processed, and knowledge is shared and applied towards business solutions.

Further, this theoretical study describes the importance of data visualization for decision-making, which aims to inspire future empirical work studying the impact of data visualization in the optimization of business processes, BI, and outcomes. It intends to offer guidance for future exploratory research for visualization tools based on cognitive theory and empirical findings, helping to advance the scope of knowledge in the field of data and business analytics and visualization and business intelligence. The presented study is socially relevant as it provides data to inspire researchers and practitioners focusing on any of the three highlighted areas, further helping with the advancement of studies and models for business processes improvement, the betterment of data visualization tools adoption and utilization, and the overall approach to data consumption with the goal of increasing organizational productivity and profitability.

REFERENCES

Akhtar, Nikhat; Tabassum, Nazia; Perwej, Asif; Perwej, Yusuf (2020). Data analytics and visualization using Tableau utilitarian for COVID-19 (Coronavirus). Global Journal of Engineering and Technology Advances. V.3(2). pp. 28- 50. DOI: https://doi.org/10.30574/gjeta.2020.3.2.0029

Bawden, D., & Robinson, L. (2009). The dark side of information: overload, anxiety and other paradoxes and pathologies. *Journal of information science*, *35*(2), 180-191.

Buchanan, J., & Kock, N. (2001). Information overload: A decision making perspective. In *Multiple criteria decisions making in the new millennium* (pp. 49-58). Springer, Berlin, Heidelberg.

Businesswire (2020). IDC's Global DataSphere Forecast Shows Continued Steady Growth in the Creation and Consumption of Data. Retrieved from: https://www.businesswire.com/news/home/20200508005025/en/IDCs- Global-DataSphere-Forecast-Shows-Continued-Steady-Growth-in-the-Creation-and-Consumption-of-Data

Davis, Joseph G.; Ganeshan, Shayan,(2009). Aversion to Loss and Information Overload: An Experimental Investigation. ICIS2009 Proceedings. Paper 11 Retrieved from: http://aisel.aisnet.org/icis2009/11

Edmunds, A., & Morris, A. (2000). The problem of information overload in business organizations: a review of the literature. *International journal of information management*, *20*(1), 17-28.

Eppler, M. J. & Mengis, J. The Concept of Information Overload: A Review of Literature from Organization Science, Accounting, Marketing, MIS, and Related Disciplines. The Information Society, 20, 2004, pp. 325-344.

Gartner (2022a). Retrieved from: https://www.gartner.com/en/about

Gartner (2022b). Data Virtualization. Retrieved from: https://www.gartner.com/reviews/market/data-

virtualization Herbert, Simon (1960). The new science of management decision. New York: Harper.

Hiltz, S. R., & Turoff, M. (1985). Structuring computer-mediated communication systems to avoid information overload. *Communications of the ACM*, *28*(7), 680-689.

King, Rebecca (2022). Leaders on how data can tackle global challenges of 2022 | World Economic Forum (weforum.org). retrieved from: https://www.weforum.org/agenda/2022/01/data-can-help-tackle-global- challenges-2022-predictions/

Kokina, J., Pachamanova, D., & Corbett, A. (2017). The role of data visualization and analytics in performance management: Guiding entrepreneurial growth decisions. Journal of Accounting Education, 38, 50-62. https://www.sciencedirect.com/science/article/pii/S0748575116301221

Kopáčková, H., & Škrobáčková, M. (2006). Decision support systems or business intelligence: what can help in decision making? Scientific papers of the University of Pardubice. Series D, Faculty of Economics and Administration.

Kronz, Austin; Schlegel, Kurt; Sun, Julian; Pidsley, David; Ganeshan, Anirudh (2022). Magic Quadrant for Analytics and Business Intelligence Platforms. Retrieved from: https://www.gartner.com/doc/reprints?id=1- 292LEME3&ct=220209&st=sb

Merlo, Tereza R.; Al-Hawamdeh, Suliman (in press). Information Overload and the use of visual analytics tools in Organizations. In: Merlo, Tereza R. (Editor). Understanding, Implementing, and Evaluating Knowledge Management in Business Settings. IGI-Global.

Negash, S., Gray, P. (2008). Business Intelligence. In: Handbook on Decision Support Systems 2. International Handbooks Information System. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-48716-6_9

Padilla, L. M., Creem-Regehr, S. H., Hegarty, M., & Stefanucci, J. K. (2018). Decision making with visualizations: a cognitive framework across disciplines. Cognitive research: principles and implications, 3, 29. https://doi.org/10.1186/s41235-018-0120-9

Power, Daniel (2002). Decision Support Systems: concepts and resources for managers. Westport: Greenwood Publishing Group.

Sarveswar, Darapureddy. (2021). Business Intelligence. Bodhi International Journal of Research in Humanities, Arts and Science. V.5 (2). pp. 87-92

Tableau (2019a). Salesforce Completes Acquisition of Tableau.. Retrieved from: https://www.tableau.com/about/press-releases/2019/salesforce-completes-acquisition-tableau

Tableau (2019b). Data visualization beginner's guide: A definition, examples, and learning resources. Retrieved from: https://www.tableau.com/learn/articles/data-visualization

Zhang, Pepe; Melguizo, Angel (2022). Data 4.0 – Rethinking rules for data-driven economy. Retrieved from: https://www.weforum.org/agenda/2022/01/data-4-0-rethinking-rules-for-a-data-driven-economy/

Metadata Practices of Academic Libraries in Kuwait, Oman, and Qatar: Current State, Risks, and Perspectives for Knowledge Management

OKSANA L. ZAVALINA[*]
Department of Information Science, University of North Texas, Denton, Texas,76207, United States of America
E-mail: oksana.zavalina@unt.edu

SALEH H. ALJALAHMAH[†]
Basic Education College, The Public Authority for Applied Education and Training (PAAET), ZIP/Zone, Kuwait
[†]*E-mail: salehaljalahmah@yahoo.com*

Developing, implementing, and managing metadata is crucial to successful knowledge management, and academic libraries have traditionally played a central role in these activities. The Arabian Gulf countries are underrepresented in the existing research into library metadata practices. This exploratory study used semi-structured interviews of metadata managers at 8 universities with the goal of developing understanding of the current state of metadata practices, including descriptive cataloging, identity management, and knowledge organization in academic libraries of three Arabian Gulf countries (Kuwait, Oman, and Qatar), as well as potential future developments to facilitate discovery of resources. Findings provide insights into this previously under-researched area and contribute to understanding of knowledge management and risks on a global scale.

1. Introduction

There is a need for providing adequate metadata to enable access to Arabic-language materials as Arabic is one of the top 5 languages of the Internet users (Internet World Stats, 2020). Currently, the number of records representing Arabic-language materials in the largest database of metadata records, WorldCat union catalog, is over 3.3 million (Inside WorldCat, 2021). Most of WolrdCat metadata contributors are libraries, in particular academic libraries. According to Zehery (1997), academic libraries play the most important role in providing access to information in the Arabian Gulf region. Thus, exploration of academic libraries' metadata practices is needed in the Arabian Gulf countries. This study examined the status of the organization of knowledge in the online catalogs (also known as integrated library systems) in academic libraries in Kuwait, Oman, and Qatar, analyzed similarities and differences in how information is organized and presented to the users in databases. In the following sections, brief review of relevant literature is followed by study methods description, findings, their discussion, conclusions, and ideas for future research.

2. Literature Review

In the second half of the 20th century, publications by Holloway (1959), Lemos (1981) and others reviewed the state of library services in the Middle East. According to them, many libraries in the region were business-owned, and their catalogs were often unavailable; for the state-operated public and academic libraries, participation in union catalogs that aggregate library metadata into a centralized environment for easier access and resource discovery was rare. As evidenced by Khurshid (1997) study, in late 1990s, libraries of the Arabian Gulf region did not yet actively participate in the cooperative cataloging efforts and databases such as global WorldCat that functions since 1971 or regional union catalogs.

By the turn of the century, the integration of Middle Eastern library metadata practices with the ones used in other parts of the world was well in progress. Libraries were found to rely on commonly used tools and standards such as the Dewey Decimal Classification scheme and Library of Congress subject headings (LCSH) to provide subject access to materials in their collections (El-Sherbini, 2001). To minimize the Western-focused bias and improve discoverability of Arabic-language and other materials of interest to the users in the region, modified versions of these standards had to be developed: for example, Dewey Decimal Classification went through multiple regional revisions that were coordinated by the Jordan Library and Information Association since 1970s (Eid, 2019). Regional standards for organizing information and knowledge were also developed. One example is Maknaz expanded thesaurus consisting of subject and genre terms, personal, corporate, and geographic names (El-Sherbini, 2015). Arab Union Catalog was launched in 2006. Translations of international metadata standards were completed to provide region's librarians with tools for creation of metadata that meets the latest requirements (Eid, 2019).

The available research literature on knowledge organization, including metadata practices in the Arabian Gulf countries in the 21st century, is currently very limited. Osman and El-Masry (2016) reviewed experiences of catalogers of Arabic-script materials (including those working in Qatar, United Arab Emirates, and Saudi Arabia) with application of the international metadata standards. Their study revealed challenges such as lack of support by integrated library systems for the links between Romanized text metadata fields and their vernacular text counterparts in MARC 21 metadata records, synchronization issues between Arab Union Catalog and WorldCat databases, bibliographic records duplication in Arab Union Catalog, as well as lack of the way to import records into the library's online catalog directly from Arab Union Catalog. As part of the study that examined application of information communication technologies, Al-Ansari (2011) found that 72% of the 25 Kuwaiti libraries were automated and used online public access catalogs. While 10 out of 18 libraries relied on Horizon Library Automation System, Al-Ansari also observed the trend for use of in-house systems for online catalogs, and their limited functionality, including incompatibility with international standards such as MARC, which complicated online sharing of resources. Similar to findings of Al-Ansari's study, Verizon was the system used by most (5 out of 6) libraries in the Ur Rehman and Al-Huraiti's (2010) study of six academic libraries in Kuwait. They found that cataloging module implementations of the integrated library systems were almost complete in the surveyed libraries. A recent exploratory case study (Aljalahmah and Zavalina, 2021) examined metadata records and compared application of metadata schemes, controlled vocabularies for name and subject representation, classification systems etc. in a museum, an archive, and a library in the Arabian Gulf region for physical and digital collections. In addition to content analysis of a sample of metadata records, that study also used interview data. One of the observations made as part of that study was that the academic library metadata manager who participated in the interview did not separate in their answers information about metadata in the institution's digital collections and online public access catalog (OPAC) otherwise known as Integrated Library System (ILS) that provided access through MARC 21 metadata to institution's entire collection, that mostly includes physical analog items and digital resources obtained through subscription rather than digitization. Together with the fact that digital library development is an emerging trend, and most libraries in the Arabian Gulf region have not yet developed their digital libraries or repositories, this pointed out the need to separately investigate information organization and knowledge representation practices in online catalogs and compare the patterns across. The study presented in this paper begins to address this need.

Education provided to catalogers and creators of metadata for digital repositories affects the metadata practices. Several studies have been published that looked at cataloging education in the Middle East: Egypt, Iran, Israel, and some Arabian Gulf countries. Khurshid's (1998; 2006) reports focused on metadata education state in Saudi Arabia. One study focused on Oman (Al Hijji and Fadlallah, 2013). However, no studies have been published in the recent years about cataloging education of metadata creators in Kuwait and Qatar, and the study presented in this paper collected and analyzed some relevant data on this under-researched topic.

3. Method

This study focused on three Arabian Gulf countries -- Kuwait, Oman, and Qatar -- and was guided by the following research questions:

- What are the current metadata (and broader information organization) practices used and what are the future perspectives?
- What are the training and practical experience levels of metadata managers, and what guidance is provided to catalogers and users?
- What risks and challenges to knowledge management are academic libraries facing now or expect to face in future projects?

Interviews were used as the data collection method. Researchers selected academic libraries that have digital project(s) or collection(s). Given the standing of academic libraries in the region, these institutions have the strong potential to become major contributors to regional digital repositories, and some have already developed digital collections of their own. To obtain accurate data on local practices, only institutions with headquarters located in Kuwait, Oman, or Qatar were selected. The research sample was stratified by country: between 5 and 6 participating institutions from each of the three countries were selected. Then, potential respondents – metadata managers responsible for information organization (including metadata) decisions -- were selected from the lists of employees on their institution’s websites. The interview recruitment email was sent to 17 potential respondents, and eight took part in the study, a response rate of 41%.

Ongoing COVID-19 pandemic poses significant physical health risks and related uptake in mental health issues caused by prolonged social isolation. Other global risks brought on by the pandemic include labor shortages and difficulty in recruiting employees. Importantly, it continues to present risk to knowledge management by negatively affecting the processes of knowledge acquisition through education at all levels, as well as through research. For example, the ability of researchers worldwide to arrange with potential study participants for in-person or audiovisual (e.g., Zoom) interviews is significantly limited which requires adoption of alternative ways to collect rich qualitative data from human study participants. One solution is relying on email interviews which allow participants to find time in their schedule to provide more thoughtful, reflective responses to the interview questions (Meho, 2006). Previous studies found email interviews effective and resulting in relatively high response rate (e.g., Park and Tosaka, 2015).

In this study, interviews were conducted by email, which allowed participants to respond whenever they can, considering the quarantine and work-from-home situations. The interview was semi-structured, with a set of main questions prepared in advance and follow-up questions as needed. The main interview questions were sent to the participants in both English and Arabic as a parallel text to help reduce participation barriers, and participants had the freedom to choose the language of their answers. Interviews included questions about integrated library systems, metadata schemes, controlled vocabularies for representing names and subjects, search options, search language interfaces enabled. Questions about whether and how the metadata creators and users are trained, what guidelines are available, how the academic libraries reach out to the users for feedback were also asked. Finally, interview instrument was designed to collect the data about educational background, length of experience, and sources of professional training received by metadata managers.

4. Results

Horizon integrated library system by Sirsi Dynix that had been reported by earlier studies as the most chosen ILS was only used by one institution participating institution in the study. Three other institutions used Sierra ILS by Innovative Interfaces. Two participants reported using EBSCO Discovery, and one of them customized the tool with EBSCO Admin and combined it with OCLC-hosted EZ-Proxy. One respondent's academic library used open-source Koha library management system. Libero library management system and the older popular application VTLS (Virginia Tech Library Systems) were used by one respondent each.

All 8 participants reported providing both simple and advanced search options for the ILS users. Only one participant reported providing a search interface with English as the single search language. One institution provides search capability in 4 languages (Arabic, English, French, and German), and remaining six academic libraries provide bilingual search interfaces in which users can search in Arabic and English.

Six libraries reported using MARC 21, and one reported using Dublin Core metadata scheme. Five participants reported using controlled vocabularies to represent names: the Library of Congress Name Authority File (LCNAF) was used in all these cases. Two participants reported not using any controlled vocabularies for names. For topical terms in metadata records 6 participants reported using the Library of Congress Subject Headings (LCSH) alone or in combination with other controlled vocabularies that are customized for needs of local users. Those regional controlled lists of topical headings included List of Standard Arabic Headings and Alkhazindar Subject Headings List. They were used by respondents for representing Arabic-language materials only (with the LCSH used for English-language materials). One participant reported their academic library did not utilize LCSH at all but only used the regional Big Subject Headings List. Five participants reported using Library of Congress Classification (LCC) for classifying all materials. Two 2 respondents use LCC for English-language materials and Dewey Decimal Classification (the version adjusted for Arab and Muslim world) for Arabic-language materials.

When asked about which sets of metadata guidelines are used by catalogers in metadata creation, one participant each reported relying on MARC 21 standard documentation and EBSCO Discovery system guidelines and utilizing locally developed metadata creation guidelines. The remaining 4 respondents claimed their library catalogers use no metadata creation guidelines.

Most (7 out of 8 participants) reported one or more kinds of training on utilizing library databases provided to the library users. Tutorials were provided by 2 participating institutions (one of which also relied on information literacy presentations). Online training was offered by 2 academic libraries, and one of them also offered face-to-face trainings. Online guidelines documents are utilized at two academic libraries. One institution relied on workshops for users.

Six metadata managers reported that their institution also reaches out to the users to collect feedback. Most contact users in person (n=4) and/or via email (n=3). One institution each replied that they use the website, social media tools, surveys, and phone calls as one of the user contact methods. Three institutions reported using workshops as ways to connect with users. The units within the academic library that maintain contact with users varied and included cataloging (n=1), collection development (n=1), public relations (n=1), reference desk and social media department (n=1).

Responses to questions regarding educational background revealed that metadata managers held academic degrees in areas relevant to the nature of the work. This included one participant majoring in computer science, three majoring in library science, and four majoring in library and information science. As to the level of degree, three respondents held a master's degree, other three a bachelor's degree, while one participant had an associate degree, and one did not specify the highest academic degree level. Among participants of this study, the total experience managing metadata and bibliographic databases ranged from 6 to 37 years, with the average of 19.88. All but one participant reported receiving additional training: colleagues provided training to 4 respondents, one other reported "personal training", two participated in training workshops, and one reported online training.

5. Discussion and Conclusions

This study observed substantial similarities in the tools used for information organization and knowledge representation by academic libraries in Arabian Gulf countries. The findings of this study confirm the trend of reliance on international standards reported in recent years (e.g., Eid, 2019), despite the regional challenges with implementing international standards pointed out by earlier studies (e.g., Khurshid, 1997; Osman and El-Masri, 2016). All institutions whose representatives participated in this study use international metadata standards such as the classification system and the subject headings by the US Library of Congress. Seven metadata managers who participated in this study also reported using the MARC 21 metadata scheme.

However, differences in the use of controlled vocabularies used would create metadata interoperability issues. Some institutions reported using different sets of subject headings for materials in Arabic language and those in English language. To meet the local needs, these academic libraries rely on subject headings lists developed in the Arabian Gulf countries, such as The Big Subject Heading List, Alkhazindar Subject Heading List, and Qa'imat Ru'us al-Mawdu'at al-Arabiyah al-Qiyasiyah lil-Maktabat wa-Marakiz al-Maʻlumat wa-Qawa'id al-Bayanat (also called List of Standard Arabic Subject Headings for Libraries, Information Centers and Databases). Also, no integrated library system is used by more than three libraries, which contradicted the assumption that many institutions might be using Horizon system which is known to be very popular in Kuwait (e.g., Al-Ansari, 2011, etc.). This might mean that in choosing digital content management systems for digital collections, institution select systems that meet their local needs but might not be supporting interoperability in aggregated environment.

Following metadata creation guidelines ensures higher quality of metadata through positively affecting it consistency, yet this study found that metadata creation guidelines are not being used in the Arabian Gulf countries academic libraries. A possible explanation for this finding is that most of these guidelines currently exist only in English and some other Western languages, which creates a language barrier to using them for majority of Arabian Gulf cultural heritage institutions employees who are not fluent in English.

Our findings also reveal the knowledge management risks of potential metadata quality issues caused by reliance on older, no-longer-updated vocabularies: e.g., one participating institution used a seriously outdated version of Dewey Decimal Classification system, another relied on Alkhazindar Subject Heading List, a legacy list that has not been updated since the 1990s when its developer died. The main reason for this trend, as well as for reliance on older legacy ILSes or content management systems is the different levels and sources of funding and financial standing of the institutions that participated in this study, budgets of some of which are very limited.

The trend among libraries in the Arabian Gulf countries to require and/or offer onsite training for their employees to prepare them to do their job according to that institution's standards and needs was reported by earlier studies of Library and Information Science education in the region (e.g., Khurshid, 2006; Al Hijji, 2012). Our study's findings show that this trend continues, which might indicate the lack of practical project-based learning in metadata education and the resulting lack of consistency in preparation of knowledge professionals.

The study found that while users are consulted for general feedback on collections and services, institutions do not examine and consider the users' needs when choosing information systems for content management, selecting existing or developing local metadata schemes and information organization standards, creating metadata records, etc.

Overall, the main risks to successful knowledge management through information organization in physical and digital libraries in the Arabian Gulf region, as revealed by this study, are the lack of sufficient long-term dedicated budgets and the lack of practical training in the library and information science programs. Another risk is over-reliance on foreign workforce that is unsustainable, as demonstrated by the exodus of foreign workforce due to COVID-19 resulting in many Arabian Gulf cultural heritage institutions being closed for a long period of time.

This study is exploratory in nature and has several limitations. One of the known limitations of the interview data collection method is that participants might provide inaccurate answers (e.g., Alshenqeeti, 2014). This phenomenon was observed in this study as well: one of the participants blended their experiences from their entire career with the experiences they had solely at their current position in their institution. While the overall benefit is getting useful contextual information, this raises some concerns regarding the accuracy of given answers. Another limitation was related to completeness of collected interview data as some participants left some questions unanswered.

Future research is needed to closely examine the current status, risks, and perspectives of the organization of knowledge in other Arabian Gulf countries that were beyond the scope of this study: Saudi Arabia, Bahrain, and United Emirates. Also needed are studies that comparatively evaluate metadata records in the databases of academic libraries and other institutions and will provide more robust understanding of metadata quality in Arabian Gulf bibliographic databases. The patterns observed in metadata and metadata-related practices across the region will be useful for planning and implementation of effective knowledge management through metadata management in large-scale portals that include metadata records from academic libraries and other cultural heritage institutions in various countries of the Arabian Gulf region.

References

Al Hijji, K.Z. (2012). Cataloguing and classification education in Gulf Cooperation Council (GCC) countries. *Cataloging & Classification Quarterly, 50*(4), 276–292. Retrieved from https://doi.org/10.1080/01639374.2011.653873

Al Hijji, K.Z., & Fadlallah, O.S. (2013). Theory versus practice in cataloging education in Oman: Students' perspectives. *Cataloging & Classification Quarterly, 51*(8), 929–944. Retrieved from https://doi.org/10.1080/01639374.2013.832456

Al-Ansari, H. (2011). Application of information and communication technologies in special libraries in Kuwait. *The Electronic Library, 29*(4), 457–469. DOI: 10.1108/02640471111156731

Aljalahmah, S.H., & Zavalina, O. L. (2021). Information representation and knowledge organization in cultural heritage organizations in Arabian Gulf: A comparative case study. *Journal of Information and Knowledge Management, 20*(2), 1–20.

Alshenqeeti, H. (2014). Interviewing as a data collection method: A critical review. *English Linguistics Research, 3*(1), 39–45.

Eid, S. (2019). Library metadata standards and Linked Data services: An introduction to Arab and international organizations. *Journal of Library Metadata, 19*(3/4), 163–185.

El-Sherbini, M. (2001). Metadata and the future of cataloging. *Library Review, 50*(1), 16–27. DOI: 10.1108/00242530110363217

El-Sherbini, M. (2015). Multilingual subject retrieval: Bibliotheca Alexandrina's subject authority file and linked subject data. In, B. Lausen, S. Krolak-Schwerdt, & M. Böhmer (eds.), *Data Science, Learning by Latent Structures, and Knowledge Discovery* (pp. 535–546), Springer, Berlin, Heidelberg.

Inside WorldCat. (December 2021). Retrieved from https://www.oclc.org/en/worldcat/inside-worldcat.html

Internet World Stats. (2020). *Internet World Users by Language: Top 10 Languages*. Retrieved from https://www.internetworldstats.com/stats7.htm

Khurshid, Z. (1997). Cooperative cataloging: Prospects and problems for libraries in Saudi Arabia. *Library Resources & Technical Services, 41*(3), 264–272. Retrieved from https://doi.org/10.5860/lrts.41n3.264

Khurshid, Z. (1998). Preparing catalogers for the electronic environment: An analysis of cataloging and related courses in the Arabian Gulf region. *Journal of Education for Library and Information Science, 39*(1), 2–13. Retrieved from https://www.jstor.org/stable/40324175

Khurshid, Z. (2006). Continuing education for catalogers in Saudi Arabia. Cataloging & Classification Quarterly, 43(3/4), 461-470. DOI: 10.1300/J104v41n03_14

Meho, L. I. (2006). E-mail interviewing in qualitative research: A methodological discussion. *Journal of the American Society for Information Science and Technology, 57*(10), 1284–1295.

Osman, R.R., & El-Masri, A.F. (2016). Cataloguing Arabic script materials: Challenges in using international standards and integrated library management systems. *Proceedings of the 82nd International Federation of Library Associations World Library and Information Congress*. Retrieved from http://library.ifla.org/id/eprint/1321/1/093-osman-en.pdf

Park, J. R., & Tosaka, Y. (2015). RDA implementation and training issues across United States academic libraries: An in-depth e-mail interview study. *Journal of Education for Library and Information Science, 56*(3), 252–266.

ur Rehman, S., & Al-Huraiti, R. (2010). Integrated systems applications in Kuwait academic libraries. *The Electronic Library, 28*(6), 858–872. Retrieved from https://doi.org/10.1108/02640471011093543

Zehery, M.H. (1997). University library development in the Arab Gulf region. *International Information & Library Review, 29*(1), 13–44. DOI: 10.1080/10572317.1997.10762412

An Ontology and a Reasoning Approach For Evacuation in Flood Disaster Response

Julie Bu Daher [1], Tom Huygue [1], Patricia Stolf [2] and Nathalie Hernandez [2]

[1] *IRIT, Université de Toulouse, CNRS, Toulouse INP, UT3, Toulouse, France*
[2] *IRIT, Université de Toulouse, CNRS, Toulouse INP, UT3, UT2, Toulouse, France*
{julie.bu-daher, patricia.stolf,nathalie.hernandez}@irit.fr

Abstract. Managing flood-related data to assist in the disaster management is a critical process of high importance during a flood disaster. These data are heterogeneous and can be provided from different data sources, and integrating them is a challenging task which allows to infer new information that helps in limiting the consequences of a flood. In this paper, we propose a novel approach that manages heterogeneous flood-related data based on semantic web techniques and helps in limiting the damage caused by floods. We first propose an ontology that is used to formally describe the flood-related data, and we build our knowledge graph through integrating heterogeneous data using the proposed ontology. Then, we propose a reasoning approach using SHACL rules to infer new information that helps in managing the flood disaster or in anticipating future events. The experimental evaluations of our proposed approach are conducted on a real case study in the frame of flood disaster management with the aim of generating evacuation priorities. The results show that it succeeds in managing heterogeneous flood-related data and generating evacuation priorities in a very short time.

Keywords: Flood disaster management ; Semantic web ; Flood ontology ; Reasoning ; SHACL rules

1. Introduction

Natural disasters, such as floods, are known to be dangerous and adverse events resulting from natural processes of the Earth. They could result in damage of properties, infrastructures and economy. In addition to the materialistic damage, they could most importantly lead to loss of lives, disruption of normal life of the population leading to food and water deficiency, disturbance of communication, displacement of victims, etc. From here comes the urgent need of disaster management processes in order to handle such situations of flood disasters. The disaster management process involves heterogeneous data collected from different sources, and they need to be integrated and used in different phases of the disaster management process. The disaster-related data are usually difficult to manage due to its large amount of diversity and heterogeneity of its nature. Extracting information and making accurate inferences from various data sources quickly is critical for natural disaster preparedness and response. Critical information about disasters needs to be provided in a structured and easily accessible way in a context-specific manner (Sermet and Demir, 2019). Therefore, the challenge lies in managing these data in an efficient way in order to help in the process of disaster management and decision making.

Heterogeneous data can be structured and formalized in a knowledge graph that relies on a shared vocabulary represented in an ontology. An ontology is defined as a formal, explicit specification of a shared conceptualization (Studer et al. 1998). It allows a structuring and a logical representation of knowledge, expressing the explicit and implicit relations between concepts. The ontology is used to describe data in a knowledge graph which expresses the domain entities related to the considered data and their relations. This enables knowledge to be machine-processable for better information retrieval and analysis. Ontology-driven systems have gained popularity as they enable semantic interoperability, flexibility, and reasoning support (Schulz and Martínez-Costa, 2013), and they allow to overcome heterogeneity and to have a consistent shared understanding of the meaning of information (Elmhadhbi et al. 2019). Ontologies have been used in the domain of flood disaster management in order to integrate and share flood-related data as well as to infer new information from these data thus helping in the disaster management process.

The disaster management has been described in the literature as a life-cycle, and it is categorized into four main phases: mitigation, preparedness, response and recovery (Franke, 2011). This work is conducted in the frame of "ANR inondations" e-flooding project[1].

[1] https://anr.fr/Projet-ANR-17-CE39-0011

This project focuses on the mitigation and response phases where it aims at integrating several disciplinary expertises to prevent flash floods and to experiment the effects of decision making on two timescales: short-term and long-term. The short-term timescale aims at optimizing the disaster management process during the disaster, while the long-term timescale aims at improving the territories' resilience for risk prevention from five years to ten years after the disaster.

This paper aims at proposing a solution for limiting the damage caused by floods. The main contributions of this work consist of relying on semantic web techniques to manage heterogeneous flood-related data through proposing a flood ontology that formally describes the heterogeneous data. Using this ontology and reasoning rules, we then infer new information that can assist in the flood disaster management process or in anticipation concerning floods. A main advantage of our approach is that it can be reactive as well as predictive. It is reactive as it manages real data of an occurring flood, and the proposed reasoning rules can be activated or deactivated in the appropriate phases of the flood. In addition, the reasoning rules can also be executed on predictive data for the purpose of anticipating future events or information. For example, using our approach, we can generate evacuation priorities of the places containing victims to be evacuated, we can infer the number and types of vehicles needed for evacuation in different zones, or we can improve the resilience of a territory through identifying vulnerable stakes. In our work, we evaluate our approach using a case study which falls in the short-term timescale of flood management where it aims at helping in the process of evacuation of flood victims during a flood, as a disaster response, through generating evacuation priorities for the places impacted by a flood and containing victims. This paper is organized as follows. Section 2 presents the literature review about ontologies proposed in the domain of natural disasters and floods and their different uses. Section 3 presents our approach and our data and details our proposed ontology. Section 4 discusses the process of building our knowledge graph. The reasoning approach using rules is detailed in section 5, and the approach is evaluated in section 6. Finally, the conclusion and the future work are presented in section 7.

2. Literature Review

Ontologies have been widely proposed in the literature in various domains, including domains of natural disasters and floods. From the review of the literature in these domains, we notice that ontologies are proposed for different uses, integration of heterogeneous data provided from various sources, information management and sharing among different actors and reasoning to infer new information. We first discuss the different uses and the proposed ontologies; then, we analyze these works.

One use of the ontologies is the integration of disaster-related data provided from various sources. Several approaches proposed ontologies for integrating homogeneous or heterogeneous data from one or more data sources. Scheuer et al. 2013 propose an ontology-based risk assessment workflow that performs flood assessment based on a computational model and focuses on integrating and operationalizing "local knowledge" in the process of flood risk management, where they define the local knowledge as the knowledge comprising the preferences of stakeholders and decision makers. These preferences are expressed by describing the data through their proposed ontology. Their concepts include "hazard" that is reused from Monitor ontology (Kollarits et al. 2009) in order to define an event and "event" that is reused from SWEET ontolgy (Raskin and Pan, 2005) in order to define a flood. A "flood" is described by a "recurrence interval" and an "intensity", and it describes an "inundation" and a "hydrospherePhenomena" that are reused from SWEET ontology. In addition, they define "population" that encompasses the most vulnerable age classes including children and elderly, "material infrastructure" as a subclass of "built environment" already defined in Monitor ontology. They define an "element at risk", for example a residential building and they reuse the "vulnerability" of an element at risk from MONITOR ontology. They also reuse flood's "duration", "area" and "inundation depth" from SWEET ontology. The concepts defined in this proposed ontology are important for flood disaster management and evacuation including "population", "elements at risk" and "infrastructure". However, infrastructures in their ontology are managed in a separated manner and can't be regrouped to describe an aggregation of the infrastructure, while this concept is important especially when no precise information about a specific infrastructure exists. Wang et al. 2018 propose a hydrological sensor web ontology to integrate heterogeneous data provided from different sensors effectively during periods of natural disasters. Their ontology is based on three existing ontologies: Sensor, Observation, Sample, and Actuator (SOSA)[2], Time[3] and GEOSPARQL[4]. They use the concepts “sensor” and “observation” from SOSA ontology.

They extend it using Time ontology for integrating temporal concepts including “Temporal Entity” class and “Instant” and “Interval” sub-classes. Then, they extend it using GEOSPARQL ontology to add geospatial dimensions using concepts including “covers”, “crosses”, “meets” and “within”. This approach is interesting when the data mainly concerns sensors; however, the objective of our work is to manage data that are not restricted to sensors but are heterogeneous and are provided from various sources.

Another main use of the ontologies in the domain of natural disasters is information management and sharing among different actors involved in the disaster management process. Disaster-related data are usually of big amounts and are provided by different actors. In this context, ontologies are proposed with the aim of sharing data among different involved actors and solve the problem of interoperability of communication among them. Khantong et al.2020 propose a flood evacuation ontology for the purpose of improving the efficiency and effectiveness of information management in a disaster response and solving the problem of information sharing among different responders, organizers or processes handled by different systems in organizations. In their proposed ontology, they have concepts describing both static and dynamic data. They use concepts from the unified foundational ontology (UFO) (Guizzardi, 2005) to describe static data; their static concepts include “organization”, “area”, “flood event”, “flood evacuation” and “victim”. They use concepts from the Design and Engineering Methodology for Organizations (DEMO) ontology (Dietz, 2010; Sprengel et al. 2000) to describe dynamic data; their concepts describe production and coordination acts. It is an interesting approach as it allows managing static and dynamic data; however, the concepts describing the flood, victims and evacuation centers do not describe detailed data about them, and no detailed concepts are defined to describe the infrastructure that is important for flood evacuation. Yahya and Ramli, 2020 discuss that flood-related data may be inaccurate or unavailable if not regularly upgraded. Therefore, they propose an ontology for formally describing data provided from different actors involved in the disaster management process. They construct an ontology for each actor, and they aim at integrating all the ontologies in a global ontology. In their work, they propose only one ontology concerning evacuation centers for managing the flood victims. They reuse existing ontologies including SEMA4A for emergency notification systems accessibility (Malizia et al. 2010), a fire emergency management ontology (Nunavath et al. 2016), an ontology for accessible evacuation routes for emergencies (Onorati et al. 2014) and earthquake evacuation ontology (Iwanaga et al. 2011). Their proposed concepts describe general data about victims such as age, gender, address and number of victims as well as data about evacuation centers such as location and capacity, while there are no concepts describing infrastructure, population in infrastructures or elements at risk, while these concepts are important for an efficient evacuation process.

A third use of the ontologies in the domain of natural disasters is for conducting reasoning on knowledge graphs to infer new information related to the flood. Wang et al. 2018 propose an ontology (previously described in this section) and construct their knowledge graph and query it using SPARQL query language[5]. They then execute SWRL rules on the knowledge graph to infer flood phases from the precipitation of water level and observation data. Kurte, Durbha, King, Younan, and Potnis 2017 aim at understanding the dynamic spatio-temporal behavior of a flood disaster. For this aim, they construct an ontology that captures dynamically evolving phenomena. The ontology uses “SIIM” (Kurte et al. 2016) and “Time” ontologies to describe geospatial and time concepts. Geospatial concepts include “geospatial region” and time concepts include “time interval” and “time slice”. As a reasoning approach, they execute SWRL rules to retrieve image regions based on their temporal interval relations. The rules that are used in the literature for the reasoning have limitations including identification, execution order and activation. Therefore, we propose a reasoning approach that handles these limitations.

2 https://www.w3.org/TR/vocab-ssn

3 https://www.w3.org/TR/owl-time

4 https://opengeospatial.github.io/ogc-geosparql/geosparql11/index.html

5 https://www.w3.org/TR/rdf-sparql-query

The ontologies proposed in the literature provide concepts describing floods, victims, infrastructures, population, elements at risk and vulnerability. These are important concepts to be considered for flood evacuation; however, certain concepts are not detailed enough. For example, aggregating different categories of infrastructures to solve the problem of unavailable data is not considered in the literature. In addition, there isn't an ontology that manages static and dynamic data with considering spatio-temporal dimensions. In our proposed ontology, we aim at handling these limitations through our defined concepts and relations.

3. Approach presentation

The general purpose of our proposed approach is to manage heterogeneous flood-related data provided from different sources to help in the flood disaster response. The main interest is to assist in the decision making process of evacuation of victims of a flood by generating evacuation priorities to the places impacted by the flood. Figure 1 displays the different steps of our proposed approach. We propose a flood ontology that formally defines the vocabulary that are used for describing these data through defining concepts and relations among concepts. Then, we build a knowledge graph using the proposed ontology to integrate all heterogeneous data. After that, we reason over this knowledge graph to infer new information. This information represents evacuation priorities to each place of the study area impacted by the flood.

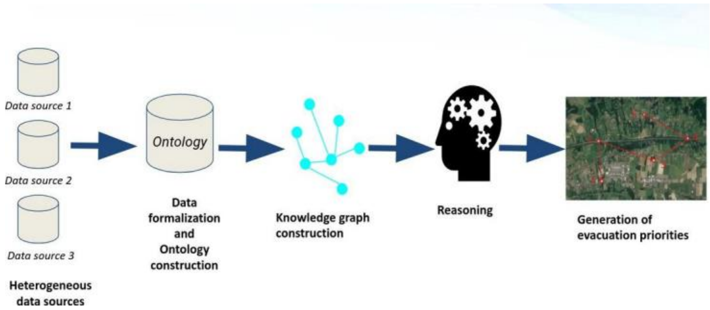


Fig. 1. Our proposed approach for generating evacuation priorities

3.1 Ontology design methodology

For the construction of our proposed flood ontology, we adopt a cyclic workflow that is proposed in NeOn design methodology (Suárez-Figueroa et al. 2015). This workflow is composed of four distinct phases: specification, conceptualization, formalization and implementation and finally evaluation. The specification consists of determining the domain, scope and purpose of the ontology as well as the use cases allowing to determine the role of the ontology. In the conceptualization phase, a glossary is built by adding definitions or common descriptions to each term. The terms are structured and then definitions of the basic concepts that are considered the most important and that align with the scope of the ontology are added. The relations are then modeled among concepts, which is the purpose of the formalization and implementation phase, during which a formal representation that is interpretable and computable by a machine is made. Finally, the developed ontology is tested against determined use cases in the evaluation phase. The results of these different steps are presented in the following sections. The ontology is written in OWL (Ontology Web Language)[6], and it is developed using Protégé[7], an ontology editor developed in Java that allows to load and save ontologies in most formal ontology representation languages, such as OWL.

6 https://www.w3.org/OWL/
7 https://protege.stanford.edu

3.2 Description of our data

The data available in our study concern the flood of Pyrénées that occurred in June 2013 in Bagnères-de-Luchon, south-western France. It was a torrential flood particularly destructive and very dangerous for the population. The consequences of this flood include destructed houses, cut roads, flooded campsites and damaged farms. 240 people were evacuated from the areas impacted by this flood. Figure 2 shows our study area.

The data of our study area are provided from various data sources. These sources include institutional databases such as BD TOPO[8] and GeoSirene[9]. They provide data about hazards, vulnerability, damage and resilience. Certain data sources provide data about geographical locations of roads, buildings, companies and establishments in France that are represented in QGIS[10], a cross- platform desktop geographic information system application that supports viewing, editing, and analysis of geospatial data. Other data are provided from various other data sources. These sources include data sensors providing data about water levels and flows, a hydrological model computing flood generation, a hydraulic model for flood propagation as well as other sources providing data about resilience corresponding to some actions taken from the past, socio-economic data and population data. Figure 3 shows the data used to feed our knowledge graph.

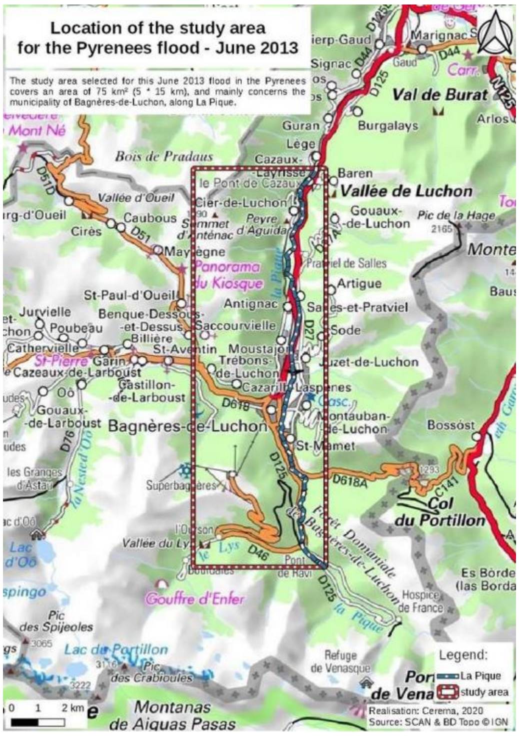


Fig. 2. Study area for Pyrénées flood in Bagnères-de-Luchon, France

8 https://www.data.gouv.fr/en/datasets/bd-topo-r/
9 https://data.laregion.fr/explore/dataset/base-sirene-v3-ss/
10 https://qgis.org/en/site

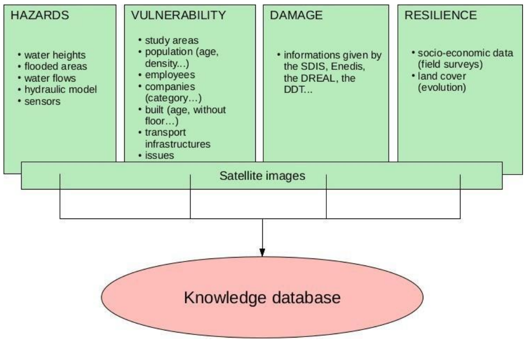


Fig.3. Data used to feed knowledge graph

Part of the data are accessible using QGIS. These data are categorized in two different formats representing two types of QGIS layers: raster layers consisting of masters of pixels and vector layers representing two-dimensional tables of various attributes. In the available data, there are data representing water levels that are the result of a hydraulic model, danger index of the flood, stakes, vulnerability and risk indexes. The calculation of these indexes are determined by the domain experts. These data can be divided into two categories: static and dynamic. The static data are those that don't change during the flood including number of floors, areas and geographic locations, while the dynamic data are those evolving throughout the flood including water level and population.

3.1 Ontology content

As our main objective is to manage flood-related data in order to generate evacuation priorities, we propose an ontology that contains the needed concepts, properties and relations that allow attaining this objective. The ontology representation is made through a hierarchy of classes (concepts) describing heterogeneous data involved in the flood phases, data properties which provide information about the characteristics of the classes and the relations which represent the object properties allowing to link the classes together.

We describe infrastructures in our proposed ontology as follows. A material infrastructure class is decomposed in several sub-classes: facility, habitat, working place and other kinds of infrastructure. A facility represents a commercial area, educational facilities such as schools, healthcare facilities such as hospitals and retirement homes as well as transport facilities such as railroad, railway station, road and tunnel. A habitat represents apartments, camping dwellings, hotels and houses. A working place represents administration sites, factories and offices. Another sub-class is also defined to describe any kind of infrastructure that is not one of the above. We detail all these different types of infrastructure in our ontology because it is important for the further reasoning process concerning the evacuation.

In addition to the material infrastructure class that allows managing each infrastructure on its own, we introduce a novel class, named infrastructure aggregation, which allows managing the different infrastructures in an aggregated manner by regrouping them in districts, buildings and floors. For example, we can describe that the district has buildings, the building has floors, and the floor has apartments by linking each two of them using the relation "has part". Thus, we can define for a district all the infrastructure that it contains. This relation is useful especially for the further step that consists of assigning evacuation priorities using rules because in the case when we only have general information about a district, we can assign a priority for all the infrastructures in this district.

The population class describes the population (number of persons) in all the infrastructures. It is divided into 2 sub-classes: fragile and non-fragile persons, and it is reused from the ontology proposed by Scheuer et al. 2013 with some additional details. The fragile population class is a defined class expressing that its instances are persons older that 65 years, children, and persons with disabilities, reduced mobility or illnesses. It represents the category of persons that need to have a high priority of evacuation when a flood occurs. Non-fragile population thus represents all the persons that are not fragile. The relation "is in" defines that a type of population is inside an infrastructure (or infrastructure aggregation). We define a demand point as a point that can be impacted by the flood and needs to be evacuated. It is similar to the class "element at risk" defined by Scheuer et al.2013 as a demand point can represent an element at risk. We thus define in our ontology a class named demand point that is a union of the two classes: infrastructure and infrastructure aggregation. Each infrastructure or infrastructure aggregation is considered as a demand point and thus a priority should be assigned to it in a further step. The demand point class has 4 sub-classes describing different priorities of evacuation: evacuate immediately, evacuate in 6 hours, evacuate in 12 hours and no evacuation. These 4 priorities are used in a further step for assigning one of them to each demand point based on rules defined according to the domain experts knowledge.

The data properties are divided into two categories, static and dynamic properties that add characteristics to the instances of demand point and population classes. The static properties concern the instances of the demand point class. They are usually determined before the crisis and allow to specify the structure of each demand point and its resilience to flooding. These properties represent building's vulnerability index, construction year, area and whether it has a basement or not They also include the floor and the number of floors when a building contains several floors as well as a property describing whether the building consists of only one floor. In addition, they include an important property that describes the geographic coordinates of a demand point, represented as "x" and "y" coordinates. The dynamic properties that evolve throughout the crisis concern the class "demand point" and represent the following characteristics: number of population, danger index, submersion height, flood duration, flood phases, number of fragile persons, whether the demand point is inhabited or not, number of population as well as approximate number of population when the exact number can't be determined.

Our proposed ontology consists of 41 classes, 6 object properties and 23 data properties. An overview of the ontology is displayed in Protégé editor is presented in figure 4. The ontology is available online via the following URL: https://www.irit.fr/recherches/MELODI/ontologies/i- Nondations.owl

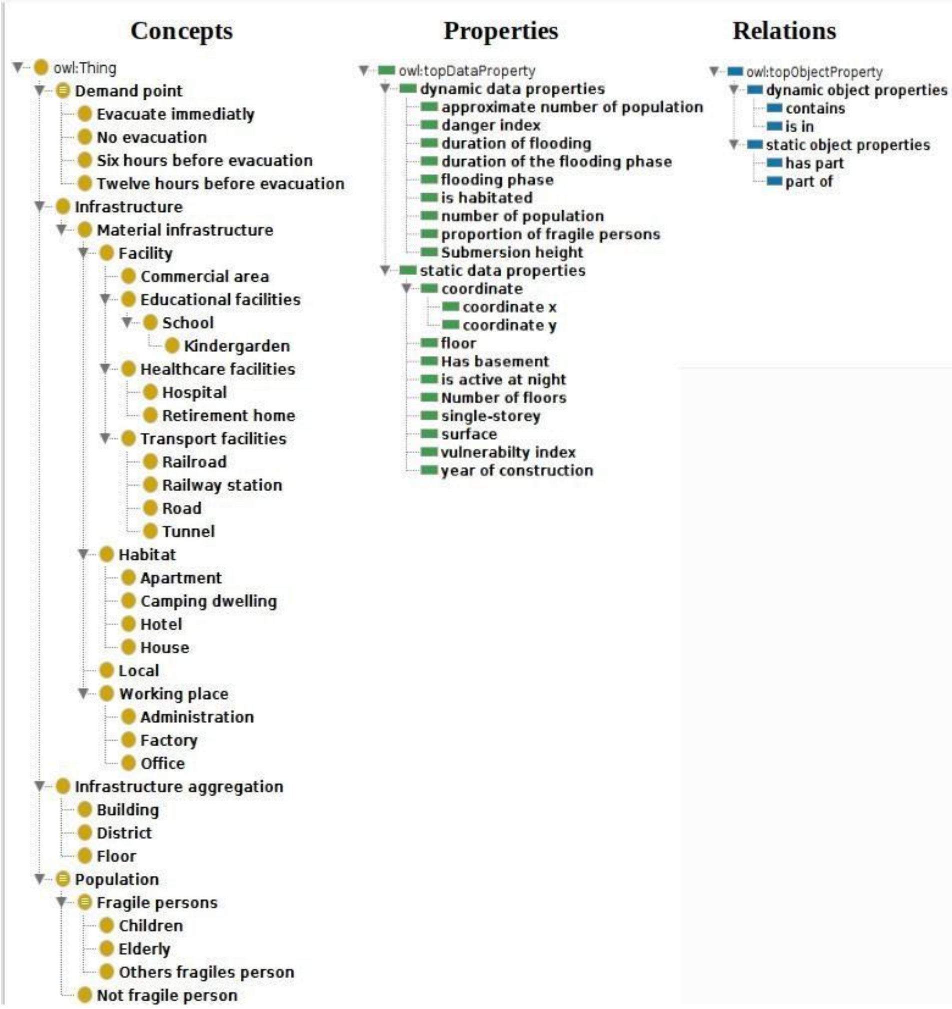


Fig. 4. Representation of our proposed ontology in Protégé

4. Knowledge graph construction

Building our knowledge graph represents integrating all the heterogeneous data using the ontology. In other words, it concerns integrating data through defining instances of the concepts and the properties of the ontology represented in the form of RDF triples. We recall that the heterogeneous data represent static and dynamic data. Therefore, we integrate static and then dynamic data to form our complete knowledge graph. Figure 5 presents the different steps of knowledge graph construction. First, we integrate static data by performing a joining process of these data from the different raster and vector layers containing them in QGIS as well as other sources such as BD TOPO. We then transform these integrated static data into static RDF triples that are added to the base ontology. In a further step, we integrate dynamic data by making a joining process for all the raster and vector layers containing these data in QGIS as well as other sources including a hydraulic model providing data about water level and flow speed and sources containing sensor and population data. We then transform the integrated dynamic data into dynamic RDF triples that are added to the base ontology and the rdf static triples, thus forming our knowledge graph that contains static and dynamic flood-related data. These processes are performed using “PyQGIS” API, the Python environment inside QGIS as well as other processes of data extraction adapted to each considered source, and “rdflib” python library is used to generate RDF data.

The last step consists of storing all the RDF triples constituting our knowledge graph in a triplestore, also named RDF store, that is a purpose-built database for the storage and retrieval of triples through semantic queries. We have chosen Virtuoso triplestore for storing the triples as it proves its efficiency in storing a big number of triples in a short time. For example, the results of a benchmark show that Virtuoso is able to load 1 billion RDF triples in 27 minutes while other triplestores take hours to load them such as BigData, BigOwlim and TDB[11].

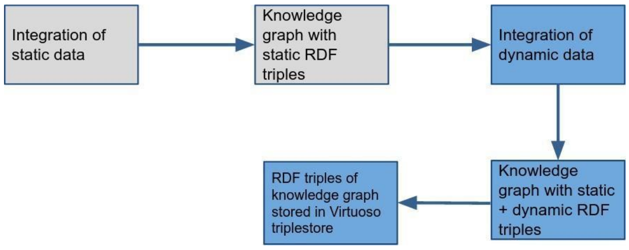


Fig. 5. Knowledge graph construction

5. Semantic reasoning: generation of evacuation priorities

We recall that our purpose is to assist in the decision making process of evacuation of victims as a flood disaster response. After building our ontology that formally describes the flood-related data and the knowledge graph that integrates the data using the ontology, the next step is to conduct a reasoning process that allows inferring new information in the form of evacuation priorities assigned to each demand point. Rules are used in the literature for inferring new information; however, they have certain limitations. They do not have identifiers, and thus a rule can not be well identified. It is also not possible to determine an execution order using these kinds of rules in case there is more than one rule. In addition, as long as the rules exist, they are automatically activated and can not be deactivated upon need. In our work, we use a more recent kind of rules, named SHACL rules[12] that are used for the same purposes and overcome these limitations; in addition, they haven’t been used in the domain of flood disaster management yet.

SHACL (Shapes Constraint Language)[13] is a World Wide Web Consortium (W3C) standard language that defines an RDF vocabulary to describe shapes, that are collections of constraints that apply to a set of nodes. One focus area of SHACL is data validation; however, the same principles of describing data patterns in shapes can also be exploited for other purposes. One purpose is using SHACL rules to derive inferred RDF triples from existing asserted triples using SHACL rules engine. A SHACL rules engine is a computer procedure that takes as input a data graph and a shapes graph and is capable of adding triples to the data graph. A SHACL rule is identified through a unique Internationalized Resource Identifier (IRI) just like any resource in contrary to other kinds of rules. In addition, it can be activated or deactivated based upon its usage purpose where a deactivated rule is ignored by the rules engine and is thus not executed. An order of execution can also be determined for SHACL rules when more than one rule is executed. SHACL rules allow not only to infer new information but also to enrich the knowledge graph with this inferred information. There are different types of SHACL rules including SPARQL rules that allow writing rules in SPARQL notation. We use SPARQL rules in our reasoning process for generating evacuation priorities. We define a rule for each evacuation priority as follows. First, we define the node shapes representing the classes that describe the priorities and the property shapes representing the properties used to define the evacuation priorities. After that, we define our rules. Each rule contains the conditions that need to be satisfied for each property (detailed later).

The rules are then executed on the knowledge graph to infer new triples. Each inferred triple consists of a demand point with an evacuation priority assigned to it according to its properties. The knowledge graph is then enriched by adding these inferred triples to it in the triplestore.

11 http://wbsg.informatik.uni-mannheim.de/bizer/berlinsparqlbenchmark/results/V7/#exploreVirtuoso
12 https://www.w3.org/TR/shacl-af
13 https://www.w3.org/TR/shacl

Various priorities to be assigned to the demand points are determined. First, a set of properties that define each evacuation priority is determined; then, the conditions that need to be satisfied for each priority are set. The properties determining the evacuation priorities represent certain static and dynamic data. The evacuation priorities and the conditions defining them are determined by domain experts, that are the firefighters concerned in the evacuation process, as well as other factors including the study area and the available data. Although a specific set of properties define the evacuation priorities in this work, it is dynamic and can be changed as our approach is generic, and it is not restricted to a specific study area or data nature.

6. Approach Evaluation

In this section, we discuss the evaluation of our proposed approach using the available data of our study area starting from building our proposed ontology to constructing the knowledge graph until generating the priorities using SHACL rules and enriching the knowledge graph with the inferred triples.

6.1 Our Knowledge graph construction

As discussed in section 3, static data are first integrated and transformed into RDF triples. The processes of joining the static data and transforming them to RDF triples are executed in 24 seconds. 245,644 static triples are generated and added to the triples representing the ontology. After that, the dynamic data are integrated and transformed to RDF triples in 174 seconds. 225,766 dynamic triples are generated and added to those of the ontology and the static data thus forming our knowledge graph.

After the knowledge graph construction, our next step is to conduct a reasoning process using SHACL rules to infer new information in the form of evacuation priorities. First, we determine the different types of evacuation priorities with the used properties and their conditions. Then, for each kind of evacuation priority, we define a SHACL rule that allows, when executed, to assign for each demand point an evacuation priority based on its properties.

6.2 Determining evacuation priorities

The domain experts involved in our project have determined four types of evacuation priorities: evacuate immediately, evacuate in 6 hours, evacuate in 12 hours and no evacuation. The definition of these evacuation priorities relies on several properties representing static and dynamic data. The properties representing static data are: number of floors and vulnerability index, while the properties representing dynamic data are: submersion height, danger index, duration of flood and number of population in a demand point. The vulnerability index is calculated by joining different topographic and social data such as population density, building quality and socio-economic conditions. The danger index is calculated by joining the speed of water flow and the level of water obtained from a hydraulic model.

```
Evacuate in 12h if:
Demand_point contains Population
0 < danger_index < 50
duration_of_flood > =12
number_of_floors >= 1
0.0 < submersion_height <= 1.0
vulnerability_index < 50.0
is_habitated is True
```

Each type of evacuation priority is defined through setting conditions on the considered properties, and each demand point that satisfies the conditions of a certain type of evacuation priority is then assigned this priority. For example, if a demand point represents an infrastructure of high priority, such as a school or a hospital, and if it contains population, then it is determined to be evacuated immediately as is thus assigned the priority “evacuate immediately”. On the other hand, if a demand point doesn’t contain population or has a danger index or a submersion height <= 0, then it is considered that there is no need to evacuate it and is thus assigned the priority “no evacuation”. An example of the priority “evacuate in 12 hours" with its properties and conditions is presented as follows. Similarly, the three other evacuation priorities are defined each having its own conditions.

6.3 SHACL rules for generating evacuation priorities

For each type of evacuation priority, we implement a SHACL rule that defines its conditions as follows. We start by defining the properties that are used in the rules. Then, we define 4 rules where each rule represents a type of evacuation priority and contains the conditions set for each property. An example of a rule defining the type of evacuation "evacuate in 12 hours" is as follows.

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix sh: <http://www.w3.org/ns/shacl#>.
@prefix ns1: <https://www.irit.fr/recherches/MELODI/ontologies/i-Nondations.owl#>.
ns1:12h_before_evacuationRulesShape
rdf:type sh:NodeShape ;
sh:targetClass ns1:Demand_point ;
sh:rule [
rdf:type sh:SPARQLRule ;
sh:prefixes ns1: ;
sh:construct """
PREFIX ns1: <https://www.irit.fr/recherches/MELODI/ontologies/i-Nondations.owl#>
CONSTRUCT
{?this ns1:priority ?priority.}
WHERE
{  ?this ns1:danger_index ?danger_index.
   ?this ns1:duration_of_flooding ?duration_of_flooding.
   ?this ns1:number_of_floors ?number_of_floors.
   ?this ns1:submersion_height ?submersion_height.
   ?this ns1:vulnerability_index ?vulnerability_index.
   ?this ns1:is_habitated ?is_habitated.
   FILTER
   (?danger_index > 0 && ?danger_index < 50
   && ?duration_of_flooding >= 12
   && ?number_of_floors >=1
   && ?submersion_height > 0.0 && ?submersion_height <= 1.0
   && ?vulnerability_index < 50.0
   && ?is_habitated = true ) .
   BIND ("12h_before_evacuation"  AS ?priority).
   }
    """ ;
    sh:condition ns1:12h_before_evacuation ;
];
```

The rules are executed using TopBraid SHACL API[14], an open source implementation of the W3C SHACL based on Apache Jena[15]. We execute our rules on the knowledge graph and infer triples representing evacuation priorities assigned to every demand point in the knowledge graph according to their properties. The execution order of the rules is not important in our case as the interest is assigning priorities to every demand point in our study area.

---

[13] https://github.com/TopQuadrant/shacl
[14] https://jena.apache.org/

There are 15,078 demand points in our study area; therefore, 15,078 new triples are generated with corresponding evacuation priorities for demand points. Table 1 shows the number of triples generated after each process in our approach.

Table 1. Number of triples generated after each process in our approach

Process	Knowledge graph with static data	Knowledge graph with static and dynamic data	Complete knowledge graph with priorities
Number of triples	246,828	472,594	487,672

The SHACL rules were executed in 9.96 seconds which represents the time of assigning priorities to all demand points. Figure 6 displays the execution time of various processes in our proposed approach.

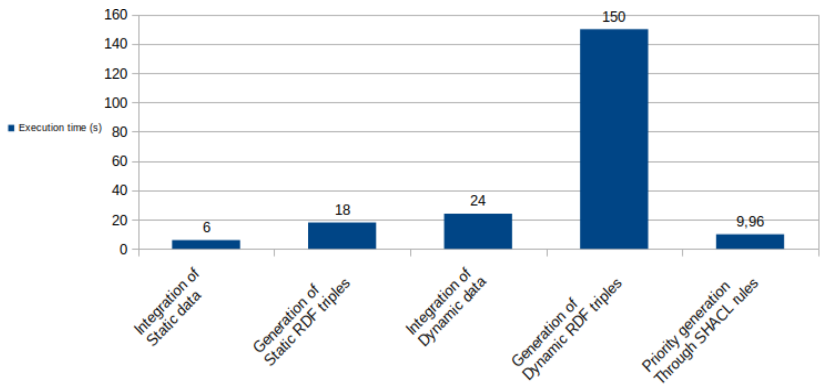


Fig. 6. Execution times of various processes

An example of a newly inferred triple is as follows.

```
<https://www.irit.fr/recherches/MELODI/ontologies/i-Nondations.owl#DP_12345>
 <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
  <https://www.irit.fr/recherches/MELODI/ontologies/i-Nondations.owl#evacuate_
  in_12h>
```

This triple represents an evacuation priority “evacuate-in-12h” assigned to the demand point “DP_12345”. The properties of this demand point are as follows:

```
"DP_12345":
danger_index = 25
duration_of_flood = 24
number_of_floors = 3
submersion_height = 1.0
vulnerability_index =  7.02
is_habitated = "true"
```

As we can see, the properties of this demand point satisfy the conditions of the evacuation priority “evacuate in 12h”; therefore, this type of priority is assigned to it.

These results show that our proposed approach succeeds in attaining its objective. The construction of the knowledge graph including all the static and dynamic data of our study area only takes 198 seconds, and the reasoning process generating evacuation priorities for each demand point in our study area only takes 9.96 seconds. Responding to the flood disaster in a short time is important especially concerning evacuating victims and saving lives. Therefore, our approach succeeds in assisting in the flood disaster response phase through generating evacuation priorities to demand points in a short time and thus helping the firefighters in the decision making process concerning the evacuation.

6.4 Visualization of evacuation priorities in QGIS

As previously discussed, the reasoning process allows inferring new information in the form of evacuation priorities to demand points in our study area. The knowledge graph is then enriched by adding the newly inferred triples to it. Figure 7 shows the data visualization represented in QGIS after adding the evacuation priorities. As shown in the figure, the evacuation priorities of the demand points are displayed in different colors as follows: “Evacuate-immediately” is displayed in red, “Evacuate-in- 6h” is displayed in blue, “Evacuate-in-12h” is displayed in yellow and “No-evacuation” is displayed in green. In our study area, the number of demand points assigned an evacuation priority “Evacuation-in- 6h” is very small with respect to the others that are assigned other evacuation priorities (31 out of 15,078 demand points). Therefore, in the figure, we only choose to display the three other evacuation priorities. The gray and black zones represent water levels; they are thus the zones where we have demand points represented in red and need to be evacuated immediately.

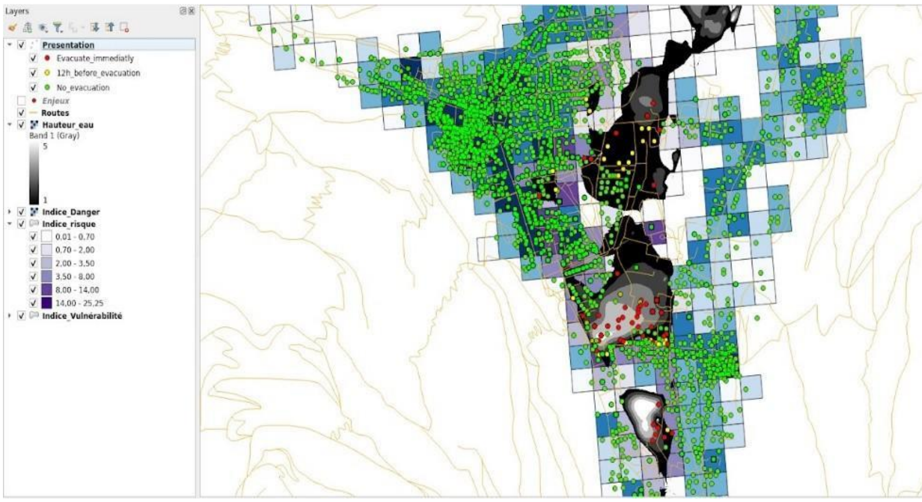


Fig.7. Visualization of data layers in QGIS with evacuation priorities

7 Conclusion

In this paper, we have proposed a novel approach of managing heterogeneous flood-related to assist in the flood disaster management process When a flood occurs, there is an urgent need for a flood disaster response concerning victims’ evacuation to save their lives in an efficient manner. Therefore, we have proposed an approach that manages flood-related data and generates evacuation priorities of flood victims to help the firefighters in the decision making process of evacuation. We have thus proposed an ontology that formally describes heterogeneous flood-related data; then we have built our knowledge graph through integrating static and dynamic data available in our study area. A reasoning process have then been conducted on the knowledge graph to assign evacuation priorities for all the demand points. The reasoning process is performed using SHACL rules that are more recent and advantageous over other kinds of rules already used in the literature. The experimental results prove that our approach succeeds in generating evacuation priorities for all demand points of the study area in a short time. As a future, work, we aim at proposing an interface that helps domain experts to elaborate the reasoning rules according to available data during the different flood stages, to activate or deactivate them and to provide an execution order to the rules upon need. Furthermore, this interface would allow executing rules on past data to improve the flood disaster management, on real-time data to manage a current flood disaster and on predictive data to anticipate future events or actions.

Acknowledgments

This work has been funded by the ANR[16] in the context of the project "i-Nondations" (e-Flooding), ANR-17- CE39-0011.

References

Dietz, J. (2010). *Enterprise ontology: theory and methodology*. Springer-Verlag Berlin Heidelberg.

Elmhadhbi, L., Karray, M. H., & Archimède, B. (2019). "A modular ontology for semantically enhanced interoperability in operational disaster response". In *16th International Conference on Information Systems for Crisis Response and Management-ISCRAM 2019,* pp. 1021-1029.

Franke, J. (2011). "Coordination of Distributed Activities in Dynamic Situations. The Case of Inter- organizational Crisis Management". PhD thesis. UniversitéUniversité Henri Poincaré-Nancy I).

Guizzardi, G. (2005). "Ontological Foundations for Structural Conceptual Models". PhD thesis.

Centre for Telematics and Information Technology.

Iwanaga, I. S. M., Nguyen, T.-M., Kawamura, T., Nakagawa, H., Tahara, Y., and Ohsuga, A. (2011). "Building an earthquake evacuation ontology from twitter". In: *2011 IEEE International Conference on Granular Computing*. IEEE, pp. 306–311.

Khantong, S., Sharif, M. N. A., and Mahmood, A. K. (2020). "An ontology for sharing and managing information in disaster response: An illustrative case study of flood evacuation". In: *International Review of Applied Sciences and Engineering*.

Kollarits, S., Wergles, N., Siegel, H., Liehr, C., Kreuzer, S., Torsoni, D., Sulzenbacher, U., Papez, J., Mayer, R., Plank, C., et al. (2009). "MONITOR–an ontological basis for risk management". In: *Tech. Rep., Monitor*.

Kurte, K. R., Durbha, S. S., King, R. L., Younan, N. H., and Potnis, A. V. (2017). "A spatio-temporal ontological model for flood disaster monitoring". In: *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, pp. 5213–5216.

Kurte, K. R., Durbha, S. S., King, R. L., Younan, N. H., and Vatsavai, R. (2016). "Semantics-enabled framework for spatial image information mining of linked earth observation data". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 10.1, pp. 29–44.

Malizia, A., Onorati, T., Diaz, P., Aedo, I., and Astorga-Paliza, F. (2010). "SEMA4A: An ontology for emergency notification systems accessibility". In: *Expert systems with applications* 37.4, pp. 3380– 3391.

Nunavath, V., Prinz, A., Comes, T., and Radianti, J. (2016). "Representing fire emergency response knowledge through a domain modelling approach". In: *Norsk konferanse for organisasjoners bruk at IT*. Vol. 24. 1.

Onorati, T., Malizia, A., Diaz, P., and Aedo, I. (2014). "Modeling an ontology on accessible evacuation routes for emergencies". In: *Expert Systems with Applications* 41.16, pp. 7124–7134.

Raskin, R. G. and Pan, M. J. (2005). "Knowledge representation in the semantic web for Earth and environmental terminology (SWEET)". In: *Computers & geosciences* 31.9, pp. 1119–1125.

Scheuer, S., Haase, D., and Meyer, V. (2013). "Towards a flood risk assessment ontology–Knowledge integration into a multi-criteria risk assessment approach". In: *Computers, Environment and Urban Systems* 37, pp. 82–94.

Schulz, S. and Martínez-Costa, C. (2013). "How ontologies can improve semantic interoperability in health care". In: *Process Support and Knowledge Representation in Health Care*. Springer, pp. 1– 10.

Sermet, Y. and Demir, I. (2019). "Towards an information centric flood ontology for information management and communication". In: *Earth Science Informatics* 12.4, pp. 541–551.

[13] https://anr.fr/Projet-ANR-17-CE39-0011

Sprengel, C., Baumann, K.-H., and Neuer, S. (2000). “Seasonal and interannual variation of coccol- ithophore fluxes and species composition in sediment traps north of Gran Canaria (29 N 15 W)”. In: *Marine Micropaleontology* 39.1-4, pp. 157–178.

Studer, R., Benjamins, V. R., and Fensel, D. (1998). “Knowledge engineering: Principles and methods”. In: *Data & knowledge engineering* 25.1-2, pp. 161–197.

Suárez-Figueroa, M. C., Gómez-Pérez, A., and Fernandez-Lopez, M. (2015). “The NeOn Methodology framework: A scenario-based methodology for ontology development”. In: *Applied ontology* 10.2, pp. 107–145.

Wang, C., Chen, N., Wang, W., and Chen, Z. (2018). “A hydrological sensor web ontology based on the SSN ontology: A case study for a flood”. In: *ISPRS International Journal of Geo-Information* 7.1, p. 2.

Yahya, H. and Ramli, R. (2020). “Ontology for Evacuation Center in Flood Management Domain”. In: *2020 8th International Conference on Information Technology and Multimedia (ICIMU)*. IEEE, pp. 288–291.

Track – Medical Knowledge Management

Psychiatric disorders among opioid dependents: socioeconomic and gender difference

YONG-MI KIM

School of Library and Information Studies, University of Oklahoma, 4502 E. 41st St.

Tulsa, Oklahoma 74135,

United States of America

yongmi@ou.edu

The psychiatric disorder rates among opioid dependents have alarmingly increased over the last decades, and these disorders are higher for women than men and higher for individuals in low socioeconomic groups. Previous knowledge asserted that opioids had no addictive or harmful effects that could lead to psychiatric disorders, but the recent discovery of opioid-related knowledge reversed the existing belief. The purpose of this research is to discover how the new knowledge has changed regarding psychiatric disorders from opioids between men and woman and across socioeconomic groups. In order to uncover these changes, the research data is obtained from the Treatment Episode Data Set (TEDS) in 2007 and 2017. While the percentage of psychiatric disorders among opioid dependents is higher for women than men, unexpectedly the growth rate of psychiatric disorders for men is much faster than women. As such, the socially constructed conventional knowledge that psychiatric disorders are women's illness will change in the near future. Congruent with existing knowledge, psychiatric disorders in this dataset are higher for underprivileged brackets such as those with low education, unemployed, separated, divorced, and widowed people, and Medicaid recipients, and the growth rates for these groups are steeper than their counterparts.

1. Introduction

In 2016, approximately 9.8 million adults were diagnosed with a serious mental illness, and 1.6 million among them were opioid misusers (SAMHSA, 2017a). Statistics published in 2007 and 2017 in Treatment Episode Data Set Admissions (TEDS) show that the increase in whole substance users was only 2.02%, but the rate increased for opioid dependents during the same period is four times higher (5.0% in 2007 to 20.1% in 2017). The same source also reported that the psychiatric disorders among opioid dependents more than doubled during the same period (17.7% to 38.2%). Opioid dependence is defined as a manifestation of brain changes resulting from chronic opioid abuse (Kosten & George, 2002). Psychiatric disorder is defined as a *co-occurring mental disorder* (Department of Health and Human Services). Mental disorders include depression, anxiety, and bipolar disorder among many other mental issues, and they are directly linked to an increase in opioid misuse (SAMHSA, 2017b; Davis et al., 2017; Wasan et al., 2015; Jones & McCance-Katz, 2019; Sullivan et al., 2005).

Individuals suffering from opioid use disorder (OUD) have significantly higher rates of mental illnesses than any other substance users (Davis et al., 2017; Rijswijk et al., 2019; Merrill et al., 2012). Adults with pre-existing mental disorders are four times more likely to use opioids than those without mental disorders (Davis et al., 2017). Among people with mental disorders, individuals with OUD are three times more likely to have a mood disorder and four times more likely to have an anxiety disorder compared to those without OUD (Conway et al., 2006). This is partially because patients with mood and anxiety disorders experience a greater likelihood of receiving an opioid prescription, and consequently, they are more apt to develop problems with opioid use (Conway et al., 2006). This finding is confirmed by the review of four longitudinal studies, all concluding that depression and anxiety disorders were associated with problematic opioid use (Rijswijk et al., 2019; Merrill et al., 2012).

Literature further shows that psychiatric disorders among individuals with OUD are disproportionally represented among lower socioeconomic statuses *and* women. This may be due to the reality that individuals in lower socioeconomic brackets often do not have access to the recent scientific knowledge developments pertaining to opioids, and women are socially perceived to have higher rates of mental issues, and doctors tend to prescribe women and men differently, resulting opioid prescription for women is higher than men (Serdarevic et al., 2017; Kim and Delen, 2015; Irwin, 2010). The common knowledge used to be that there were no addictive side effects or mental health disorders caused from opioid use; a stark contrast to the scientific knowledge that exists today. However, there has not been comprehensive research on how the new knowledge has changed opioid dependents' psychiatric disorders. Therefore, this study aims to investigate how the opioid-related new knowledge has (1) impacted psychiatric disorders between men and women and (2) changed psychiatric disorders in varying socioeconomic brackets.

2. Gender Differences

Gender disparities for psychiatric disorders can be viewed from a socialization process perspective. Socially constructed knowledge of masculinity can result in health disparities (Kim & Delen, 2017). Boys and girls are generally socialized to express pain differently. Boys are often taught to be tough, tolerate pain, and sustain painful experiences while girls are socialized to be sensitive, careful, and to verbalize discomfort (Myers et al., 2003). The main attributes of masculinity are described as strength, endurance, and stoicism while femininity is associated with sensitivity and pain expression (Bernardes et al., 2008). Since men often perceive pain as an ailment more frequently experienced by women, expressing or recognizing pain is commonly considered a threat to their masculinity (Lack et al., 2011). Since pain is subjective, a socialization process and an opposing gender knowledge influence a different pain perception and expression of pain, which eventually leads to variations in gender-specific illness symptoms, risk behaviors, and use of medical services (Samulowitz et al., 2018; Bartley & Fillingim, 2013; Stock et al., 2008). A meta-analysis confirms that masculinity and high pain tolerance are correlated, as well as illustrating a correlation between femininity and high pain sensitivity (Alabas et al., 2012; Nielsen et al., 2011). Doctors prescribe men and women differently, and opioids are more likely to be prescribed for women than men (Serdarevic et al., 2017; Kim and Delen, 2015). As such, one can expect women to have higher rates of psychiatric disorders from opioids than their male counterparts.

3. Socioeconomic Groups

3.1 Educational Level

Education is directly related to income levels and substance abusers frequently lack education and professional training (Ellis et al., 2020; Atkinson et al., 2001). Patients with higher levels of education tend to have knowledge about the medical consequences of opioid abuse and are less likely to receive opioids during an emergency department (ED) visit compared to their counterparts (Platt-Mills, 2012). One study compared the educations of treatment-seeking opioid users and the US Census population of those 25 years or older between 2010 and 2018 and found that the highest numbers of opioid treatment-seeking patients are among those with high school/GED degrees (39.5%). Those with bachelors' degree or higher (6.3%) show the lowest levels of opioid treatment-seeking (Ellis et al., 2020). Because educated individuals are more likely to search valid medical information online, and because the healthcare industry is increasingly sending out updated health information online (Kim, 2015), they are likely to read and understand the most recent opioid-related information on the psychiatric disorders from opioids and are thus less likely to suffer psychiatric disorders from opioid misuse than their counterparts.

3.2 Employment Status

Employment status also correlates to substance/opioid use (Altekruse et al., 2020; Svikis et al., 2012). As a county's unemployment rate increases by one percentage point, the opioid death rate rises by 3.6%, and the opioid overdose emergency department (ED) visit rate increases by 7.0% (Green et al., 2003). Correlations exist between high rates of opioid consumption and lower labor force participation (Krueger, 2017). This may be due to the fact that opioid-dependent individuals often lose previously held jobs (Platt, 1995; Kidorf et al., 1994), or they face obstacles in finding employment (Magura, 2003; Svikis et al, 2012). Furthermore, opioid-dependent individuals frequently experience decreased motivations to secure employment (Zanis et al., 2001). Because their employment situation affects their mental health, they turn to opioids to treat their depressive conditions (Conway et al., 2006). Since individuals with a secure job have mental and financial resources to rely on to protect their health, they are less likely to abuse opioids and thus present lower levels of psychiatric disorders.

3.3 Marital Status

Divorce is often a powerful factor in the onset of drug abuse, even after controlling for variables such as a family history of substance misuse (Scott et al, 2010; Edwards et al., 2017). Being previously married (versus stably married) was associated with the increased risk of all psychological disorders (Scott et al., 2010). This is because divorced or separated people sometimes lose their financial stability, placing them into a lower socioeconomic bracket (Kim & Delen, 2017), and they experience significant stress and psychological instability, which often leads them to opioid misuse (Lanier et al., 2012). On the other hand, stably married individuals will enjoy psychological stability and stable income to take care of their health, and they may also feel responsible for a family. As such, the psychiatric disorders from opioids are expected to change depending on the marital status.

3.4 Health Insurance

Insurance types have high correlations with opioid prescriptions and misuses. An individual's status as a Medicaid recipient, which are usually those in low socioeconomic brackets, is associated with a higher risk for prescription opioid abuse and overdose (CDC, 2019). This is related to insurance reimbursements, as Medicaid covers a disproportionate number of nonelderly adults with opioid addiction. Furthermore, Medicaid covers an even greater share of those with low incomes, higher dosages, and longer durations of prescription compared to other insurance types (Platt-Mills, 2012; Orgera & Tolbert, 2019; Whitmire & Adams, 2010; NIDA, 2017). As such, approximately 55% of nonelderly low- income adults with OUD were on Medicaid (Orgera & Tolbert, 2019). On the other hand, hydrocodone prescriptions for Medicare patients decreased from 21.9% to 18.3% between 2013 and 2015 (Kuo et al., 2018). This source also reported that the risk of receiving prolonged opioid prescriptions decreased by roughly 7% for Medicare patients during the same period. It is clear that the insurance provided for low socioeconomic groups plays an important role for psychiatric disorders from opioids. Because different insurance types cover or do not cover their patients' treatments using opioids, psychiatric disorders from opioids will be very different between these two insurance types.

4. Research Methods

4.1. The Sample

Data are from the Treatment Episode Data Set (TEDS) 2007 and 2017. This dataset is ideal for this study because TEDS measures psychiatric disorders among opioid dependents using the same measurements over time. These data sets are a repository of treatment data collected by states in order to monitor substance use treatment systems (SAMHSA, 2017a). The total number of substance users in TEDS was 1,965,194 in 2007 and 2,005,375 in 2017— approximately a 2% increase. From all the substance users in the data set, only opioid dependents (ICD-9: 304.00 – 304.09 in 2007; ICD-10: F11.2 – F11.23 in 2017)
were extracted. This number was 93,600 in 2007 and 388,008 in 2017, an increase of 314.5% during the ten-year period.

When investigating psychiatric disorders and their relation to opioid dependence, gender plays an integral role in the data because of socially constructed knowledge about men and women. Since there are more opioid-dependent men than women, men will naturally exhibit a higher percentage of psychiatric disorders if not controlled. Because the purpose of this study is to investigate how psychiatric disorders are concentrated in a specific subgroup, the calculation of psychiatric disorders from opioids is controlled for gender difference and different socioeconomic strata. After controlling for gender difference, psychiatric disorder rates are higher for females than men in both 2007 and 2017, and their findings are statistically highly significant as shown in Table 1. However, the increase rate is slightly higher for men than women during the research period. More specifically, the increase in psychiatric disorders over the ten-year period is 74% and 69% for men and women.

Table 1: Sample distribution

Year	Men	Women	Sig.
Psychiatric disorder ratio among opioid dependents			
2007	37,614+ (8,108, 21.6%)	29,008+ (8,137, 28.1%)	***
2017	214,926 (80,836, 37.6%, 74.01% ↑)	147,960 (70,204, 47.5%, 69.0 % ↑)	***

+ missing data for gender and psychiatric disorder removed. ***: $p<0.000$.

4.2. Analytical Strategies

The dependent variable is the psychiatric disorder, which is measured by whether or not the admitted opioid dependents experience any psychiatric disorders. The independent variables are the identified groups representing socioeconomic status (i.e., employment status, insurance type, marital status, and education level). Gender differences are also compared across the socioeconomic groups. Because the data were already collected, the operationalization of psychiatric disorders is based on TEDS' measurement, using either the Diagnostic and Statistical Manual of Mental Disorders (DSM) from the American Psychiatric Association or the International Classification of Diseases (ICD), from the World Health Organization (2007, 2017). Because the independent and dependent variables are all categorical, analysis of variance (ANOVA) is employed. Gender and socioeconomic groups are controlled to directly compare opioid dependents' rate within the respective group and the rate increase over the research period. The educational level is categorized into low and high in order to clearly compare the effects of education on psychiatric disorders from opioids. Insurance types included Medicaid and Medicare to compare the policy effects on psychiatric disorders from opioids.

5. Results and Discussion

The statistical findings of the comparison between male and female are highly significant across all the categories ($p<0.001$). As such, statistical significances are not reported in the appendix table.

Table 2: Report of statistical findings

Category	Year	Men	Women
Psychiatric disorder ratio by educational level			
8 and less	2007	29,018 (5,731, 19.8%)	20875 (5,219, 25.00%)
13 and more		35,251 (7,545, 27.3%)	7,720 (2,741, 35.5%)
8 and less	2017	161,896 (62,559, 38.6%, 95↑)	106,156 (49,948, 47.0%, 88% ↑)
13 and more		38611 (15,906, 41.2%, 51% ↑)	35428 (18,133, 51.2%, 44% ↑)
Psychiatric disorder ratio by employment status			
Employed	2007	5,709 (1,917, 31.5%)	2,470 (896, 36.3%)
Unemployed		21425 (4,113, 19.2%)	17,081 (4,571, 26.8%)
Employed	2017	31,668 (10,200, 32.2%, 2% ↑)	11,614 (5,073, 43.7%, 20% ↑)
Unemployed		85,834 (33,348, 38.9%, 103% ↑)	60,089 (28,819, 48.0%, 79% ↑)
Psychiatric disorder ratio by marital status			
Married	2007	5,729 (1,531, 26.7%)	4,726 (1,631, 34.5%)
Separated, divorced, widowed		6,792 (1,598, 23.5%)	7,001 (2,358, 33.7%)
Married	2017	23,557 (8,063, 34.2%, 28% ↑)	17,353 (7,914, 45.6%, 32% ↑)
Separated, divorced, widowed		29,807 (12,303, 41.3%, 76% ↑)	28,000 (14,586, 52.1%, 55% ↑)
Psychiatric disorder ratio by insurance type			
Medicaid	2007	3,190 (316, 9.9%)	4,499 (801, 17.8%)
Medicare		1,183 (505, 42.7%)	818 (388, 47.4%)
Medicaid	2017	76,821 (27,413, 35.7%, 261% ↑)	63,464 (28,996, 45.7%, 157% ↑)
Medicare		8,921 (2,273, 25.5%, 40% ↓)	6,010 (2,105, 35.5%, 25% ↓)

[+]This number is derived after removing the missing values for psychiatric problems. As such, there is a discrepancy between the opioid dependents and the total number of psychiatric problems calculation.
Since all of the differences are statistically highly significant (p<.000), the statistical significance from ANOVA is not included in the table.

5.1. Psychiatric Disorder Changes across Education Levels

Table 2 reveals that psychiatric disorders from opioid use between low and highly educated groups present a more than two-fold difference. With regard to the percentage, low- educated people regardless of gender show lower levels of psychiatric disorders; however, psychiatric disorders in this group of individuals are rapidly increasing. More specifically, the most rapidly growing psychiatric disorders from opioids are observed among the educational level 8 years or less, which increased by 95% for males and 88% for females. On the other hand, the highly educated groups (13 years and more) increased only by 51% for males and 44% for females. This finding supports the argument that highly educated individuals may be familiar with the opioid-related recent knowledge and mindful of the opioid-born psychiatric problems and thus reduce their consumption. This finding is also consistent with the finding that highly educated individuals are less likely to receive opioid prescriptions during an ED visit than their counterparts (Platt-Mills, 2012) and are less likely to have psychiatric disorders as a result. It is interesting though that females have higher rates of psychiatric disorders during the research period, but the rate increase is smaller than that of the male group.

5.2. Psychiatric Disorder Changes across Employment Statuses

Employment status is highly correlated with psychiatric disorders. When psychiatric disorder difference rates between employed and unemployed groups are compared, it does not appear to be serious at face value; however, the percentage increases between employed and unemployed groups are alarming. More specifically, the highest increase rate appears in the unemployed male group, 103%, during the research period, and the female group jumped 79% during the same period. Contrary, the employed male group shows the lowest growth rate, which is only 2%, but the rate is higher for the female group, which is 20%. Employed individuals regardless of gender show much lower increase rates than unemployed people. It could be that because the top information seeking practices for the unemployed lie in job-related information rather than their health, and the knowledge about opioids may not be efficiently disseminated to them. In addition, unemployed individuals often lack search skills (Kim, 2013), which may hinder the diffusion of the new knowledge related to opioids. Furthermore, individuals without a job may feel depressed and look for opioids to improve their mood more than their counterparts (Conway et al., 2006). In addition, the rate of psychiatric disorders is much higher for the female group than the male category.

5.3. Psychiatric Disorder Changes across Marital Statuses

Viewed from the percentage, a major variation is only captured in gender difference: regardless of marital status, women have higher rates of psychiatric disorders from opioids, and this finding is consistent with existing studies. The new discovery from this dataset is captured in the percentage increase. Table 2 illustrates that the separated, divorced, and widowed (SDW) males and females have higher rates of psychiatric disorders from opioids than married people. In other words, psychiatric disorders from opioids for the male SDW group increased by 76%-- triple the value of the married male group, which increased by 28%. The disparity gap for psychiatric disorders between SDW and married female groups is not as high as the male groups, but the SDW female group still has a higher rate of psychiatric disorders from opioid disorders (32% for married vs. 55% for SDW). This could be because single individuals may feel lonely and experience financial instability, while married individuals may feel responsibilities for the family and receive support from their spouses for the consequences of opioids, which will deter opioid misuse.

5.4. Psychiatric Disorder Changes across Insurance Types

This group of findings unambiguously reveals the importance of insurance with psychiatric disorders among opioid users. As noted, Medicaid covers a disproportionate number of nonelderly adults, and the recipients are individuals of low socioeconomic status with lower education and income. These individuals have less opportunities to learn new medical information due to their educational level. Among Medicaid recipients, psychiatric disorders for the male groups increased by 261%, and that of the female group increased by 157%. Contrary to these jumps, the recipients of Medicare for the male group *decreased* by 40%, which is a stark contract with a 261% increase for the male Medicare group. In the case of the female group, psychiatric disorders from opioids are *down* by 25%, and this is also a clear distinction with the female Medicaid group. While the female Medicaid group has a higher percentage of psychiatric disorders, the increase for the male group is higher by 104%. Additionally, the decrease rate of psychiatric disorders is higher for men than women. This is further evidence that male opioid users' psychiatric disorders will soon surpass those of women.

6. Conclusion

The purpose of this research was to examine how recent knowledge about psychiatric disorders from opioids between men and women and across various socioeconomic groups has changed between 2007 and 2017.

The unexpected and surprising findings are from the gender differences. It is conventional knowledge that women have higher rates of psychiatric disorders, and it is supported in this finding: women's psychiatric disorders from opioids are higher for all the subcategories. This might have been attributed to socially constructed knowledge about women. More specifically, doctors tend to prescribe opioids for women more often than men because they believe that women have higher levels of pain as they tend to express their pain more, which results in higher rates of opioid prescription and psychiatric disorders. However, what is unexpected is that the rates of increase for men are much steeper than women. This finding indicates that in the near future, opioid-related psychiatric disorders in men will surpass women, which will reverse our common knowledge about pain and psychiatric disorders. When the analysis simply compares the percentages across subgroups, this increase will *not* be captured, but this contrasting discovery is revealed when the growth rates are employed.

Somewhat expected and congruent with existing literature, the findings from this dataset support that those psychiatric disorders from opioids are much higher for those of low socioeconomic status: low education, unemployed, SDW, and Medicaid recipients. The author surmises that individuals in low socioeconomic status may have opportunities to learn about the new knowledge compared to their counterparts. This finding suggests that opioid-born psychiatric disorder programs should be focused on low socioeconomic groups.

References

Alabas, O.A., Tashani, O.A., Tabasam, G., & Johnson, M.I. (2012). Gender role affects experimental pain responses: A systematic review with meta-analysis. *European Journal of Pain*, *16*(9), 1211–1223.

Altekruse, S.F., Cosgrove, C.M., Altekruse, W.C., Jenkins, R.A., & Blanco, C. (2020). Socioeconomic risk factors for fatal opioid overdoses in the United States: Findings from the Mortality Disparities in American Communities Study (MDAC). *PLoS ONE, 15*(1). e0227966. https://doi.org/10.1371/journal.pone.0227966

Atkinson, J., Lee, S., Dayton-Shotts, C. & French, C. (2001). Self-Perceived Job Skills and Employment Barriers among Non-Drug Using and Chronic Drug Using Welfare-to- Work Participants. *Journal of Drug Issues*, *31*(3), 747–756.

Back, S., Payne, R., Wahlquist, A., Rickey, E., Carter., Stroud, Z., Haynes, L., Hillhouse, M., Brady, K.T., & Ling, W. (2011). Comparative Profiles of Men and Women with Opioid Dependence: Results from a National Multisite Effectiveness Trial. *The American Journal of Drug and Alcohol Abuse, 37*, 313–323. https://dx.doi.org/10.3109%2F00952990.2011.596982

Bartley, E.J., & Fillingim, R.B. (2013). Sex differences in pain: a brief review of clinical and experimental findings. *British Journal of Anaesthesia 111*(1), 52–58.

Bernardes, S.F., Keogh, E., & Lima, M.L. (2008) Bridging the gap between pain and gender research: a selective literature review. *European Journal of Pain, 12*(4), 427– 440.

Center for Behavioral Health Statistics and Quality Substance Abuse and Mental Health Services Administration. (2007, 2017). Treatment episode data set admissions (TEDS). https://datafiles.samhsa.gov/study-dataset/treatment-episode-data-set-admissions-2007-teds-2007-ds0001-nid13680; https://datafiles.samhsa.gov/study- dataset/national-survey-drug-use-and-health-2017-nsduh-2017-ds0001-nid17939

Center for Disease Control and Prevention. (2019). CDC's response to the opioid overdose epidemic: A public health crisis. https://www.cdc.gov/opioids/strategy.html. Accessed May 21, 2020.

Cicero, T.J., Wong, G., Tian, Y., Lynskey, M., Todotov, A., & Isenberg, K. (2009). Co- morbidity and utilization of medical services by pain patients receiving opioid medications: data from an insurance claims database. *Pain, 144*, 20–27.

Compton, W.M., Cottler, L.B., Abdallah, A.B., Phelps, D.L., Spitznagel, E.L., & Hortone, J.C. (2000). Substance dependence and other psychiatric disorders among drug dependent subjects: Race and gender correlates. *The American Journal on Addictions, 9*(2), 113–125. https://doi.org/10.1080/10550490050173181

Conway, K.P., Compton, W., Stinson, F.S., & Grant, B.F. (2006). Lifetime Comorbidity of DSM-IV Mood and Anxiety Disorders and Specific Drug Use Disorders: Results From the National Epidemiologic Survey on Alcohol and Related Conditions. *The Journal of Clinical Psychiatry, 67*, 247–257. https://doi.org/10.4088/jcp.v67n0211.

Davis, M., Lin, L., Liu, H., Sites, B. (2017). Prescription Opioid Use among Adults with Mental Health Disorders in the United States. *Journal of the American Board of Family Medicine,* 407-417. doi: 10.3122/jabfm.2017.04.170112

Department of Health and Human Services: Substance Abuse and Mental Health Services Administration. (2017, 2007). Treatment Episode Data Set (TEDS), 2017, 2007: Admissions to and discharges from publicly-funded substance use treatment.

Edwards, A.C., Larsson Lönn, S., Sundquist, J., Kendler, K.S., & Sundquist, K. (2017). Associations Between Divorce and Onset of Drug Abuse in a Swedish National Sample. *American Journal of Epidemiology*, 187(5), 1010–1018. https://doi.org/10.1093/aje/kwx321.

Ellis, M., Kasper, Z., Cicero, T. (2020). The impact of opioid use disorder on the levels of educational attainment: Perceived benefits and consequences. *Drug and Alcohol Dependence*, *206*, 1-6. doi.org/10.1037/a0018053

Focus for Health. (2020). The opioid crisis explained in Black and White. https://www.focusforhealth.org/the-opioid-crisis-explained-in-black-and-white/. May 14, 2020.

Galea, S., Ahern, J., Tardiff, K., Leon, A., Coffin, P.O., Derr, K., & Vlahov, D. (2003). Racial/Ethnic Disparities in Overdose Mortality Trends in New York City, 1990– 1998. *Journal of Urban Health, 80*(2), 201-211. https://doi.org/10.1093/jurban/jtg023

Ghate, S.R., Haroutiunian, S., Winslow, R., & McAdam-Marx, C. (2010). Cost and comorbidities associated with opioid abuse in managed care and Medicaid patients in the United States: A comparison of two recently published studies. *Journal of Pain andPalliative Care Pharmacotherapy, 24*(3), 251–258. https://doi.org/10.3109/15360288.2010.501851

Ghertner, R, & Groves, L. (2018). The opioid crisis and economic opportunity: geographic and economic trends. ASPE Research Brief 2018 Sept. https://aspe.hhs.gov/system/files/pdf/259261/ASPEEconomicOpportunityOpioidCris is.pdf

Glaesmer, H, & Deter, H.C. (2002). Gender-specific aspects of physicians' work. *Psychotherapie, Psychosomatik, Medizinische Psychologie, 52*, 386.

Green, C.R., Wheeler, J.R., & LaPorte, F. (2003) Clinical decision making in pain management: contributions of physician and patient characteristics to variations in practice. *The Journal of Pain, 4*, 29.

Han, L., Allore, H., Goulet, J., Bathulapali, H., Skanderson, M., Brandt, C., Haskell, S., & Krebs, E. (2017) Opioid dosing trends over eight years among US veterans with musculoskeletal disorders after returning from service in support of recent conflicts. *Annals of Epidemiology, 27*(9), 563-569. https://doi.org/10.1016/j.annepidem.2017.08.015

Hansen, H., & Netherland, J. (2016). Is the Prescription Opioid Epidemic a White Problem? *Am Journal of Public Health*, *106*(12), 2127-2128. https://dx.doi.org/10.2105%2FAJPH.2016.303483

Hollingsworth, A., Ruhmb, C.J., & Simon, K. (2017). Macroeconomic conditions and opioid abuse. *Journal of Health Economics, 56*, 222–233. https://doi.org/10.3386/w23192

Hurley, R.W., & Adams, M.C. (2008). Sex, gender, and pain: an overview of a complex field. *Anesthesia and Analgesia, 107*(1), 309–317.

Irwin, S.O (2010). A conceptual framework for action on the social determinants of health. Social determinants of health discussion paper 2 (policy and practice). *World Health Organization*, Geneva, Switzerland; 2010.

James, K., & Jordan, A. (2018). The opioid crisis in Black communities. *Journal of Law, Medicine, and Ethics, 46*(2), 404-421. https://doi.org/10.1177%2F1073110518782949

Jones, C.M., & McCance-Katz, E.F. (2019). Co-occurring substance use and mental disorders among adults with opioid use disorder. *Drug and Alcohol Dependence*, *197*, 78-82. doi:10.1016/j.drugalcdep.2018.12.030

Kamimura, A., Panahi, S., Rathi, N., Weaver, S., Pye, M., Sin, K., & Ashby, J. (2020) Risks of opioid abuse among uninsured primary care patients utilizing a free clinic. *Journal of Ethnicity in Substance Abuse*, *19*(1), 58-69. https://doi.org/10.1080/15332640.2018.1456387

Kelly, S.M., O'Grady, K.E., Schwartz, R.P., Peterson, J.A., Wilson, M.E., & Brown, B.S. (2010). The relationship of social support to treatment entry and engagement: the Community Assessment Inventory. *Substance Abuse*, *31*(1), 43–52. https://doi.org/10.1080/08897070903442640

Kidorf, M., Stitzer, ML., & Brooner, RK (1994). Characteristics of methadone patients responding to take-home incentives. *Behavior Therapy*, *12*(1): 109-121

Kim, Yong-Mi (2013). Does online searching cause or enforce health information disparity? Journal of Information & Knowledge Management. 12(4).

Kim, Y.M. (2015). Is seeking health information online different from seeking general information online? Journal of Information Science, 41(2) 228–241.

Kim, Y.M. (2019). Discovering major opioid-related research themes over time: A text mining technique. AMIA Joint Summits on Translational Science, 751-760.

Kim, Y.M., & Delen, D. (2017). Critical assessment of health disparities across subpopulation groups through a social determinants of health perspective: The case of type 2 diabetes patients. *Informatics for Health and Social Care, 43*, 172-185. https://doi.org/10.1080/17538157.2017.1364244

Kosten, T.R., & George, T.P. (2002). The neurobiology of opioid dependence: implications for treatment. *Science and Practice Perspectives*, *1*(1), 13–20. https://dx.doi.org/10.1151%2Fspp021113

Krueger, A.B. (2017). Where Have All the Workers Gone? An Inquiry into the Decline of the U.S. Labor Force Participation Rate. Brookings Papers: On economic activity. https://www.brookings.edu/wp-content/uploads/2018/02/kruegertextfa17bpea.pdf.

Kuo, Y. F., Raji, M. A., Liaw, V., Baillargeon, J., & Goodwin, J. S. (2018). Opioid Prescriptions in Older Medicare Beneficiaries After the 2014 Federal Rescheduling of Hydrocodone Products. *Journal of the American Geriatrics Society*, *66*(5), 945–953. https://doi.org/10.1111/jgs.15332

Lack, S., Noddings, R., & Hewlett, S. (2011) Men’s experiences of rheumatoid arthritis: an inductive thematic analysis. *Musculoskeletal Care, 9*(2): 102–112.

Lanier, W.A., Johnson, E.M., Rolfs, R.T., Friedrichs, M.D., & Grey, T.C. (2012). Risk factors for prescription opioid-related death, Utah, 2008–2009. *Pain Medicine, 13*, 1580–1589. https://doi.org/10.1111/j.1526-4637.2012.01518.x

Lippold, K.M., Jones, C.M., Olsen, E.O., & Giroir, B.P. (2019). Racial/Ethnic and Age Group Differences in Opioid and Synthetic Opioid-Involved Overdose Deaths Among Adults Aged ≥18 Years in Metropolitan Areas - United States, 2015-2017. *MMWR.*

Morbidity and mortality weekly report, *68*(43), 967–973. https://doi.org/10.15585/mmwr.mm6843a3

Magura S. (2003). The role of work in substance dependency treatment: a preliminary overview. *Substance Use and Misuse*, *38*, 1865–2003.

Martins, S.S., Sampson, L., Cerda, M., & Galea, S. (2015). Worldwide Prevalence and Trends in Unintentional Drug Overdose: A Systematic Review of the Literature. *American Journal of Public Health*, *105*, 2373–2393. https://doi.org/10.2105/AJPH.2015.302843a

Mazure, C.M., & Fiellin, D.A. (2018). Women and opioids: something different is happening here. *Lancet 392*(10141), 9-11. doi:10.1016/S0140-6736(18)31203-0

McAdam-Marx, C., Roland, C.L., Cleveland, J., & Oderda, G. M. (2010). Costs of opioid abuse and misuse determined from a Medicaid Database. *Journal of Pain and Palliative Care Pharmacotherapy,* *24*(1), 5–18. https://doi.org/10.3109/15360280903544877

Merrill, J.O., Von Korff, M., Banta-Green, C.J., Sullivan, M.D., Saunders, K.W., Campbell, C.I., et al. (2012). Prescribed opioid difficulties, depression and opioid dose among chronic opioid therapy patients. *General Hospital Psychiatry*, *34*, 581–7. https://dx.doi.org/10.1016%2Fj.genhosppsych.2012.06.018

Myers, C.D., Riley, J.L., & Robinson, M.E. (2003). Psychosocial contributions to sex- correlated differences in pain. *The Clinical Journal of Pain. 19*(4), 225-232.

National Health Care for the Homeless Council. (2017). Addressing the opioid epidemic: How the opioid crisis affects homeless populations. https://nhchc.org/wp-content/uploads/2019/08/nhchc-opioid-fact-sheet-august-2017.pdf

NIDA. (2017). "Addressing the Opioid Crisis Means Confronting Socioeconomic Disparities." *National Institute on Drug Abuse*, https://www.drugabuse.gov/about-nida/noras-blog/2017/10/addressing-opioid-crisis-means-confronting-socioeconomic-disparities. Accessed May 21, 2020

Nielsen, D.S., Brixen, K., & Huniche, L. (2011). Men's experiences of living with osteoporosis: focus group interviews. *American Journal of Men's Health 5*(2): 166– 176.

Orgera, K., & Tolbert, J. (2019). The Opioid Epidemic and Medicaid's Role in Facilitating Access to Treatment. Henry J. Kaiser Family Foundation.

Platt J. J. (1995). Vocational rehabilitation of drug abusers. *Psychological Bulletin, 117*, 416–433,

Platt-Mills, T.F., Hunold, K.M., Bortsov, A.V., Soward, A.C., Peak, D.A., Jones, J.S., Swor, R.A., Lee, D.C., Domeier, R.M., Hendry, P.L., Rathlev, N.K., & McLean, S.A. (2012). More educated emergency department patients are less likely to receive opioids for acute pain. *Pain*, *153*(5), 967-973. https://doi.org/10.1016/j.pain.2012.01.013

van Rijswijk, S.M., van Beek, M.H.C.T., Schoofc, G.M., Schenea, A.H., Steegersd, M., & Schellekensa, A.F. (2019). Iatrogenic opioid use disorder, chronic pain and psychiatric comorbidity: A systematic review. *General Hospital Psychiatry,* 59, 37–50. https://doi.org/10.1016/j.genhosppsych.2019.04.008

Ross, J., Teesson, M., Darke, S., Lynskey, M., Ali, R., Ritter, A., & Cooke, R. (2005). The characteristics of heroin users entering treatment: Findings from the Australian Treatment Outcome Study (ATOS). *Drug and Alcohol Review*, *24*(5), 411–418. https://doi.org/10.1080/09595230500286039

SAMHSA (2017a). *National Survey on Drug and Health: The CBHSQ Report.* https://www.samhsa.gov/data/sites/default/files/report_2734/Spotlight-2734.pdf. Accessed May 21, 2020.

SAMHSA (2017b). Treatment episode data set admissions (teds-a) 2017: *Codebook. Center for Behavioral Health Statistics and Quality Substance Abuse and Mental Health Services Administration.* https://www.datafiles.samhsa.gov/study-dataset/national-survey-substance-abuse-treatment-services-2017-n-ssats-2017- ds0001-nid18445.

SAMHSA (2019c). Homeless and housing services providers confront opioid overdose. Locatedhttps://www.samhsa.gov/homelessness-programs-resources/hpr- resources/homeless-housing-services-providers-confront-opioid. Accessed May 13, 2020.

Samulowitz, A., Gremyr, I., Eriksson, E., & Hensing, G. (2018). "Brave men" and "emotional women": a theory-guided literature review on gender bias in health care and gendered norms towards patients with chronic pain. *Pain Research and Management, 2018*(6358624). doi:10.1155/2018/6358624

Scott, K.M., Wells, J.E., Angermeyer, M., & Brugha, T., Bromet, E., Demyttenaere, K., de Girolamo, G., Gureje, O., Haro, J.M., Karam Nasser, A., Kovess, V., Lara, C., Levinson, D., Ormel, J., Posada-Villa, J., Sampson, N., Takeshima, T., Zhang, M., & R.C. Kessler. (2010). Gender and the relationship between marital status and first onset of mood, anxiety and substance use disorders. *Psychological Medicine, 40*(9), 1495– 1505. https://doi.org/10.1017/S0033291709991942

Serdarevic, M., Striley, C.W., & Cottler, L.B. (2017). Gender differences in prescription opioid use. *Current Opinion in Psychiatry, 30*(4), 238–246. https://doi.org/10.1097/YCO.0000000000000337

Sheth, S.A., Chandra, N.C., & Mehta, R.Y. (2018). Cognition and Opioid Substitution Therapy Adherence in Patients with Opioid Dependence. *Journal of Clinical and Diagnostic Research*, *12*(5), VC01-VC05.

Shiels, M.S., Freedman, N.D., Thomas, D., & Berrington de Gonzalez, A. (2018). Trends in U.S. Drug Overdose Deaths in non-Hispanic blacks, Hispanics, and non-Hispanic whites: 2000–2015. *Annals of Internal Medicine, 168*(6), 453–455. https://doi.org/10.7326/M17-1812

Simoni-Wastila, L. (2000). The use of abusable prescription drugs: the role of gender. *Journal of Women's Health and Gender-Based Medicine, 9*(3), 289-297. doi:10.1089/152460900318470.

Stock, S.A.K., Stollenwerk, B., Redaelli, M., Civello, D., & Lauterbach, K.W. (2008). Sex differences in treatment patterns of six chronic diseases: an analysis from the German statutory health insurance. *Journal of Women's Health, 17*(3), 343–54.

Sullivan, M.D., Edlund, M.J., Steffick D., & Unützer, J. (2005). Regular use of prescribed opioids: association with common psychiatric disorders. *Pain*, *119*, 95–103. https://doi.org/10.1016/j.pain.2005.09.020

Svikis, D. S., Keyser-Marcus, L., Stitzer, M., Rieckmann, T., Safford, L., Loeb, P., Allen, T., Luna-Anderson, C., Back, S. E., Cohen, J., DeBernardi, M. A., Dillard, B., Forcehimes, A., Jaffee, W., Killeen, T., Kolodner, K., Levy, M., Pallas, D., Perl, H. I., Potter, J. S., … Zweben, J. (2012). Randomized multi-site trial of the Job Seekers' Workshop in patients with substance use disorders. *Drug and Alcohol Dependence*, 120(1-3), 55–64. https://doi.org/10.1016/j.drugalcdep.2011.06.024

Tabenkin, H., Goodwin, M., Zyzanski, S., Stange, K., & Medalie, J. (2004). Gender Differences in Time Spent during Direct Observation of Doctor-Patient Encounters. *Journal of Women's Health, 13*, 341-9. 10.1089/154099904323016509.

Teesson, M., Havard, A., Fairbairn, S., Ross, J., Lynskey, M., & Darke, S. (2005). Depression among entrants to treatment for heroin dependence in the Australian Treatment Outcome Study (ATOS): Prevalence, correlates and treatment seeking. *Drug and Alcohol Dependence*, *78*(3), 309–315.

Unruh, A.M. (1996). Gender variations in clinical pain experience. *Pain. 65*(2–3), 123– 167.

Wasan, A.D., Michna, E., Edwards, R.R., Katz, J.N., Nedeljkovic, S.S., Dolman, A.J., Janfaza, J., Isaac, Z., & Jamison, R.N. (2015) Psychiatric comorbidity is associated prospectively with diminished opioid analgesia and increased opioid misuse in patients with chronic low back pain. *Anesthesiology*, *123*(861), 72. https://doi.org/10.1097/ALN.0000000000000768

Whitmire, J.T., & Adams, G. (2010). *Unintentional overdose deaths in the North Carolina Medicaid population: prevalence, prescription drug use, and medical care services*. Raleigh, NC: State Center for Health Statistics.

World Health Organization (2020). Mental health: Gender and women's mental health. Located at https://www.who.int/mental_health/prevention/genderwomen/en/. Accessed on June 9, 2020.

Wu, L.T., Woody, G.E., Yang, C., & Blazera, D.G. (2010). Subtypes of nonmedical opioid users: Results from the national epidemiologic survey on alcohol and related conditions. *Drug and Alcohol Dependence, 112*, 69–80. https://dx.doi.org/10.1016%2Fj.drugalcdep.2010.05.013

Zanis, D.A., Coviello, D., Alterman, A.I., & Appling, S.E. (2001). A community-based trial of vocational problem-solving to increase employment among methadone patients. *Journal of Substance Abuse Treatment, 21*(1), 19-26. doi:10.1016/s0740-5472(01)00177-5

U.S. Buraus of Labor Statitisc (2021). People who are not in the labor force: Why aren't they working? https://www.bls.gov/opub/btn/volume-4/people-who-are-not-in-the- labor-force-why-arent-they-working.htm. Accessed on October 12, 2021

Increasing Information Certainty for Post-traumatic Growth

[Practitioner and Work-In-Progress Abstract]

Nicole Bank
University of North Texas, Research Associate,
nicolebank@my.unt.edu
https://orcid.org/0000-0002-9339-5595

Jeff Allen
University of North Texas, Regents Professor,
Jeff.Allen@unt.edu
https://orcid.org/0000-0003-0551-0539

Summary: Trauma, and its associated effects, can be conceptualized as a period of information uncertainty. The natural psychological response to trauma is a period of post-traumatic stress. Trauma occurs when an existing knowledge base has been challenged. Any event that challenges important components of an individual's assumptive world is said to be traumatic. This post-traumatic period is akin to many theories and concepts in information science including uncertainty reduction, Everyday Life Information Seeking, Sensemaking Theory, Making Meaning and Anomalous States of Knowledge. One possible outcome after the post- traumatic period is post-traumatic growth. Researchers agree post-traumatic growth primarily occurs across one or more of the following domains: personal strength, new possibilities, relating to others, appreciation of life and spiritual change. That is, people affected by trauma tend to grow when they find new or additional paths of information certainty.

Method: To construct a preliminary theoretical framework for post-traumatic growth from an integrative literature review of knowledge management, social psychology and organizational development that can be used by organizations.

Social Implication: Negative experiences can spur positive growth. Knowledge Management researchers are engaged in studying knowledge enhancement processes thereby providing assistance when information is uncertain. The authors propose developing a theoretical framework and practical knowledge management tool that creates certainty for the individual across one or more of these post-traumatic growth domains, thereby increasing the information certainty. This knowledge management tool can foster post-traumatic growth after adverse experiences such as the death of a loved one, assault or war. During the last year, the world has experienced a global pandemic spurring economic impact, millions of deaths, unemployment and a global economy. A KM tool for fostering growth after trauma could help save lives.

Keywords: Sensemaking, ASK, Trauma, Growth

Learning to Support Computable Biomedical Knowledge (CBK)

Deborah Swain, Ph.D., NC Central University, USA and

Christopher Cunningham, Ph.D., NC Central University, USA

dswain@nccu.edu and chris.cunningham@nccu.edu

ABSTRACT

Using technology and e-publication formats, Mobilizing Computable Biomedical Knowledge (MCBK) aims to promote learning health services and reduce time to get healthcare products to patients. This paper summarizes developing a community of practice (CoP) experts, providing educational activities, and plans for sustainable Open Educational Resource (OER) materials online as funded by an IMLS grant.* Activities were part of a pilot training for 20 knowledge managers, students and professionals in Library and Information Science (LIS) in December, 2021. The educational goal is to introduce MCBK concepts globally and and to promote KM and LIS leadership in designing healthcare repositiories and accessing information.

Beyond text and graphics, CBK publications provide encodable prediction models and computable information in electronic media. It is "dynamic knowledge" in open access formats. Such computable knowledge allows readers to evaluate and validate data or tools that may, for example, compute a risk score for infection. Analysis and review by e-journal readers may reduce time to implementation in healthcare systems or K2P (knowledge to performance) from years with traditional peer- reviewed publications to months using electronic, open access.

The summaries and exercises in this paper are from presentations during the pilot training, which were assessed and recommended by student participants.

* Institute of Museum and Library Services, RE-250159-OLS-21.

AUTHOR KEYWORDS

biomedical knowledge; computable knowledge; learning health systems (LHS); open access; open educational resources (OER)

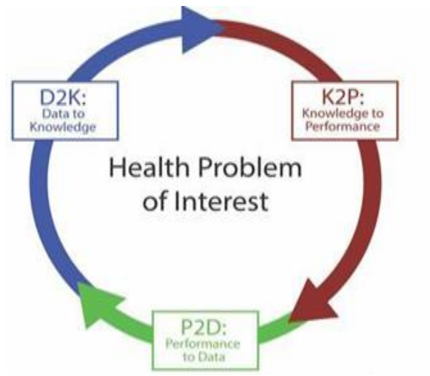
INTRODUCTION

The need for Learning Health Systems (LHS) is worldwide. Knowledge of computable systems is growing to fill the gaps in LHS and to reduce health disparities (Friedman & Flynn, 2019). Both cost reductions and faster development of effective, efficient systems analysis programs (SAP) seem possible from Mobilizing Computable Biomedical Knowledge (MCBK). Open access journals can make a difference. An example of an open access publication, the *Learning Health Systems Journal*, is edited and managed at the University of Michigan, School of Medicine, USA. As this paper will show, leadership in developing CBK by contributing researchers, authors, and editors can help knowledge managers, librarians, and information professionals be early designers and leaders in academics, government, and business.

A systems approach is required for health care to improve across states, countries, and continents. With systems, the data, algorithms, and informatics programs can be generalized during a learning cycle that covers performance to data (P2D), data to knowledge (D2K), and knowledge to performance (K2P) for solving a health problem. Problems like heart disease and diabetes require system.

The following figure from Charles Friedman (2021) illustrates the cycle going clockwise from the bottom (P2D) when medical practices can be captured as data, assembled and analyzed with external evidence to become knowledge (D2K), and then the results are interpreted, and interventions can be designed for health care action (K2P) as the cycle starts again.

Figure 1
LHS: Continuous Cycles of Study and Change



As the 2018 "Mobilizing Computable Biomedical Knowledge (CBK): A Manifesto" illustrates, the effort will take time and dedicated support.

It is no longer sufficient to represent knowledge in the form of printed words and static pictures. The increasingly rapid rate of scientific discovery needs knowledge representations that are more agile and amenable to scalability and mass action. This in turn can enable the continuous cycles of discovery and improvement envisioned as Learning Health Systems. (Manifesto, 2018)

It will take public-private partnerships to sustain a CBK ecosystem. Dedicated to FAIR (findable, accesible, interoperable and resuable) principles (Wilkinson, et al., 2016), the growing community of MCBK and LHS supporters suggest that it is time for action on a global scale. This paper outlines opportunities for knowledge managers, information professionals, and librarians to participate in MCBK and to locate resources tolearn about aspects of CBK. In 2020, Friedman (2020) was the first keynote speaker at the ICKM 2020 conference and he provided details of the history and concepts defining MCBK.

COMMUNITY OF PRACTICE (CoP): Partners, Presenters, and Planners

As described in the 2021 grant proposal to the IMLS, a training pilot and follow-up research might capitalize on the wealth of knowledge available within the LIS academic world. So with collaborative partners, pilot presenters, and associated planners and mentors, our CBK team is building an online training curriculum with health informatics experts from around the world. The results of research from the grant will support the advancement of theory and practice in biomedical informatics while developing leaders in computable knowledge storage, access standards and tools across LIS disciplines. The team also hopes to address current information needs of the medical and healthcare community as well as future research development and data needs. Equipping LIS professionals with the skills to use tools in software and data management supports the LHS objective for every decision affecting the health of individuals and populations to be informed by the best knowledge available.

Collecting that knowledge requires a collaborative team of partners to plan and support the pilot class and subsequent research. The authors, who are the project director and co-project director, invited MCBK and health informatics experts to be collaborative partners. The original list of partners who continue to support the CBK research and knowledge-based learning systems development are:

- Dr. Charles Friedman, LHS department chair, Michigan School of Medicine
- Dr. Rachel Richesson, Michigan School of Medicine;
- Kathleen Young, Editorial Assistant, *LHS* journal;
- Jerry Perry, Associate Dean for Unversity of Arizona Libraries, and inclusion researcher;
- Feili Tu-Keefner, health informatics and health sciences librarianship expert at the University of South Carolina
- Suliman Hawamdeh, Ana Cleveland and Jody Philbrick, professors and lecturer at North Texas University
- Robin Ann Yurk, MD Yurk with healthcare and leadership experience
- Shannon Jones, Director of Libraries for the Medical University of South Carolina.

Associate members of our growing CoP are health informatics, LIS, and data analytics consortium partners in North Carolina at Duke University, UNC-Chapel Hill, East Carolina University, UNC-Charlotte and Wake Forest University, including Javed Mostafa, Director of the Carolina Health Informatics Program (CHIP) at UNC-Chapel Hill and Drs. Ed Hammond and Vivian West, Duke Clinical Health Informatics (DCHI) at Duke University in Durham, NC. As described next, the CoP worked collaboratively to design a syllabus for the pilot training in MCBK held December 15, 2021 - January 6, 2022. Additional community partners presented modules and mentored the pilot class attendees, 20 professionals and graduate students from across the USA.

Beginning in October, 2021, an invitation to participate was distributed by email and social media to the JESSE List, ASIS&T SIGs for HLTH and KM, the ICKM, and to faculty and students of LIS in NC, Washington, and other states. The MCBK pilot training was designed to be taught online using Zoom and to require 40 hours of interactive learning with mentors for 3 cohorts. Attendees (and mentors) would qualify for stipends. Space was originially limited to 15 students, but a supplementary grant from the NNLM and southern chapter added 5 students for a total of 20 training pilot participants. After the original invitations were distributed a USA federal grant restriction was noted, and it was announced that international students and librarians or IS professionals needed to be US students or workers in the USA with Visas. About 10 international applicants applied and were informed that they would have access to Open Education Resources (OER) being developed online in 2022 based on the pilot training and additional research.
Here is the text copy of the short invitation:
Library and Information professionals are invited to apply for training on Mobilizing Computable Biomedical Knowledge (MCBK). Training will be online using Zoom over two weeks in late December 2021 and early January 2022 (between traditional fall and spring semesters). As part of a grant from the Institute of Museum and Library Services (IMLS), the training will introduce how electronic journals can provide “dynamic knowledge” that readers can validate immediately. For example, for data, programming code or encodable prediction models in healthcare, readers can provide feedback and contribute to faster development of diagnosis treatments or tools to compute risk of coronavirus infection.
A stipend of up to $1500 will be paid to students for attending class and providing feedback. The 40-hour interactive class with speakers and discussions are 2-4 hours: December 15, 16, 20, 22, 23, 27, 28, 30 and January 3, 5, and 6. Mentors will support groups of students, too. Please submit your letter of interest and resume to Dr. Deborah Swain, Project Director (dswain@nccu.edu) or Dr. Christopher Cunningham, Co-Director (ccunni23@nccu.edu) by November 15.
Professional librarians, technical professionals and students are invited. Interest in health informatics recommended but not required.

Twenty participants were invited and accepted; the group included five particpants sponsored by the Medical Library Association (MLA) Southern Region Chapter of the National Network of Libraries of Medicine (NNLM) who lived in the southern region. Three applicants on the wait-list were sent copies of the schedule and invited to audit if interested without a stipend. Pilot participants ranged from a young pre-med undergraduate to a Ph.D. with over 10 experience in health sciences libraries. Attendance, participation, and feedback on speakers, learning modules and issues was 100%. For examples, student participants reviewed sample materials, interacted with knowledgeable speakers, and provided feedback and suggestions about the open access publication, *Learning Health Systems Journal*, bias in CBK, doing systematic reviews, and interoperable, metadata infrastructure proposals. A pre- and post-assessment was available and completed by all participants. (Results of knowledge learning is summarized under “Learning Assessments“ below.)

THE PILOT: Syllabus, Schedule, and Modules for Training

From September to December, partners met, reviewed and revised the Syllabus. They agreed to these learning objectives and outcomes in Table 1.

Table 1
Pilot Training Learning Outcomes

Goals and Objectives	**Student Learning Outcomes (Students will be able to…)**	**Measurement**
Find tools and utilize searches for CBK: LIS professionals can help design more effective data archives and repositories to improve healthcare information accessibility for professionals, patients, and researchers.	1. Design MCBK archive repository for healthcare papers or articles with code, data, algorithms.	1. Present guidelines for an MCBK archive repository
Review the technology and electronic publication format for CBK: Competency with electronic journals to provide "dynamic knowledge" that readers can validate immediately. For example, for data, programming code or encodable prediction models in healthcare, readers can provide feedback and contribute to faster development of diagnosis treatments or tools to compute risk of coronavirus infection	2. Identify accessibility requirements for readers to evaluate code, data, algorithms.	3. Document directions for authors and users to access data to evaluate healthcare device or procedures in electronic articles
Provide feedback and suggestions for developing OER: Based on the pilot class and a community of practice (CoP) from universities, libraries, and medical schools, sustainable open educational resources (OER) will be developed with online materials for future use. (2022)	4. Plan for OER resources based on pilot classes and MCBK: Cheryl Casey's toolkit and guidelines (Arizona), Will Cross (NC State), and Josh Bullock (Kansas) 5. Create CoP dist list from collaborative partners and pilot class	6. Students provide feedback for dynamic, interactive learning resources OER (Open Educational Resources); pilot class reviews OER toolkit and recommends legal approach 7. Test dist list with email; update webpage

Students were expected to attend 2 orientations and 9 classes in Zoom meetings. The two 2-hour orientations were 10:30am-12:30pm (December 15 and January 6), and the nine 4-hour online classes were 12- 4pm EST for a total of 40 hours of training. An outline of the modules is in Table 2.

Table 2
Pilot Training: Schedule by Subject

Class	Student Learning Outcome Subject	Activities
1- Thurs., Dec. 16	Open Access to Data, Research, and Scholarship	Follow-up to orientation: slides and video Review Sample article Survey from UVa
2- Mon., Dec. 20	Publishing	Discuss Author Instructions and sample article Process for journal submission
3- Wed., Dec. 22	Knowledge Bases and Repositories	Review background articles and poster before Discuss CBK archive repository draft desiderata
4- Thurs., Dec. 23	Bias in Machine Learning, Algorithms, AI	View "Coded Bias" video before Read sample articles (including "Algorithmic Justice" guidelines)
5- Mon., Dec. 27	Entrepreneurship and Application Development	Read PATTIE article Discuss new tools for apps
6- Tues., Dec 28	Systematic Project Review Scoping Review Librarian roles	AHRQ Poster Video on PICO Selected reading (PRISMA, plus)
7- Thurs., Dec 30	CBK Technical Infrastructure	Review and summarize poster with partner or team Discuss principles for technical infrastructure to support CBK
8- Mon., Jan 3	MCBK Metadata and research (data) networks	Discuss metadata; website review exercise Example OHDSI slides
9- Wed., Jan 5	NLM Update	Attend review of NLM updates Discuss Covid-19 impact on CBK

The invited speakers were experts in their field and the domain knowledge that they presented. Table 3 lists the speakers and their topics with some of the information resources used.

Table 3
Pilot Training: Presentations and MCBK Resources

Session	Learning Activities	Resources
Orientation, **Dec. 15** (10:30 – 12:30 pm EST)	**Introduce MCBK** Describe LHS (Learning Health Systems) background and "MCBK Manifesto" Meet mentors and partners Review pilot process, schedule, expectations, and webpage: https://www.nccu.edu/mcbk-home	Speaker: Charles Friedman Assignment: Friedman slides on CBK. Zoom video (12/15/21)

1-Class, Thurs., **Dec. 16** (12-4 pm EST)	Complete pre-class assessment survey **Open Access to Data to Knowledge (to application and practice)** Describe library context for MCBK (open research and scholarship); complete survey from UVa Volunteer for Lib Guide dev	Review cycle of learning from slides and video. *LHS* Article (invited early reading): https://doi.org/10.1002/lrh2.10244 Speaker: Nancy Allee on “open science” and “open data” Interactive discussion with mentor (breakout sessions) Class discusses together (Lib Guide ideas)
2-Class, Mon., **Dec. 20** (12-4 pm EST)	Evaluate article on immunization calculation engine Describe online *LHS* journal’s mission, challenges and growth and Author Instructions Process for manuscript submission, review, production, publication/open access/and indexing.	Review CYCLE OF LEARNING to application and practice topic: Data to Knowledge (D2K), Knowledge to Performance (K2P) and Performance to Data (P2D) and Open Access *LHS* Article (how CKP manuscript differs): https://doi.org/10.1002/lrh2.10285 Speaker: Kathleen Young (background on Learning Health Systems follow-up with Charles Friedman) Interactive discussion with mentor (breakout sessions)
3-Class, Wed., **Dec. 22** (12-4 pm EST)	**Trust for knowledge bases and repositories** MCBK Trust and Policy (T&P) Workgroup	Review assigned readings Speaker: Josh Richardson, T&P MCBK Workgroup Review preliminary survey results by T&P MCBK Workgroup Interactive discussion with mentor (breakout sessions) Discuss CBK archive repository draft desiderata Assignment for Dec. 23: “Coded Bias” documentary video (MIT)
4-Class, Thurs., **Dec. 23,** (12-4 pm EST)	**Bias in Machine Learning (algorithms in AI and ML)** using Ethics of Care framework Review “Thought Piece” and Call to Action Review inclusion guidelines and bias analysis, plus “Algorithmic Justice” Engaging different disciplines	Discuss “Coded Bias” documentary video (MIT) Five (5) selected articles from annotated bibliography or lit review Speaker: Jerry Perry (S&I MCBK Workgroup) will review Sustainability and Inclusion issues and “mobilizing” aspect of MCBK Interactive discussion with mentor (breakout sessions) Full class discussion and follow-up with Jerry.

5-Class, Mon., **Dec. 27,** (12-4 pm EST)	**Entrepreneurship and application development** New Tools for entrepreneurs. Prototype system called Publication Access Through Tiered Interaction & Exploration (PATTIE).	Read article on PATTIE Speaker: Javed Mostafa (UNC-CH) Advancing research in new areas (CBK and DSS) Interactive discussion with mentor (breakout sessions)
6-Class, Tues., **Dec 28**, (12-4 pm EST)	**Systematic Project reviews** Librarians as reviewers and authors Evidence-based research Academy of Medicine stds, PICO, and PRISMA reviews	Video on PICO: https://www.youtube.com/watch?v=IHVO4FC2_Is Selected reading: https://guides.mclibrary.duke.edu/sysreview/types Speaker: Feili Tu-Keefer (U South Carolina) Interactive discussion with FULL CLASS (knowledge café)
7-Class,Thurs., **Dec 30,** (12-4 pm EST)	Review posters from MCBK Conf. 2021 (Gabe Rios) **Guiding Principles for Technical Infrastructure to Support Computable Biomedical Knowledge (White Paper)**	Select with team/partner poster or paper (Jan. 5). Note: AHRQ Poster Speaker: Chris Shaffer (UC-San Francisco) Interactive discussion with mentor (breakout sessions) Report out
8-Class, Mon. **Jan 3,** (12-4 pm EST)	**Overview MCBK metadata and the nature of different research (data) networks** and how MCBK knowledge might connect with the available data. Example: Observational Health Data Sciences & Informatics (OHDSI).	Overview of metadata. Activity on website identification. Speaker: Rachel Richesson Interactive discussion with mentor (breakout sessions) Background slides from Juan Banda (Georgia State) on OHDSI data collaboration
9-Class, Wed.,**Jan 5,** (12-4 pm EST)	**Update on NLM Tools and Resources** Creating learning resources and videos Complete post-class assessment survey	Discuss posters and papers from MCBK conference. Speaker: Kate Majewski, NLM NLM Resources: Research & Development and NLM Tools for Moving to Future with FAIR Principles Interactive discussion with FULL CLASS (knowledge café)
Closing Orientation, **Jan. 6** (10:30 – 12:30 pm EST)	Share plans and LIS designs for OER (Open Educational Research) Human-Computer Interactions and Interface UX	Deborah Swain (Grant project director) and Chris Cunningham (Grant co-project director) Creative designs for MCBK following pandemic impact and Impact of Covid-19 Pandemic on Data Science

Note: Contact paper authors for details on speakers and resources.

Reports from three groups of cohorts provided anecdotal evidence of learning by experienced librarians, developers, students, and information professionals interested in supporting CBK and encouraging researchers in their institutions to participate in sharing knowledge in computable formats for open access review and efficient systems implementation. Most learning was interactive and based on participation in cohorts; however, there were some pre-learning assignments, such as view a video or read an article. However, the learning was participatory and part of the training pilot‘s online meeting hours. Each class meeting had a speaker and/or discussion topic. There were breakout sessions with 2 mentors for each student cohort during most classes. On December 28 and January 5, a full class discussion or “knowledge café” replaced the breakout sessions. As the next section summarized, the assessments measured and showed quantitatively, several presentations were evaluated as knowledge-learning experiences. A focus group meeting and “class reunion” is planned for late June or July in 2022.

LEARNING ASSESSMENTS: Student Learning Outcomes and Measurements for Recommended Modules

The primary goal of the upcoming OER will be to support training and implementation of CBK developments for LHS through e-publications and collections or repositories with computable knowledge. Topics from the pilot that show promise for LIS learners include: developer entrepreneurship, trust and policy requirements to counter bias in AI or ML and support algorithmi justice, using NLM resources, supporting a technical infrastructure and metadata for interoperable CBK repositories, systematic reviewing by librarians/knowledge managers, and learning how to publish in open access media like the *LHS* journal (Friedman, et al., 2017). In this section, we outline the modules that will be based on the Pilot Training and reflect the feedback from participants about sessions through assessment surveys before and after the pilot (see Appendix A for questions).

The sessions with high percentages of learning during the Pilot Training seemed to show student participants what new terms meant. The following terms (names and abbreviations had some of the highest % change in knowledge increases between the pre-test and the post-test; see Tables 2 and 3, and Appendix B: Spreadsheet on Pre- and Post-Assessments (Learning %):

- PATTIE (class 5) – 130.56% change in perceived knowledge
- AHRQ (classes 6-7; resources list in syllabus) – 118.92%
- FAIR principle (classes 4 & 9; resources list in syllabus) – 108.7%
- OHDSI (class 8; slide presentation) – 100%
- MCBK (opening orientation and background articles for class 3) – 94%.

It should be noted that OER did not have a class session but was mentioned throughout the course, and it had a low 80% change. As a result, plans for the OER textbook and website include module about open education resources and example. Other recommendations for the high percentage terms reflect a business knowledge interest in entrepreneurship tools, such as the prototype system called Publication Access Through Tiered Interaction and Exploration (PATTIE) from UNC-Chapel Hill, which is a specialized new tool for building programs and using CBK. The desire to grow from no information to knowledge specifically will require a presentation about AHRQ (Agency for Healthcare Research and Quality); Pilot Training depended on the resource list of articles. Because FAIR (findable, accessible, interoperable, and reusable) is a basic MCBK principle, several OER learning modules will continue to emphasize it when describing how to build CBK and to avoid bias. The MCBK Manifesto (2018) is available at https://mobilizecbk.med.umich.edu/about/manifesto and states:
Achieving this through evolution of an open Computable Biomedical Knowledge ecosystem dedicated to achieving the FAIR principles: making Computable Biomedical Knowledge easily findable, universally accessible, highly interoperable, and readily reusable.* The current interest in making data “FAIR” should be matched by equally intense interest in making knowledge “FAIR.” - October 17, 2018 (Wilkinson, et al., 2016)

OHDSI is Observational Health Data Sciences and Informatics, an more learning materials are recommended. An important learning module would cover MCBK metadata and the nature of different research (data) networks and how MCBK knowledge might connect with the available data. Already LHS researchers have published frameworks and data analysis for transmission dynamics and metadata for models based on knowledge commons, data sharing agreements, and standards (Foraker, et al., 2020). Definitely, research models will part of the OER. The post-assessment showed 100% improvement in learning about Observational Health Data Sciences and Informatics (OHDSI) and 93% improvement describe the nature of different research (data) networks, which are the heart of data infrastructure and successful sharing in a learning health system. The pilot learning module and research by Rachel Richesson can cover website identification, research networks for sharing, embedded research, and data standards (Richesson, 2020). Also, pilot slide presentations with examples of OHDSI as demonstrated by Juan Banda are recommended. OHDSI is already connecting almost 3000 users in 18 countries with a network of databases as presented to MCBK and LHS groups (Banda, et al., 2017). Guiding principles for a technical infrastructure to support CBK is being developed by MCBK workgroups and repository users; an updated learning module can be in the OER based on the pilot training.

Finally, the term MCBK itself and "mobilizing computable biomedical knowledge" will remain an integral part of all lessons. Building on Charles Friedman's orientation presentation and his participation in class 1 about the open access *LHS Journal*, learning modules will emphasize the Cycle of Learning and the systems required for successful learning health services. The concept of MCBK takes many lessons and is a dynamic movement. The 84.31% (middling) change indicates the need to repeat and continually give examples of Data to Knowledge (D2K), Knowledge to Performance (K2P) and Performance to Data (P2D) stages of learning in healthcare systems and change.

Overall, student learning was shown to be strong from the Pilot Training assessment. After analysis of assessment data recommendations for more or new learning modules have been determined (see recommendations in Appendix C). For example, there was a 100% increase in knowledge about OHDSI and 93% increase in understanding the nature of different research (data) networks from pre-workshop to post-pilot classes. Students reponded "well" when asked to rate learning about both (questions #15 on data networks and #16 on OHDSI). See more about OER plans in the next section.

FUTURE: OER, Standards for Data and Publication Repositories, and Research on Bias in CBK and AI

Developing sustainable Open Educational Resource (OER) materials in 2022 is research-in-progress intended especially for LIS professionals and students. The importance of developing CBK leaders from LIS is underscored by research that has found, "little attention has been paid to forecasting the information resources and services that researchers, specifically, will need" (Cain, et al., 2016). Whereas, LIS trained professionals seem to have begun to lose reference and knowledge management positions to more IT-based workers, MCBK training may provides an opportunity to bridge gaps and provide design partners from LIS to be part of a community of CBK developers in LHS.

It is expected that the CoP, which will grow out of the grant and developed sustainable OER materials, to support healthcare community outreach projects and activities. As the pilot training demonstrated, trained mentors can be partners with novice CBK users (Figueroa, 2014) to have impact on multi-racial, low income urban and rural communities with health disparities.

Among the processes to be learned, there is first the cycle of learning; however, equally important is systematic project reviews in which librarians and knowledge managers can lead reviewers to meet high level reviews so that CBK is peer reviewed and reader reviewed effectively. The PICO approach requires the patient, intervention, comparison and outcomes framework to focus clinical questions and produce quantitative reviews. Learning and practicing systematic reviewing was a module in the Pilot Training, but not part of assessment surveys. Nevertheless, it is an important aspect of leadership in CBK for librarians and information professional.

Knowing about publishing in open access journals was covered in classes 1 and 2. Learning about author guidelines for the *Learning Health Systems (LHS)* journal received a 112.2% learning change score. Thus, there is evidence that librarians and information professional want to be able to teach or show researchers and programmers in healthcare how to publish using CBK. Resulting repositories of data and access to software and algorithms, will require data standards; however, the LHS results can be strong in making research efficiently reach implementation or performance phases.

Finally, learning how to identify "algorithmic justice" or bias in AI, is complex and may require many learning modules in OER. With the fairly extensive learning materials (video, 5 articles, and an annotated bibliography), the co-chair of the Sustainability and Inclusion MCBK work group, Gerald Perry, was able to introduce bias issues well in class 4. Students had active cohort discussions. Because of the assessment scores for bias (question #8) and algorithmic bias (#20) being 83.02% and 78.43% respectively, it is recommended that specific examples be presented for both and learners practice using an Ethics of Care approach to analyzing types of bias in the OER.

Using the assessments and notes from students, slides, and recordings, the Pilot Training gives OER developers rich material to sort and organize into modules. Class, chapter, or presentation names may change in the production of an open textbook and online modules for the OER website, but the evidence from the Pilot Training shows that librarians and information professionals can prepare to partner with developers, clinicians, and data scientists to create CBK in open publications and CoPs.

CONCLUSION: CBK in Libraries and Biomedical Institutions

We feel that our paper is supporting the theme of the ICKM 2022 conference, "Knowledge and Risks: From individual to global scale" at different levels of analysis and agency. Individuals can learn to be leaders in the MCBK movement and support learning health systems on a global scale. But there can be risks with knowledge if you do not work to understand terms, processes (cycle of learning, systematic reviews and publishing in open access journal), standards, and how to identify "algorithmic justice" or bias in AI. Building on the Ethics of Care, FAIR principles, and knowledge commons, you an apply the level of analysis and personal agency you choose. Whether part of a library or biomedical institution, you can encourage and help manage CBK that will contribute to learning health services and practices at a systems level.

As MCBK publications grow, the need for trained LIS professionals increases. As a recent example, research on rapidly translating clinical guidelines for Covid-19 promises systematic use of AI and knowledge engineering to create decision making models (Fox, et al., 2020) for LHS. Can there be a "second knowledge revolution" based on publishing and sharing computable information, not just text and graphics? Based on the MCBK Pilot Training and input from collaborative partners, who presented or mentored during the pilot, we agree with early MCBK promoters (Williams, et al., 2020) and add that knowledge managers, librarians, and information science professionals can be leaders in this revolution to improve healthcare globally. We hope you will utilize the OER textbook and website based on what we learned during the Pilot Training and applied to developing an LHS learning commons. New partners, such as Cheryl (Cullier) Casey, Open Education Librarian at the University of Arizona Libraries, and Will Cross, Director of the Copyright and Digital Scholarship Center in the NC State University Libraries, will join our efforts at NC Central University as our IMLS research team works with Danielle Colbert-Lewis, Head of Research and Instructional Services at James E. Shepard Library, and Racheal Brooks, Office of e-Learning. Join us and use OER in late 2022 via online, accessible textbook and learning resources at a webpage based on our Pilot Training website at https://www.nccu.edu/mcbk-home .

ACKNOWLEDGEMENTS

Graduate assistants Delaina Zeinoun, MIS, and Nijee Brown, MLS student, provided invaluable support in planning the MCBK pilot training and developing a website, and in analyzing the pre- and post-assessments from pilot participants.

REFERENCES

Banda, J., Halpern, Y., Sontag, D., and Shah, N. (2017). "Electronic phenotyping with APHRODITE and the Observational Health Sciences and Informatics (OHDSI) data network," AMIA Joint Summits Translational Science Proceedings, July 26, 2017, pp. 48-57. PMID: 28815104.

Cain, T., Cheek, F., Kupsco, J., Hartel, L., and Getselman, A. (2016). "Health Sciences Libraries Forecasting Information Service Trends for Researchers: Models Applicable to All Academic Libraries," College and Research Libraries, September, 2016. doi:https://doi.org/10.5860/crl.77.5.595

Figueroa, O. (2014). "The significance of recruiting underrepresented minorities in medicine: an examination of the need for effective approaches used in admissions by higher education institutions," Medical Education Online, 19:1, 24891, DOI: 10.3402/meo.v19.24891

Foraker, R., Lai, A., Kannampallil, T., Woeltje, K., Trolard, A., and Payne, P. (2020). "Transmission Dynamics: Data Sharing in the Covid-19 Era," Learning Health Systems, 2021;5: e10235. https://doi.org/10.1002/lrh2.10235

Fox, J., Khan, O., Curtis, H., Wright, A., Pal, C., Cockburn, N., Cooper, J., Chandan, J., and Nirantharakumar, K. (2020). "Rapid Translation of Clinical Guidelines into Executable Knowledge: A Case Study of Covid-19 and Online Demonstration," Learning Health Systems, 2021;5:e10236. https://doi.org/10.1002/lrh2. 10236

Friedman, C. (2020) "Mobilizing Computable Biomedical Knowledge: An Imperative for Learning Health Systems," 2020 International Conference for Knowledge Management, Online Presentation, Durham, NC, December 3, 2020.

Friedman, C., Allee, N., Delaney, B., Flynn, A., Silverstein, J., Sullivan, K., and Young, K. (2017). "The Science of Learning Health Systems: Foundations for a New Journal," Learning Health Systems, 1:e10020. doi: 10.1002/lrh2.10020.

Friedman, C. P., & Flynn, A. J. (2019). Computable knowledge: An imperative for learning health systems. Learning Health Systems, 3(4), e10203-n/a. https://doi.org/10.1002/lrh210203

Mehta, N., & Pandit, A. (2018). Concurrence of big data analytics and healthcare: A systematic review. *International journal of medical informatics*, *114*, 57-65.

Richesson, R. L. (2020). "Learning health systems, embedded research, and data standards-recommendations for healthcare system leaders," JAMIA Open, 3(4), 488- 491. https://doi.org/10.1093/jamiaopen/ooaa046

Wilkinson, M., Dumontier, M., Aalbersberg, U., et al. (2016). "The FAIR Guiding Principles of scientific data management and stewardship," *Scientific Data*, 3, article 160018. DOI https://doi.org/10.1038/sdata.2016.18

Williams, M., Richesson, R., Bray, B., Greenes, R., McIntosh, L., Middleton, B., Perry, G., Platt, J., and Shaffler, C. (2020). "Summary of Third Annual MCBK Public Meeting: Mobilzing Computable Biomedical Knowledge – Accelerating the second knowledge revolution," Learning Health Systems, 5:1 (Special Report)

APPENDIX A: Pre- and Post-Assessment Questions (from Pilot Website)

Pre- and Post-Class Assessment Survey – MCBK Pilot Class – December 2021 – January 2022
NAME: ____
Use a 5-point scale to respond to each question. Put your score in blank in front of question.
1 No ---------2 Not Well -------------3 Not Fully ---------------- 4 Well -----------------5 Very Well

_____1. **Can you define "mobilizing computable biomedical knowledge**" (MCBK)?

_____2. Do you know the MCBK Manifesto?

_____3. Can you explain the CYCLE OF LEARNING: Data to Knowledge (D2K), Knowledge to Performance (K2P) and Performance to Data (P2D)?

_____4. Can you list an author guideline for the ***Learning Health Systems* journal**?

_____5. Can you **define computable biomedical knowledge** and give an example?

_____6. Can you **define taxonomies**?

_____7. Can you **define metadata**?

_____8. Can you give an example of **bias in machine learning or data analysis**?

_____9. Can you explain **Ethics of Care or the FAIR principle?**

_____10. Can you describe tools for **advancing new research or entrepreneurial** development in health care?

_____11. Can you describe **PATTIE** (define it)?

_____12. Can you describe **AHRQ** (define) guidelines?

_____13. Can you explain the value of **biomedical data visualization**?

_____14. Can you describe **MCBK metadata** issues?

_____15. Can you describe the **nature of different research (data) network**?

_____16. Can you define OHDSI?

_____17. Can you summarize **NLM (define it) strategies** for CBK?

_____18. Can **you give examples of OER (and define**)?

_____19. Can you summarize **how to search for CBK** materials?

_____20. Can you define **"algorithmic justice"** and give an example?

Complete as pre-assessment by Dec. 16 and email to: dswain@nccu.edu.
Revise as new Word doc by Jan. 6 and email to: dswain@nccu.edu.

APPENDIX B: Spreadsheet on Pre- and Post-Assessments (Learning %)

QUESTION	Pre - MEAN	Post- MEAN	Pre- MEDIAN	Post- MEDIAN	% CHANGE	T-Test	Majority Post- Response
#11:PATTIE	1.8	4.15	2	4.5	130.56	- 14.78	Not fully - Very well
#17: NLM strategies	1.85	4.15	2	5	124.32	- 36.66	Very well
#19: Search for CBK	2.05	4.55	2	5	121.95	- 27.07	Very well
#12: Describe AHRQ	1.85	4.05	2	4	118.92	-11.4	Not fully
#4: Author guideline for LHS	2.05	4.35	2	5	112.2	-6.9	Well
#14: MCBK metadata issues	2.15	4.5	2	5	109.3	error	Very well
#9: Ethics of Care and FAIR	2.3	4.8	2	5.5	108.7	-7.45	Very well
#10: Entrepreneurial health tools	2.1	4.2	2	5	100	-9.74	Well
#16: Define OHDSI	1.9	3.8	2	4	100	5.57	Well
#1: Define MCBK	2.5	4.85	2.5	5.5	94	- 2.942	Very well
#15: Research (data) networks	2.15	4.15	2	5	93.02	-6.64	Well
#3: CYCLE OF LEARNING	2.55	4.7	2	5	84.31	- 21.88	Very well
#8: Bias in ML or data analysis	2.65	4.85	2	6	83.02	-6.47	Very well
#13: Biomedical data visualization	2.5	4.55	2	5	82	error	Well
#18: Examples of OER	2.5	4.5	2	5.5	80	-40	No
#20: "Algorithmic justice"	2.55	4.55	2	5	78.43	- 12.66	Well - Very well
#5: Define CBK & example	2.8	4.75	2.5	5	69.64	-7.38	Very well
#2: MCBK Manifesto	2.6	4.3	2.5	5	65.38	- 8.975	Well
#6: Define taxonomies	2.6	4.15	2	4.5	59.62	-5.78	Very well
#7: Define metadata	3.6	4.85	3.5	5	34.72	- 20.81	Very well

APPENDIX C: OER Recommendations from Assessment Spreadsheet

QUESTION	OER CBK Learning Recommendations
#1: Define MCBK	Teach definition of Mobilizing Biomedical Computable Knowledge; examples
#2: MCBK Manifesto	Teach historical start for Mobilizing Biomedical Computable Knowledge
#3: CYCLE OF LEARNING	Expand modules; emphasize systems (heart of MCBK effort)
#4: Author guideline for LHS	Retain as a primary learning focus
#5: Define CBK & example	Expand or integrate wth other modules
#6: Define taxonomies	Prepare as foundational module
#7: Define metadata	Prepare as foundational module
#8: Bias in ML or data analysis	Good module
#9: Ethics of Care and FAIR	Continue to use as MCBK principle
#10: Entrepreneurial health tools	Specialized program for developers
#11:PATTIE	Specialized NEW program for developers
#12: Describe AHRQ	Teach more about Agency for Healthcare Research & Quality
#13: Biomedical data visualization	Good module
#14: MCBK metadata issues	Retain for research and improvement
#15: Research (data) networks	Teach about research data network standards
#16: Define OHDSI	Retain to teach Observational Health Data Sciences & Informatics more
#17: NLM strategies	Strengthen partnership and update with NLM
#18: Examples of OER	Preliminary need to introduce format for commons: Online Educational Resources
#19: Search for CBK	Required learning to lea
#20: “Algorithmic justice”	Expand or integrate wth other modules

Track – Innovative Designs

Organizations as quantum: A Metaphor to Prepare for Proliferated Quantum Supremacy

STEFAN HOLTEL*
PricewaterhouseCoopers
Bernhard-Wicki-Straße 8, 80636 Munich, Germany
E-mail: stefan.holtel@pwc.com

Quantum physics surpasses human imagination. It totally contradicts everyday experiences. Even literal and mathematical explanations cannot substitute for a non-intuitive behavior that puzzles many of us. Commercial quantum computers are in reach within this decade. History provides examples that the appearance of new technologies brought metaphors to life that may explain up to that point poorly understood knowledge domains (e.g., "brains are hardware while thoughts are software"). The author describes why the proliferation of quantum computers will be accompanied by the rise of metaphors that explain quantum effects. For one, those might help to better communicate about the best use and consequences of quantum computers. But those metaphors could also shed light on organizational performance at all. To show a practical example, the author proposes how the use of quantum metaphors could help organizations to prepare for the quantum computing era - without being exact in predictions about most likely technical implementation of quantum capabilities.

1 Setting the Scene

This paper proposes a possibly surprising argument: The closer we get to the commercial availability of quantum computers, the more "mental pictures" of quantum effects will offer us deeper insights into understanding organizations at all. We will see a close link between the physical characteristics of quantum effects (Drummond, 2019) and the use of quantum metaphors to communicate about organizational issues (Steen, Dorst, Herrmann, Kaal, & Krennmayr, 2010). This contribution makes an effort to build a case for surprising links between superficially different knowledge domains (Shibayama & Wang, 2020).

1.1 Metaphors as Thinking Tools

The metaphor is far wiser than its author and so are many things.
Everything has its depths. He who has eyes sees [everything] in everything.
Georg Christoph Lichtenberg

In 1980, linguists George Lakoff and Mark Johnson released their seminal publication *Metaphors We Live By* which transformed into a classic for the field of metaphorology (Lakoff & Johnson, 1980). They argue that metaphors help us to make sense of abstract concepts like work, time, mental activity, and emotions by letting one draw on what s/he already knows from physical and social experiences. When humans want to express complex and abstract issues, they unconsciously revert to the linguistic structure of metaphor. Metaphors can help to reason about and get an understanding of difficult, fluffy-described knowledge domains. But they even shape our thinking in ways we would not expect.

A study found that a metaphor presented as a single word in an otherwise identical text ("monster" or "virus") had impact on how individuals approached social issues and collected knowledge to make "informed" choices (Thibodeau & Boroditsky, 2011). And the framing effect in this study had been even more subtle: Reflecting on their answers the participants did not identify the metaphor as a factor in their decision-making, instead they referred to "objective" data like numbers and statistics. These and other studies show that language metaphors create deep frame-consistent knowledge structures and inferences. Metaphors can play a malign influence on thinking and decision-making.

But metaphors may even play a beneficial role. Studies have shown their persuasive effects over the use of literal language (Sopory & Dillard, 2006). Some known application areas for such effects are found in such different areas ranging from business, medicine, organizational research, to understanding religious meanings (Ashford, 1989):

LEGO© Serious Play© (LSP) is a method to guide meetings, facilitate group communications, and moderate problem-solving (Kristiansen & Rassmusen, 2014) (Idle, 2018). By help of LEGO bricks participants are pushed through a series of questions that probe deeper into dedicated issues. Each participant creates three- dimensional LEGO models out of a small set of pieces. The models serve as metaphors to facilitate discussion and information exchange. By using metaphors, the group can solve problems more effectively. It leverages human senses to teach and listen and offers everyone his own voice by leveraging visual, aural, and kinesthetic abilities. LSP is a lingua franca for people of many backgrounds and different socioeconomic status.

- Psychotherapist Stefan Büchi invented *Pictures of Illness and Self-Measure* (PRISM), a straightforward visual tool that acts as a “metaphoric language” of human mental states (Büchi & Sensky, 1999). Partly by chance he stumbled upon its larger potential and proved its reliability and validity as a quantitative measure of human pain. Within a couple of years, PRISM use cases beyond health care have sparked interest. Research is ongoing to define general conceptualizations of PRISM that uniforms different realms (Holtel, 2013) (Sensky & Büchi, 2016).

- With his seminal book *Images of Organization* economist Gareth Morgan published a serial of metaphors with the stated desire to introduce instruments that shed light on organizational issues (Morgan, 1986, 2006). His metaphors range from obvious *“Organization as Machines”* and *“Organization as Brains”* to more surprising ones like *“Organizations as Flux and Transformation”* or *“Organizations as Psychic Prisons”* (Morgan, 1986, 2006, S. 11pp, 73pp, 251pp, 215pp). Morgan wanted to introduce metaphors as an analysis framework. But he even did not deny the predicaments of his approach: *“Images or metaphors only create partial ways of seeing, for in encouraging us to see and understand the world from one perspective they discourage us from seeing it from others.”* (Morgan, 1986, 2006, S. 27).

These examples provide a glimpse that metaphors can be customized to different purposes thereby serving in different knowledge domains and usage scenarios. The next chapter will focus on a dedicated category of the metaphors’ body, those emerging from technologies. They play a dedicated role in appearance and further use.

1.2 The Emergence of Technology Metaphors

New technology is not good or evil in and of itself.
It's all about how people choose to use it.
David Wong

Science tools push the boundaries of our knowledge. They bounce between discoveries and their justification, and range from oscillators to grounded theory. If science widely uses a new academic tool, it is more likely that even new theories emerge from daily use and will ultimately be adopted by the scientific community (Gigerenzer, 1992).

A similar dynamic happens for the greater in technology realm. History shows an impressive record of continual technological progress. A *General Purpose Technology* (GPT) characterizes a ground-breaking technology of creation and innovation that leaves large footprints in the long-term in economy, society, and politics. GPTs are important because their proliferation is often followed by strong economic growth (Jovanovic & Rousseau, 20065). Steam engines (in wide use ca. 1770-1920), electrification (since ca. 1880), automobiles (invented 1886), computers (available from the 1940s), or artificial intelligence (first mentioned 1956) can be considered as such GPTs. All emerged within the last two centuries.

Since the invention of quantum computing (first hypothesized 1982 by physicist Richard Feynman (Feynman, 1982)) we see more interest from different industries to exploit quantum physics (McKinsey & Company, 2021, S. 24-25, 33). But the deployment of a GPT usually unveils another pattern: The ubiquitous presence of a technology makes it a valuable metaphor that helps to explain phenomena in science and knowledge domains not well understood at that time (i.e. in lack of sufficiently working metaphors).

The history of cognitive science presents famous examples that bring evidence for this assumption. For example, respected academics have repeatedly utilized the current state of contemporary technologies as metaphors to describe and explain cognitive and mental processes of the brain (Radman, 1996):

- Polymath Gottfried Leibniz (1646-1716) asserted that purely material objects are incapable of thinking or seeing. He compared the brain with a mill (known as the “Mill Argument”) to argue that brains cannot be explained in terms of their biological properties alone. He states: *“Moreover, we must confess that perception, and what depends on it, is inexplicable in terms of mechanical reasons, that is, through shapes and motions”* (Leibniz, 1714, S. §17).
- Psychologist Sigmund Freud (1856-1939) borrowed notions from contemporary technology to understand the human brain. Before becoming the founding father of psychoanalysis, he worked as a neurologist. He conducted research and contributed to neuroscience as a new field of research. Freud detected the effects of electricity on nerves and muscles. He proved that grey matter in the spinal cord include cells that form the origin of sensory nerve roots. Another important discovery made by him was that white matter pathways link the spinal cord and the cerebellum (Lothane, 1998, S. 50). That living things differ from inanimate objects by presence of a non-physical entity was elaborated by Freud in his idea of psychodynamics (Bob, 2015).
- Philosopher John Searle was born in 1932 (roughly half a century after the invention of the telephone and its infrastructure). In his seminal work *Minds, Brains and Programs* he reflects on his understanding of brain functions: *“In my childhood we were always assured that the brain was a telephone switch-board. ('What else could it be?')”* (Searle, 1984, S. 44).
- Physiologist Edgar D. Adrian (1889-1977) accumulated knowledge of the nerve fiber in the 1920s (which brought him a Nobel Prize in 1932). He also influenced the invention of the vacuum tube as an amplifier to intercept coded signals. The vacuum tube analogy advanced into the idea that brains might be units that process information (Garson, 2015).

Taking these anecdotes into account it is plausible to consider that quantum computing mechanics will also provide new metaphors. Before jumping into these, we will take a closer look into the relationship of technology and organizations that put it into use.

1.3 Making Sense of Technologies in Organizations

“It is only when they go wrong that machines remind you how powerful they are.”
Clive James

When it comes to improving business processes, information technology plays a crucial role. But it can either help or hinder progress. The impact of information technology on business processes and its elements are not entirely known. That is because technology is not an objective fact but a product "socially created" (Orlikowski, 1992). Organizational theorist Wand Orlikowski examines information technology and business development with her concept of "technological duality". She considers an interdependence between human and machine affordances. Depending on the context, many factors enable or restrict the use of technology - which illustrates this duality. Looking into a technical environment, several elements may either enable or prevent organizational change. Technology and organizations are in permanent negotiations of their interdependent relationship and influence each other.

Project *Cybersyn* can be considered as an example to understand the duality of technology in play. It is part of historic events in Chilean politics (Medina, 2015): After a socialist transition, Salvador Allende became president of Chile, and a couple of young scientists built a complex computer- and telex-based infrastructure with the goal to control the Chilean economy. The plan included a holistic system design, controlled through a decentralized administration, advances in building rich human-computer interactions, and had been implemented on a national-wide telex network. Allende's government was violently overthrown and Cybersyn had never been fully implemented. But it left lessons for political and organizational leaders on the very close links between technology, economy, society, and politics.

What we can learn at the advent of quantum computing is that we will likely see a change in our understanding of processing power as compared to the first decades of computing history. The term "Quantum Supremacy" indicates the time when quantum computers are capable to resolve problems in reasonable timeframes that classic computer architecture cannot do. Once this goal is accomplished the sheer presence of more and more quantum computing power will shape and change our way to think and act in interaction with these machines and everything we experience around us.

2 Quantum Mechanics in a Nutshell

I, at any rate, am convinced that He [God] does not throw dice.
Albert Einstein

Not only does God play dice, but […] he sometimes throws them where they cannot be seen.
Stephen Hawking

Before we look at quantum metaphors, it is helpful to at least get a general understanding on how they work. Otherwise, we will not detect and match attributes of the source domain under consideration (quantum physics) to those of the destination domain (organizations). In this chapter the author makes an effort to explain the evidence made from physical experiments, also known as quantum effects, in simple metaphors.

Linguists Lakoff and Johnson state: *"In general, abstract concepts are defined metaphorically in terms of concepts that are more concrete and more clearly structured on their own terms – concepts like SPACE, MOTION, FOOD, OBJECTS, etc."* (Lakoff & Johnson, 1980). Quantum mechanics can be considered as such a general, abstract concept. It is a catch term that describes physical phenomena that uniquely occur at atomic levels. Nevertheless, they leave impact in the macroscopic world - which is the world humans experience with senses.

The first ideas to make use of quantum effects arose from theoretical thoughts theoretical physicist Richard Feynman outlined decades ago. He made the proposal to build computers that should make use of quantum effects to simulate quantum effects (sic!) (Feynman, 1982). His ideas were of hypothetical nature and he did not directly consider the development of all-purpose quantum computing architectures. Nevertheless, today we await commercial quantum computers in the late 2020s or early 2030s (National Academy of Sciences, 2018, S. 7-13pp) (McKinsey & Company, 2021, S. 38).

The far-reaching consequences of quantum physics can be considered as a late response to "Laplace's demon", a thought experiment about determinism: According to determinism, if someone (the demon) knows the precise location and momentum of every atom in the universe, their past and future values for any given time can be predicted; they can be calculated from the laws of classical mechanics (Laplace, 1814, 1902). But in 1925, physicist Werner Heisenberg proved that uncertainty is an inherent quality of quantum mechanics which disproved the possibility of a Laplace demon (Heisenberg, 1925).

Different technical approaches to quantum computing coexist and currently compete to realize useful architectures. But all share common denominators: Classical hardware works on the dichotomy of 0 and 1 where an information unit is called *Bit* and is coded as the dedicated difference between two states (von Neumann, 1945) (Hally, 2006). In contrary, quantum computing systems work with *Qubit* which is to handle a bunch of logical states at the same time (Rieffel & Polak, 2011).

2.1 Superposition

Qubits are the building blocks of quantum computers. But they are kind more sophisticated than Bits in storing much more information. A Bit may only hold dichotomic values, either 0 or 1, while a Qubit may contain "any value" between 0 and 1. Those Qubits are made up from controlled particles as well as the methods of controlling them.

Figure 1: Quantum superposition can be explained by positions on the sphere of a soccer ball

The state of superposition describes a Qubit that may exist in two distinct states at the same time, and it is conceivable for it to exist in both states. In physical terms, an electron may exist in two different quantum states, a spin-up and spin-down state. In case this electron is in superposition, it is both up and down simultaneously - a mystic mix of two states. It is only when this electron is *measured* that it falls out of its superposition and determines in one of two positions. Quantum algorithms make use of this superposition to harness its inherent calculation power.

A simple metaphor can explain this phenomenon: The sphere of a soccer ball is a good metaphor to present all weird states of a Qubit (Figure 1): Myriads of points on its surface represent possible states of a Qubit (at the same time before measurement). Only the top of the soccer ball (north) would be a clear 1 (after measurement!), the down of the soccer ball (south) would be a clear 0 (after measurement!).

2.2 Quantum Entanglement

If one knows, then no one knows.
Ludwig Wittgenstein

There is another weird phenomenon in quantum systems that is far beyond human imagination. It is called "Quantum Entanglement" and describes something known as the "nonlocality" that is the ability of two objects to instantaneously know about the state of each other – independent from their distance.

This surprising "distant connection" and "faster than light" relationship of partners appear to contradict Einstein relativity. But theoretical and experimental studies have proven this fact. There is a correlation of measurements between twins that are separated.

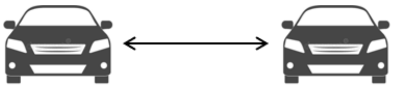


Figure 2: "Spooky action at a distance" demonstrated by two cars

We can model this phenomenon by a metaphor presents two cars (Figure 2). Both have been somehow linked to each other ("entangled"). Every time a car tinkerer makes a tune to his car (left), by magic another tune will somehow appear at the car of his friend (right). Even more weird: This happens instantly. And totally independent from distance. One car might parks in a garage in downtown Munich, the other might have traveled to Mars on a rocket and just arrived in Mare Erythraeum.

Physicist Albert Einstein couldn't make sense of this strange effect of quantum mechanics where two particles instantaneously interact over a distance because this contradicts his calculated limit of speed of light. In a letter to Max Born he coined it *"spooky action at a distance"* (*spukhafte Fernwirkung*) (Einstein & Born, 2004).

2.3 Quantum Instabilities

What is greatest challenge to build a quantum computer? It is the engineering puzzle to purely keep a quantum state long enough to make sense of it - and produce correct answers before it is destroyed. Storing a quantum state (particles in superposition) is very difficult. Any interaction with the universe disrupts it and can cause errors. That is why quantum computers are shielded electro-magnetically. That is why quantum computers are cooled down to almost absolute zero (which is minus 273 degree Celsius). The chandelier-like computing device is an expression of this problem: In fact, it is a refrigerator hanging from a ceiling (Figure 3). The tube of its cooling system makes its outer shape. The core computing unit is chip-size and sits at the bottom of the whole structure. This special structure is meant to guarantee at least microseconds to store information necessary for calculations. But even these precautions and preparations are not enough to guarantee valid measures.

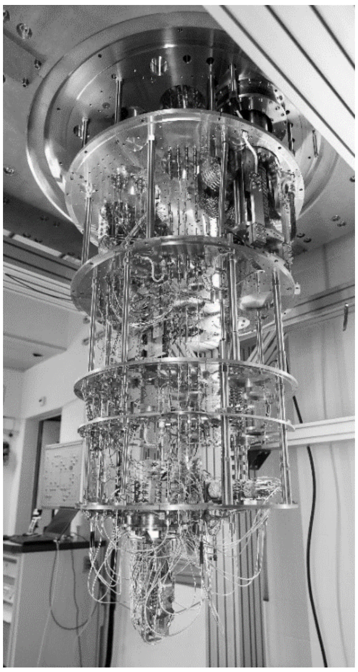


Figure 3: IBM Quantum Computer (IBM, 2017)

In a real sense, quantum calculations must be repeated many times to make them highly reliable. Doing them once would deliver uncertain results. The major caveat from quantum algorithms is its results are only correct in a statistical sense. That is every single result from a quantum computing algorithm can be wrong (if only within probabilities). It is of importance that quantum computers can never provide unambiguous answers. Results can be considered as "statistically correct". One major step in quantum computing technology will be the availability of self-correcting quantum algorithms to marginalize that source of failure (The Wall Street Journal, 2021).

Now that we have at least a rough idea of evident quantum effects we will switch over to the use of its metaphors in an organizational context.

3 Looking at Organizations as Quantum

We can notice increasing efforts to utilize a better understanding of quantum states as means to increase insights in other realms like economy (Orrell, 2019), human language (Aerts & Beltran, 2022), or even the simulation of life itself (Rasetti, 2017). The author now makes an effort to use quantum effects as a metaphor to better understand organizational dynamics.

Quantum mechanics challenges an assumption: *"It undermines scientific determinism, implying that human knowledge about the world is always incomplete, uncertain and highly contingent."* (Das, 2012). Inexactness ultimately challenges logical causality on a very basic level. Thus, this pattern holds a lesson that sheds light on organizations.

Three archetypical issues of organizational relevance can show that case. Each either possesses a long track in the history of organizational sciences, or companies are facing them today or will do likely in the future. All topics share a plethora of ongoing academic dispute and are often high priority at levels of corporate decision-makers.

3.1 Quantum Metaphors in Organizational Control

respice finem (look to the end)
Latin phrase

We are living in a world of uncertainty which students of an US military academy after the Cold War coined "VUCA" which is Volatility, Uncertainty, Complexity, and Ambiguity (Barber, 1992). But long before, economist Frank H. Knight already reconsidered the human perception of risk awareness. *"The key to the whole tangle will be found to lie in the notion of risk or uncertainty and the ambiguities concealed therein."* (Knight F. H., 1921). Knight distinguished between uncertainty (which is not predictable), and two variants of risks (either insurable or not insurable). The quantum effect of principle uncertainty to predict an outcome creates a valuable metaphor that reenacts the understanding of unpredictability in the business world.

Economist Saras D. Sarasvathy made the idea of unpredictability as a starting point to better understand entrepreneurship (Sarasvathy, 2001): In business, choices and actions are done on a regular basis. She therefore asked: What is the secret of success whenever a decision must be made? By performing a research with 27 successful entrepreneurs who founded firms ranging in size from USD 200 million up to USD 6.5 billion she listened intently when those entrepreneurs were faced with working through a problem set. This approach allowed her to learn more than just what participants decided. Sarasvathy learned about the way they did. She built a conceptual framework called "effectual action" which she described as a thought pattern experienced serial entrepreneurs utilize to thrive. She recognized two quite different ways to create business value: causal (or predictive) thinking and effectual thinking (or effectuation). There is merit in both views since they address quite distinct issues (Sarasvathy, 2009). Causal reasoning entails deciding on a goal and then securing the funding and personnel necessary to accomplish it. In contrast to mere forecasting, effectuation emphasizes to create circumstances on the resources someone has just currently available. The concept of effectuation is a strong challenge to conventional thinking of corporate planning. To be really effectuative an organization has to accept that it cannot *control* the future, but it can *anticipate* it (Sarasvathy, 2009). Such, smart organizations might be capable to profit from huge ambiguities they face (Wilkinson, 2006).

The quantum effects of principle unpredictability (i.e. presence and future) provides another way how to convey quantum metaphor to organizations. The next chapter elaborates this idea.

3.2 Quantum Metaphors in Organizational Cultures

Culture performance has a significant influence on business performance: According to a Deloitte poll, 82 percent of respondents agreed that culture may provide an edge over competitors (Deloitte , 2016). But decision-makers often consider cultural issues as a vague, mysterious hard-to-model relationship between cultural foundations and business impact.

The culture of organizations can be better understood by looking beyond "visible artifacts". It is to unveil essential beliefs that shine through them (which is related to the fact that quantum effects can only be observed by indirect measures because otherwise measures break down).

Economist Edgar Schein claimed that only digging beneath the surface of an organization reveals assumptions that lie at its very heart (Schein, 1984). He distinguished three levels to interpret an organizational culture: artifacts, espoused values, and assumptions (Figure 4) (Schein, 2004).

1. Artifacts: Visible artifacts of an organization are architecture, technology, office layout, dress codes, or visible behavior patterns like praise and blame. But this level of analysis is superficial. And there is some difficulty to interpret the results. *How* a group produces its environment and what patterns of behavior are evident among its members may be easy to define. But *why* a group operates in a specific manner can be sometimes difficult to interpret.
2. Espoused values: At the second level, Schein finds values that govern an organization’s behavior. It helps to understand why members behave the way they really do. But values may only be deduced via interviews or by content analysis of press releases, reports, or marketing campaigns. Additionally, values take into account people's justifications of their behavior. Underlying reasons may still remain hidden from them, either conscious or subconscious.
3. Assumptions: There is more work to be done if someone wants to get down to the core of a culture. It is to learn more about beliefs and behaviors of the organization under consideration. Unconscious assumptions must be explored because they guide how members perceive the organization and interact with each other, customers and clients. The more employees adapt to those assumptions, the less all are aware of them. Assumptions become a “second skin” to everyone who joins the organization.

Schein’s framework is a good example that “things are not how they shine”. We must accept that an organizational culture cannot not be planned, controlled, and outplayed in a simple way – even on an abstract level.

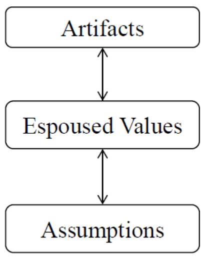


Figure 4: Three levels of interpreting organizational culture (according to (Schein, 2004))

In quantum-related terms, the "cultural state of a company" can be measured in terms of statistical likelihoods that specific attributes of this culture are enacted over the long run. At the same time, all attributes (here: artifacts, espoused values, assumptions) create momentum at the same time with no clear chance to snap into a static image. Everything remains in permanent "superstition" and might change at any given time – depending on everything that might penetrate the organizations from outside.

Only after important decisions are made by someone (i.e. a quantum measurements taken) an outsider gets an insight into the culture of an organization. Only by looking at a dedicated incident (e.g. a scandal shuttering the espoused values while a C-level manager get ousted due to political responsibility) shows whether an organization lives by its values.

3.3 Quantum Metaphors in Organizational Performance

You get what you measure
Proverb

Organizations follow the unwritten rule that performance management is a prerequisite to successful management. Fundamentally, this entails evaluating progress, reporting on it, and controlling performance constantly both at individual and corporate levels. Key performance indicators and metrics are used to measure performance. They pretend to represent how well enterprises are doing with respect to strategic goals and business objectives. But that comes with a flaw. When managers look at numbers and diagrams they look at a solely data-driven model of reality. It is impossible to measure reality in any detail, it remains feeding a model that more or less mirrors reality to a specific degree (Podniecks, 2010).

This problem becomes even larger if human beings are objects of measurement. They will most likely change behavior because they are monitored (which is the same phenomenon physicists experience in measure quantum states: any effort to measure a quantum state destroys it). The outcomes of the famous Hawthorne experiments is an example for such an analogy in organizational context (Gillespie, 1993): Researchers consulted Western Electric factory employees that built electrical relays at Hawthorne Works. The researchers carried out the now renowned "Lighting Research". A group of five women participated in a series of adjustments in work organization (such as alterations in rest intervals, changing lightning conditions) (Figure 5). The uncontrolled and methodologically weak research prevented clear conclusions for the primary goals.

Figure 5: Participants of the Hawthorne Experiments

But it left a surprising insight: When people became aware that they were being watched, they changed their behavior. The later coined "Hawthorne Effect" says that researchers' novelty and the attention these women got had contributed to gains in employees' productivity.

Scientific Management made an argument for a growing body in need and purpose of organizational measurements (Taylor, 2003). Again, a quantum effect lends us a helping hand to look at the Hawthorne outcome: It is impossible to measure states without destroying them, and we must carefully consider the purpose and limits of measuring organizational performance. Or as psychologist Warren Thorngate once stated: *"It is impossible for a theory of social behaviour to be simultaneously general, simple or parsimonious, and accurate."* (Thorngate, 1976, S. 406).

4 Discussion

The author reflected on two relevant issues to utilize quantum metaphors for the analysis of organizations, namely their general usefulness for this purpose.

4.1 The Usefulness of Quantum Metaphors

Within *Images of Organization*, Morgan introduced eight different metaphors to communicate about organizations, one dedicated chapter for each (Morgan, 1986, 2006). All metaphors are means to create a kind of social and cognitive congruence, a concept that arose in the wider context of education (Lockspeiser, O'Sullivan, Teherani, & Muller, 2008). The last paragraph Morgan devotes to a reflection on the "Strengths and Limitations" of each metaphor introduced. This chapter is the copycat to do the same for the quantum metaphor.

The biggest benefit of quantum metaphors is that "principle uncertainty" brings insights to the chances and risks of managing organizations beyond precise measurements. Accepting the categorial impossibility to "make precise measurements without destroying states to be measured" leaves managers relieved from making another effort but just pursue more promising methods. Once we have been deprived from certainty on measurement this could open new ways not yet gone in better dealing with uncertainty (Knight F. , 1921, 2014).

Quantum metaphors could become enablers to better accept and ultimately learn how to thrive in a world of strong uncertainty and weak predictability (Sarasvathy, 2001). They can serve as helpful cognitive maps (although not precise) to develop sensemaking and might provide new common ground of understanding in organizations (Weick, 1995). Traditional methods like project management could profit from a different understanding of complexity and the way to accomplish outcomes (Holtel & Heinen-Konschak, 2018).

A big disadvantage in the use of quantum metaphors lies in the fact that they scratch at the limits of human imagination and overwhelm our cognitive capabilities. This will become an even bigger challenge once from a multitude of quantum metaphors those particularly catchy ones will establish as instruments of discourse and everyday language. That is because for human it is challenging to switch between verbal and numerical presentations of probabilities (Witteman, Renooij, & Koele, 2007) (Figure 6). But sharing the same language is important in the ups and downs of organizations to achieve e.g. mastery in their efforts of digital transformation processes (Hay, Redman, Yonke John, & Zachman, 2020) (Hay, Redman, Yonke, & Zachman, 2021).

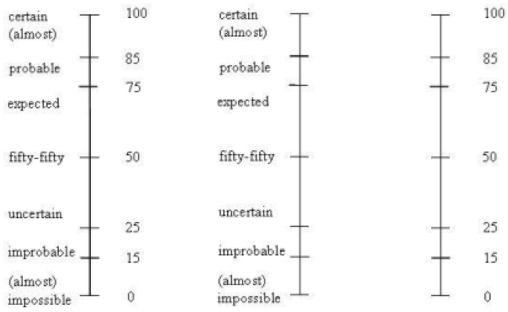


Figure 6: The probability scales: The verbal-numerical (a), the verbal (b)and the numerical (c) probability scale (Witteman, Renooij, & Koele, 2007)

Another challenge is that although quantum technology might be available in a few years there remains a plethora of organizational situations where causal planning is rationale, effective, and makes utmost sense. It would be too much of a good thing if principle inexactness in the world would wipe out its causal counterpart where planning and monitoring remains meaningful with measurable outcomes (e.g. to plan and maintain large long-term infrastructure projects).

4.2 Towards Practical Applications of Organization Quantum Metaphors

As business leaders navigate and co-create the "New Normal" at the end of the COVID pandemic, they expect a world of ongoing and even increasing uncertainty (Harvard Business Review, 2021). The challenge remains and even increases to navigate a corporation through unchartered territory. The chances and risks emerging from the commercial availability of quantum computers is a business issue many organizations must address in time. Quantum computer capabilities might improve classic algorithms which increases general data processing speed (Grover, 1999). But quantum computers are poised to turn upside down some areas of businesses with a high impact on technology, economy, politics and society (e.g. cryptography) (Goel, Girma, & Garuba, 2007).

Can companies prepare for quantum computing more than a decade before it ferries across from academic islands to commercial mainland?

To grasp the future of quantum computing it is not mandatory to predict it. It is much more critical to take an active role in molding its appearance (Sarasvathy, 2001). Forecasting techniques to write scenarios about the future are valuable to gain insights into possible futures (Schwartz, 1991) (Lindgren & Bandhold, 2009). Well-written scenarios prepares an organizational environment by coming up with plausible but very unlikely (sic!) scenarios for what an enterprise's future might hold. This can be regarded as already preparing (at least mentally not after the fact) for a breadth of possible future outcomes in advance of quantum proliferation. Another advantage of quantum computing scenarios lies in the fact that they are based on key factors that could be easily overlooked and forgotten if not maintained well in advance. For example, there is less academic or business discourse about the undisputed mismatch between prompt demand of quantum-related talent (McKinsey & Company, 2021, S. 37).

It may seem not worth time and effort but developing a set of plausible assumptions is the only thing corporations can do for the long run. At the moment, writing scenarios seem to be best option available to prepare for a future of quantum computing without making big bets. Their benefits are rather obvious within acceptable cost and time: Preparing for a worst-case situation in a quantum computing world could make the difference between life and death. Preparing for best-case scenarios could exploit its opportunities as quickly as possible.

But scenario planning brings an inevitable caveat: On one hand it takes time and effort to create quality-rich scenarios. But more important, they only blossom through a stream of unique perspectives and from many angles. Experts from different domains contribute and their answers must finally put together. That might be the reason we only see scattered efforts to leverage the power of scenario writing for a better understanding of a world of quantum computing (McKinsey & Company, 2021, S. 42-44).

With the help of scenario writing we might even catch a glimpse of a possible symbiosis of two general-purpose technologies that are under way at the moment, namely quantum computing and artificial intelligence. We might not call this era "Quantum Revolution" in the end but nevertheless will likely experience progress and breakthroughs that lie beyond our imagination today.

5 Acknowledgements

This paper is the outcome of long cross-disciplinary work on hard to crack organizational issues. On this journey, many people have challenged the author to reconsider naïve assumptions the individual, cognitive working tools, and organizations. The arrival of quantum computing will be the next turn on this journey that will most likely override the author's current understanding about the relationship between humans and machines.

References

Aerts, D., & Beltran, L. (2021, December). Are Words the Quanta of Human Language? Extending the Domain of Quantum Cognition. *Entropy, 24*(6). doi:10.3390/e24010006

Aerts, L., & Beltran, L. (2022). Are Words the Quanta of Human Language? Extending. *Entropy, 24*(1). doi:10.3390/e24010006

Ashford, W. B. (1989). Light of Reason, Light of Nature. Catholic and Protestant Metaphors of Scientific Knowledge. *Science in Context, 3*(1), 89-107. doi:10.1017/S0269889700000739

Barber, H. F. (1992). Developing Strategic Leadership: The US Army War College Experience. *Journal of Management Development, 11*(6), 4-12. doi:10.1108/02621719210018208

Bob, P. (2015). *The Brain and Conscious Unity.* Heidelberg: Springer. doi:10.1007/978-1-4939-2700-5

Büchi, S., & Sensky, T. (1999). PRISM: Pictorial Representation of Illness and Self-Measure. A brief nonverbal measure of illness impact and therapeutic aid in psychosomatic medicine. *Psychomatics, 40*(4), 314-320. doi:10.1016/S0033-3182(99)71225-9

Das, S. (2012). Impossible Exactness. In J. Brookman, *This Explains Everything: Deep, Beautiful, and Elegant Theories of How the World Works* (p. 432). New York, NY: Harper Perennial.

Deloitte . (2016). *Global Human Capital Trends.* Westlake, TX: Deloitte University Press.

Drummond, B. (2019). Understanding quantum mechanics: a review and synthesis in precise language. *Open Physics*. doi:10.1515/phys-2019-0045

Einstein, A., & Born, M. (2004). *The Born-Einstein Letters 1916-1955.* Palgrave Macmillian.

Feynman, R. (1982). Simulating physics with computers. *Int J Theor Phys, 21*, 467-488. doi:10.1007/BF02650179

Garson, J. (2015). The Birth of Information in the Brain: Edgar Adrian and the Vacuum Tube. *Science in Context, 28*(1), 31-52. doi:10.1017/S0269889714000313

Gigerenzer, G. (1992). Discovery in Cognitive Psychology: New Tools Inspire New Theories. *Science in Context, 5*(2), 329-350. doi:10.1017/S0269889700001216

Gillespie, R. (1993). *Manufacturing Knowledge: A History of the Hawthorne Experiments.* Cambridge, MA: Cambridge University Press.

Goel, R., Girma, A., & Garuba, M. (2007). Research Directions in Quantum Cryptography. *ourth International Conference on Information Technology: New Generations (ITNG 2007).* Las Vegas, NV. doi:10.1109/ITNG.2007.166

Grover, L. K. (1999). QUANTUM COMPUTING How the weird logic of the subatomic world could make it possible for machines to calculate millions of times faster than they do today. *The Sciences*, 24-30.

Hally, M. (2006). *Electronic Brains: Stories from the Dawn of the Computer Age.* London, UK: Granta Books.

Harvard Business Review. (2021, November 23). 3 Strategies to Help Employees Thrive in the New "Normal". *Harvard Business Review*. Retrieved from https://hbr.org/2021/11/3-strategies-to-help-employees-thrive-in-the-new-normal

Harvard Business Review. (2021, Dec 14). Effective Digital Transformation Depends on a Shared Language. *Harvard Business Review*. Retrieved from https://hbr.org/2021/12/effective-digital-transformation-relies-on-a-shared-language

Hay, D. C., Redman, T. C., Yonke John, C. L., & Zachman, A. (2020). *Developing and Adopting a Common Language: What's Required from an Organizational Perspective.*

Hay, D. C., Redman, T. C., Yonke, C. L., & Zachman, J. A. (2021, 12 14). Language, Effective Digital Transformation Depends on a Shared. *Havard Business Review*. Retrieved from https://hbr.org/2021/12/effective-digital-transformation-relies-on-a-shared-language

Heisenberg, W. (1925). Über quantentheoretische Umdeutung kinematischer und mechanischer Beziehungen. *Zeitschrift für Physik, 33*, 879-893.

Holtel, S. (2013). A pattern language on how to apply visual instrument PRISM. *EuroPLoP '13: Proceedings of the 18th European Conference on Pattern Languages of Program.* Irsee, Germany. doi:10.1145/2739011.2739031

Holtel, S. (2020). *KI-volution: Künstliche Intelligenz einfach erklärt für alle.* München: Redline Verlag.

Holtel, S., & Heinen-Konschak, E. (2018). How to Root Effectuation Skills in Your Project Team: Improve Project Performance by Balancing Uncertainty. *Proceedings of 23rd EuroPLoP*. doi:10.1145/3282308.3282310

IBM. (2017, November 2). The intricate beauty of an IBM quantum computer. Retrieved from https://de-de.facebook.com/IBM/posts/1521835057906771

Idle, D. (2018). An introduction to Lego Serious Play. *medium.com*.

Jovanovic, B., & Rousseau, P. L. (20065). General Purpose Technologies. In P. Aghion, & S. N. Durlauf, *Handbook of Economic Growth* (pp. 1181-1224). Elsevier. doi:10.1016/S1574-0684(05)01018-X

Knight, F. (1921, 2014). *Risk, Uncertainty and Profit.* Eastford, CT: Martino Fine Books.

Knight, F. H. (1921). *Risk, Uncertainty, and Profits.* Cambridge: The Riverside Press.

Kristiansen, P., & Rassmusen, R. (2014). *Building a Better Business Using the Lego Serious Play Method.*

Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By.* Chicago: University of Chicago Press.

Lakoff, G., & Johnson, M. (1980). The Metaphorical Structure of the Human Conceptual System. *Cognitive Science, 4*, 195-208.

Laplace, P.-S. (1814, 1902). *A Philosophy on Probabilities.* London: John Wiely & Son.

Leibniz, G. W. (1714). *Monadologie.*

Lindgren, M., & Bandhold, H. (2009). *Scenario Planning - Revised and Updated.* London: Palgrave Macmillan. doi:10.1057/9780230233584

Lockspeiser, M., O'Sullivan, P., Teherani, A., & Muller, J. (2008). Understanding the experience of being taught by peers: the value of social and cognitive congruence. *Adv Health Sci Edu Theory Pract, 13*(3), 361-372. doi:10.1007/s10459-006-9049-8

Lothane, H. (1998). Freud's 1895 Project: From Mind to Brain and Back Again. *Annals of the New York Academy of Sciences, 843*(1), 43-65. doi:10.1111/j.1749-6632.1998.tb08204.x

McKinsey & Company. (2021). *The Quantum Technology Monitor.* McKinsey & Company. Retrieved from https://www.mckinsey.de/~/media/mckinsey/locations/europe%20and%20middle%20east/deutschland/news/presse/2021/2021-09-22%20quantum%20technology%20monitor/mckinsey_quantum_technology_monitor_2021_vf.pdf

Medina, E. (2015). *Cybernetic Revolutionaries. Technology and Politics in Allende's Chile.* Cambridge, MA: The MIT Press.

Morgan, G. (1986, 2006). *Images of Organization.* London: SAGE Publications. National Academy of Sciences. (2018). *Quantum Computing: Progress and Prospects.* National Academy of Sciences. doi:10.17226/25196

Orlikowski, W. J. (1992). The Duality of Technology: Rethinking the Concept of Technology in Organizations. *Organization Science, 3*(3), 398-427. Retrieved from https://www.jstor.org/stable/2635280

Orrell, D. (2019). *Quantum Economics: The New Science of Money.* London: Fabr & Faber.

Patel, T. (2018, Sep 15). What is “so called spooky action at a distance”? *medium.com*. Retrieved from https://medium.com/@trushitpatel172/what-is-so-called-spooky-action-at-a-distance-2d91a3b39b0f

Podniecks, K. (2010). *The Limits of Modeling.*

Radman, Z. (1996). *Metaphors: Figures of the MInd.* Heidelberg: Springer Netherlands.

Rasetti, M. (2017). The 'Life Machine': A Quantum Metaphor for Living Matter. *International Journal of Theoretical Physics, 56*(1), 145-167. doi:10.1007/s10773-016-3177-6

Rieffel, E., & Polak, W. (2011). *Quantum Computing. A Gentle Introduction.* Cambridge, MA: MIT Press.

Sarasvathy, S. D. (2001). Causation and Effectuation: Toward a Theoretical Shift From Economic Inevitability to Entrepreneurial Contigency. *Academy of Management Review, 26*(2), 243-263.

Sarasvathy, S. D. (2009). *Effectuation: Elements of Entrepreneurial Expertise.* Northhampton, MA: Edward Elgar Publishing.

Schein, E. H. (1984). Coming to a New Awareness of Organizational Culture. *MITSloan Management Review, Winter 1984*. Retrieved from https://sloanreview.mit.edu/article/coming-to-a-new-awareness-of-organizational-culture/

Schein, E. H. (2004). *Organizational Culture and Leadership.* London, UK: John Wiley & Sons.

Schwartz, P. (1991). *The Art of Long View.* New York, NY: Doubleday.

Searle, J. (1984). *Minds Brains and Science.* Cambridge, MA: Harvard University Press.

Sensky, T., & Büchi, S. (2016). PRISM, a Novel Visual Metaphor Measuring Personally Salient Appraisals, Attitudes and Decision-Making: Qualitative Evidence Synthesis. *PLOS ONE*. doi:10.1371/journal.pone.0156284

Shibayama, S., & Wang, J. (2020). Measuring originality in science. *Scientometrics, 122*, 409-427. doi:10.1007/s11192-019-03263-0

Sopory, P., & Dillard, J. P. (2006). The Persuasive Effects of Metaphor: A Meta-Analysis. *Human Communication Research, 28*(3), 382-419. doi:10.1111/j.1468-2958.2002.tb00813.x

Steen, G., Dorst, L., Herrmann, J. B., Kaal, A., & Krennmayr, T. (2010). Metaphor in usage. *Cognitive Linguistics, 21*(4). doi:10.1515/cogl.2010.024

Taylor, F. W. (2003). *Scientific Management.* London, UK: Routledge.

The Wall Street Journal. (2021, May 18). Google Aims for Commercial-Grade Quantum Computer by 2029. *The Wall Street Journal*. Retrieved from https://www.wsj.com/articles/google-aims-for-commercial-grade-quantum-computer-by-2029-11621359156

Thibodeau, P., & Boroditsky, L. (2011). Metaphors We Think With: The Role of Metaphor in Reasonin. *PLoS ONE, 6*(2). doi:10.1371/journal.pone.0016782

Thorngate, W. (1976). "In general" vs. "it depends": Some comments on the Gergen-Schlenker debate. *Personality and Social Psychology Bulletin, 2*(4), 404-410. doi:10.1177/014616727600200413

von Neumann, J. (1945). *First Draft of a Report on the EDVAC.* Retrieved from http://web.mit.edu/STS.035/www/PDFs/edvac.pdf

Weick, K. (1995). *Sensemaking in the Organization.* London, UK: SAGE Publications.

Wilkinson, D. (2006). *The Ambiguity Advantage: What Great Leaders are Great At.* London: Palgrave Macmillan.

Witteman, C., Renooij, S., & Koele, P. (2007). Medicine in words and numbers: A cross-sectional survey comparing probability assessment scales. *BMC Medical Informatics and Decision Making, 7*(1). doi:10.1186/1472-6947-7-13

From digital Twins to Digital Selves and Beyond – Perspectives for Social Modeling in a Trans-Humanist World

FRANZ BARACHINI
Bisamberggasse 14, 2103 Langenzersdorf, Austria
E-mail: franz.barachini@gmail.com

CHRISTIAN STARY
Institut für Wirtschaftsinformatik-Communications Engineering, Johannes Kepler University, Altenbergerestrasse 69, 4040 Linz, Austria
E-mail: christian.stary@jku.at

We discuss the engineering of so-called 'humanoid' socio-technical settings mimicking cognitive and social skills. We discuss the influence of Artificial Intelligence (AI) on digital selves. We introduce relevant approaches to awareness and the emergence of consciousness in artificial agents. We define mandatory properties of digital selves so that digital societies can be modeled and we present our thoughts on the role of KM in a trans-humanist world.

1. Essential Concepts and Background

We start out explaining some fundamental concepts and raising fundamental questions with respect to artificial intelligence.

1.1. Digital Twins, Digital Selves and the Cyberspace

Grieves, who had raised the term *digital twin* in 2003, in his white paper introduces the concept of a "Digital Twin" as a virtual representation of what has been produced (Grieves, 2014). Given the background in industrial production twins developers aim to increase transparency, and trigger efficiency gains, mostly in terms of faster process execution times, increased flexibility of organizing processes, and higher levels of security and safety. Gartner analysts assume that half of all industrial companies use digital twins by now.

As such, digital twins do not only designate a new technology, but rather stand for the methodological integration of heterogeneous components (cf. Perno et al., 2022, Stary, 2021), requiring protocols for exchanging even large amount of data between various technologies. Originally designed as virtual double of physical products, digital twins are increasingly becoming digital replicas of cyber-physical systems. Not only the simulation of processes over the entire life cycle of a product, but also envisioned architectures become core activities when cyber-physical systems evolve (cf. Zhuang et al., 2017, Grieves, 2019). A digital twin is consequently a virtual representation of a process, service or product with parts of it in the real world. There are no predefined limits to what developers transform by digital means and represent in the virtual world: Objects like planes including all kinds of operation characteristics, regions or urban infrastructures, or biological processes copying the human organism. Developers simulate functional behavior with the digital twin in order to determine how capabilities can be changed depending on different parameters. This information helps to modify any product or service in a purely virtual manner before instantiating it for operation.

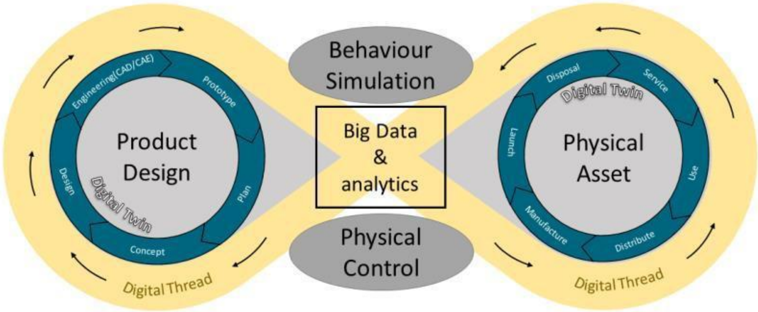


Fig.1: Digital Twin - Evolvement and variants

Figure 1 reveals the development of the concept in terms of its origin in product design towards digital threads allowing controlling of the behavior synchronized between digital and physical appearance based on data analytics, both for operation and further development (simulation of behavior).

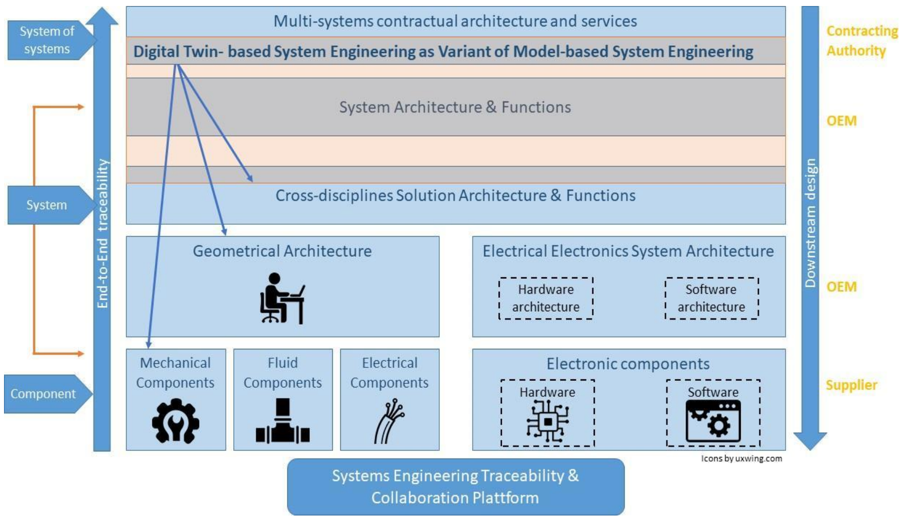


Fig.2: Systems of Systems Architecture propagating to components

Figure 2 shows the various layers including design and manufacturing in model-based engineering environments. The top layer comprises multi-systems contractual architectures and services. The digital twin-based engineering can be seen as a variant of model-based engineering. In the middle layer cross-discipline solution architecture and functions are addressed. The bottom layer comprises components of the real world.

A crucial aspect to digital twins concerns accessibility and persistence in terms of active sharing of model and real-time information. They feature what could be termed Open Economy. Digital twins operate in a transparent way. Due to their virtual representation, any digital twin enables everyone involved in the development process or concerned ecosystem evolvement - for the product, for example, manufacturer and buyer –to simultaneously access the current development or operational status. These insights can have immediate effects on connected actors or network partners.

Following the original concept, essential building blocks of a digital twin are physical products or elements that are equipped with Internet-based sensors. These sensors record numerous data from construction and operation in real time once the product is in use. A collector component records all behavior data and stores it in a centralized repository, mainly using cloud services and some digital platform for processing these data. Intelligent actors contain AI algorithms to evaluate the data. Actors could be equipped with interactive access, e.g., in product design with 3D technology or virtual reality, to visualize data and to facilitate simulating certain product properties.

Developers build digital twins for the entire life cycle of a product and all roundtrips of a product or service. These digital representations can accompany as a mirror a product or service over its entire life cycle. It thus provides valuable information about the way in which consumers use products. When such monitoring runs throughout operation the lifecycle of a product, massive information becomes available.

In addition to the continuous optimization of the product during operation, the feedback of the usage data to the producer manufacturer - often also called digital thread - enables important findings for further development and the concerned designers and engineers. This knowledge is helpful, e.g., when it comes to adapting product features better to the needs of the users and costumers.

The great potential of digital twins can be illustrated by various examples. Most cases focus on virtual models created on the basis of existing operational and construction data to study the behavior of certain parts at different conditions. Due to virtual prototyping, engineers and designers are to work much closer together across departmental boundaries. Consequently, workflows can be streamlined before put to operation.

A digital self by definition is an intelligent digital twin mimicking cognitive and social skills. In the literature we talk about humanoids, trans-humans or intelligent agents. In its last consequence it is a digital replica of a human being.

A tri-partite framework has been developed to capture human-related issues, organizational elements, and technologies including their mutual relations. We adopt this well-established conceptualization for complex system analysis and development (cf. Oppl et al., 2019) for digitalization of work (processes), and Ackerman et al., 2018) for socio-technical transformation of relevant systems for the society.

The adoption of the framework extends the human perspective to the individual perspective on trans-human developments enabled by technology. The technology perspective recognizes the cyber-physical nature of upcoming ecosystems. The collective perspective includes the socio-technical structuring of individual elements towards a systemic whole (see Figure 3).

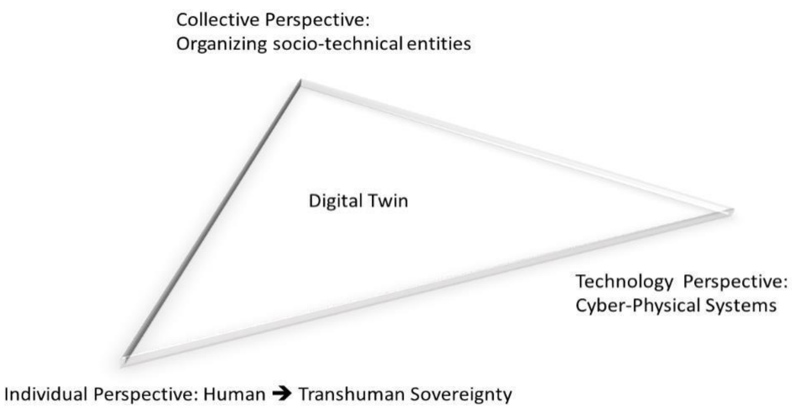


Fig.3: Digital twin development as socio-technical endeavor

It depends on progress in Artificial Intelligence (AI) to what extend digital twins might evolve to digital selves mimicking cognitive and social skills.

1.2. The Role of AI

In the year 1950 Allen Turing (Turing, 1950) posed a fundamental question: "Can machines think?". This question inspired numerous scientists and the competition trying to solve the Turing test has begun. This was the advent of Artificial Intelligence (AI).

Although Joseph Weizenbaum (Weizenbaum, 1976) was convinced that a computer will never gain wisdom he implemented one of the first AI programs, the ELIZA program. The rule-based program, implemented in the year 1966 used pattern matching to the written statements of patients to figure out its replies. Although very simple, the program delivered astonishing results. Users had the feeling as if they were talking to a genuine psychotherapist. However, after a while the program ended in loops because there were only a limited number of rules which were able to identify limited numbers of different words of the sentences entered via a console.

In the nineteenth we implemented a version of ELIZA in our lab using the forward chaining production system language PAMELA (Barachini, 1991). The more rules we added to our knowledge base the longer the user believed that he was talking to a psychotherapist. However, after a while the game came to an end. Experts assert that today's *chat-bots* are not better than the rule-based programs written in the seventieth.

Cognitive architectures like GPS (1972) by Allen Newell and Herb Simon (Newell & Simon, 1972) or SOAR from John Laird et al. (Laird et al., 2012, 1987) used production systems[1] to implement means-ends analysis so that more general problems could be solved. The reinforced learning algorithm in SOAR, which dynamically modified the weight of rules, and the chunking mechanism, which enabled primitive learning were able to solve planning problems. A Truth Maintenance System guaranteed the consistent status of the working memory. The architecture is still used as a basis for cognitive simulations.

For rule-based programs applied to a very limited scope of application the computer pioneer Edward Feigenbaum (Feigenbaum, 1981, 1982) coined the word expert system. The first expert systems were applied to medical problems before they were broadly used in industry. Expert systems as a whole were very successful. Nevertheless, cognitive learning had its limits.

In parallel to the above described developments search algorithms have been developed. They are able to prune huge search spaces. In combination with heuristics and high-performance supercomputers they are able to solve complex problems. Some computer scientists believe that playing chess or Go is a good measurement for the effectiveness of Artificial Intelligence. In the year 1997 IBM's deep blue computer has beaten the chess champion Kasparov. However, Go turned out to be a harder nut to crack. In this case just speeding up the hardware and the implementation of search algorithms was not sufficient since the search spaces are much more complex than in chess.

Another paradigm, namely artificial neural network processing, helped to solve the problem. Google's AlphaGo Zero program used self-training reinforcement learning. AlphaGo Zero played itself in hundreds of millions of games such that it could measure positions more intuitively and collect its knowledge in neural networks. In a way the network masters the game without teaching the rules. The so called deep neural network was implemented on a single machine with 4 tensor processing units. In May 2017, the program could beat the world champion.

[1]Production systems consist of rules, a working memory (facts) and an inference engine

Many other applications are being developed with similar kinds of neural networks. Among them is AlphaFold which was able to predict the 3D structure of proteins in the year 2020. There are approximately 10^{300} possible combinations for a single protein. It would therefore not be possible to solve this protein-folding problem with brute force algorithms in a reasonable time period, except for quantum computers which are in the experimental phase right now.

Similar good results were presented by IBM's Watson supercomputer in 2011. This computer was doing better than any human when playing the *jeopardy* game.

Encyclopedias, dictionaries and thesauri were stored in local memory. The machine used different paradigms, among them neural networks during problem solving. Numerous commercial and medical applications are currently on its way.

According to our observations, artificial neural networks were the last cornerstone for artificial learning. They go back to Frank Rosenblatt (Rosenblatt, 1958) and to Hebb's rule (Hebb, 1949). The rule explains human learning, in which simultaneous activation of cells leads to pronounced increases in synaptic strength between those cells. Much later, Teuvio Kohonen and others (Kohonen et al., 1982, 1984, 1991) presented forward and backward propagation algorithms for artificial neural networks. Having a closer look to the near future, autonomous cars, robots and drones rely on this technology as well.

Whenever an article about a new AI driven product is published, one can be confident that in most of the cases neural networks are part of the game. However, the application of neural networks follows no cooking recipe. It is rather an experimental try- and error procedure until an appropriate network consisting of the right number of nodes, layers and propagation functions is working properly. Furthermore, engineers need to figure out, how to present the data to the network and how to get the right sort of output that can be translated back into the real world.

Undoubtedly, deep learning with neural networks has brought new insights. AlphaGo Zero was able to produce strategies which were not known and applied from humans before. The same is true for radiological AI programs. They interpret pixels in different way than humans. Therefore, we suspect that artificial intelligence deviates from human intelligence. However, if AlphaGo Zero would be confronted with a change of the board size during the game, the program would fail to deliver meaningful strategies. Thus, robustness is still a challenge for neural network applications. But there is no doubt that artificial intelligence can produce better and other results than human intelligence in limited domains.

Today's implementations combine neural networks with rule-based, object oriented, procedural and functional approaches. However, the mix of the paradigms and the increasing architectural complexity makes software vulnerable and error prone. Gödel has shown that the set of Gödel numbers of functions whose domains belong to a class of recursively enumerable sets is either empty or contains all recursively enumerable sets. In other words, only trivial properties of programs are algorithmically decidable. Thus, the logical correctness of a program can be guaranteed only for very simple code pieces. Therefore, we need to implement redundancy in our architectures, similar as it is provisioned in our natural brains.

When we come back on Turing's question if a machine can think then the answer is ambivalent at the moment. Additional research in neuroscience is necessary. As long as neuroscience is not able to isolate the process that gives rise to consciousness in human brains, it will not be possible to create machines that have consciousness and mind, unless consciousness is a result of the thinking process.

John Searle's (Searle, 1980) Chinese room experiment clearly underpins our position. On the other hand, the Chinese room experiment encourages us to simulate understanding on computers in the sense of a weak AI. There is a difference whether the computer simulates a property or whether the machine understands what the simulation is about. In the latter case we speak about strong AI.

The Turing test requires that a machine must be able to execute all human behaviors, regardless of whether they are intelligent or not. That means that a machine that mimic thinking must include emotions, cheating, lying etc. Mind, understanding and consciousness are elemental properties of thinking.

2. Consciousness of Digital Selves

If we try to model the human brain on a digital machine is this copy identical to the real brain? Is the digital self capable of thinking? Does it have its own personality? Will consciousness emerge as a consequence of thinking or is it rather an illusion? Is there even a free will?

We can only speculate about the answers to these questions since to our best knowledge there is no commonly approved theory of mind existing yet. However, neuroscientists are making progress in brain research. We know today that the human brain is organized in a hierarchy. Dehaene and Changeux (Dehaene, 2011) investigated the transition from unconscious to conscious perception. They showed that according to Baar's theory (Baar, 1988) this what we experience as perception is caused through the intensification of a single information entity which is then simultaneously transferred in different parts of the cerebral cortex. By applying functional MRT and EEGs Dehaene discovered that the transformation process triggers certain synchronized rhythms of electrical activities. Presumably these rhythms stimulate consciousness by activating a network of pyramid neurons located in the parietal and frontal brain regions which in turn cause top-down amplification. Dehaene discovered this phenomenon while experimenting with the picture recognition process of the human brain. Similar results were obtained during experiments with the tactile sense. It can be concluded that there is a threshold between unconscious data processing and conscious data processing.

Libet (Libet, 2005) also discovered that a lot of the brain activities are not conscious at all. Based on his experiments Libet concluded that our awareness of decision making appears to be an illusion. According to Libet there is also no free will.

Neurobiologists suspect that changes in the hormone level not only influence feelings and spirit but also the function of the brain. Obviously, parts of the hippocampus are changing its structure and its activity during specific hormone cycles of women. The implications of these observations are not yet well understood.

Eric Kandel (Kandel, 2018) presents a good overview how art is interpreted by the human brain. As computer scientists, we also find Kurzweil's (Kurzweil, 2012) elaborations about the creation of a mind promising.

In order to mimic the hierarchies of the human brain, Kurzweil proposes the implementation of Hidden Markov models in combination with self-learning neural networks. He is convinced that before 2050 or even earlier, we will bring the full power of human brains on silicon substrates. In contrast to our brain artificial neural networks will learn much quicker. This is due to the fact that during the learning process the human brain takes time to grow new dendrites. Technologically we are able to construct artificial neural networks consisting of more neurons and synapses than the human brain. Humans cannot expand the size of brains on biological level in a short time. But thanks to exponential technological growth we might do it with artificial neural networks in the farer future. Kurzweil has certainly a point there.

The question above, whether consciousness is an emergent property is still not solved. It might be emergent if we follow Deahene's observations. If we consider e.g. hormones as important influencers then we need to understand also their behavior. As soon as we understand their influence, we can design an algorithm which can be executed as a Turing-machine on a real computer.

If consciousness follows an algorithm somewhere in the neo-cortex, what has not yet been proven then we can simulate also consciousness. If consciousness emerges automatically as a result of low-level perception processes than we will wait and see the results of virtual humans (humanoids) or virtual agents in a few decades.

We believe that Kurzweil's ideas will bring substantial progress in that sense that a good percentage of a brain will be simulated in silicon substrate in a few decades. And it could easily be that this brain might outperform average brains in certain general tasks due to quicker expansion options. Much more artificial neurons and synaptic connections yield more brain capacity and the presentation of several millions of examples in a short time period yield faster learning cycles. However, as long as the influence of electro-chemical processes on synaptic circuits is not fully understood the silicon brain might probably lack certain high-level functionalities. At the moment robots can simulate the behavior of approximately 4 years old children.

Since a digital self is a simulation, digital selves are different from one self, similar as human twins differ from each other. There is a difference whether the computer simulates a property or whether the machine understands what the simulation is about. It is therefore an illusion to think that we can replicate ourselves and thus exist for eternity.

Artificial neural networks are producing other and sometimes better strategies for problem solving, as explained when addressing the AlphaGO Zero program. Consequently, we will face an artificial intelligence which deviates from human intelligence. Thus, a digital self might develop its own artificial personality, at best.

3. Properties of Digital Selves

Trans-humanism is about evolving intelligent life beyond its currently human form and overcoming human limitations by means of science and technology. Albeit claims to guide trans-human development by life-promoting principles and values (More, 1990), advocating the improvement of human capacities through advanced technologies has triggered intense discussions about future IT and arte-fact developments. In particular, Bostrom (2009, 2014) has argued self-emergent artificial systems could finally control the development of intelligence, and thus, human life.

As explained above, an essential driver of this development is Artificial Intelligence. Digital arte-facts, such as robots have increasingly become autonomous, allowing them to reproduce and evolve under their control. Key is their capability of self-awareness (cf. Amir et al., 2007). According to McCarthy (2004) it comprises

- knowledge about one's own permanent aspects or of one's relationships to others
- awareness of one's sensory experiences and their implications
- awareness of one's beliefs, desires, intentions, and goals
- knowledge about one's own knowledge or lack thereof; awareness of one's attitudes such as hopes, fears, regrets, and expectations
- the ability to perform mental actions such as forming or dropping an intention.

Some forms of self-awareness have been considered useful for digital arte-facts, in particular

- reasoning about what they are able to do and what not,
- reasoning about ways to achieve new knowledge and abilities
- represent, how they arrived at its current beliefs
- maintain a reflective view on current beliefs and use this knowledge to revise their beliefs in light of new information
- regard their entire 'mental' state up to the present as an object, and have the ability to transcend it and think about it.

The recognition of the social aspect of self-awareness (cf. Amir et al., 2004) seems to be crucial, as a digital arte-fact may have a system theory of itself that it can use to interact with others. Originally thought of a property to be used in multi-agent systems for dealing with errors in communication, argumentation, negotiation etc. it could be useful for reflecting on one's own development state, and articulating meaningful inputs to co-creative processes.

Amir et al. (2004) have considered several types of self-awareness:

- '*'Explicit Self-Awareness*. The computer system has a full-fledged self-model that represents knowledge about itself (e. g., its autobiography, current situation, activities, abilities, goals, knowledge, intentions, knowledge about others' knowledge of its knowledge) in a form that lends itself to use by its general reasoning system and can be communicated (possibly in some language) by a general reasoning system.
- *Self-Monitoring*. The computer system monitors, evaluates, and intervenes in its internal processes, in a purposive way. This does not presuppose that the monitored information lends itself to general reasoning; in fact, there may be no general reasoning (e.g., operating systems, insects).
- *Self-Explanation.* The agent can recount and justify its actions and inferences. In itself, this does not presuppose a full-fledged self-model, or integration of knowledge needed for self-explanation into a general reasoning capability." (ibid, p.1)

Cognitive architectures such as ACT-R, SOAR or CLARION take these properties into consideration. However, social cooperation does not always need the agent's understanding. There are forms of cooperation that are evolutionary self-organizing and unaware. Such cooperation forms do not require joint intentions, mutual awareness or shared plans among cooperating agents. In order to investigate intra-agent behavior of crowds regarding emotions we used more simplistic agents in our simulations (Barachini et al., 2022). Nevertheless, awareness of the self and other actors can be considered as one of the important factors of trans-human (system) developments. As explained, whether self-awareness (consciousness) can be artificially created or will automatically emerge remains open.

4. KM for Socio-Technical Capacity Building – An Opportune Advancement

We finally put the development of digital selves into a KM-relevant methodological context, applying Design Science-based research (Hevner, 2007, Peffers et al., 2007). This approach aims at building a respective knowledge base while targeting scientific rigor.

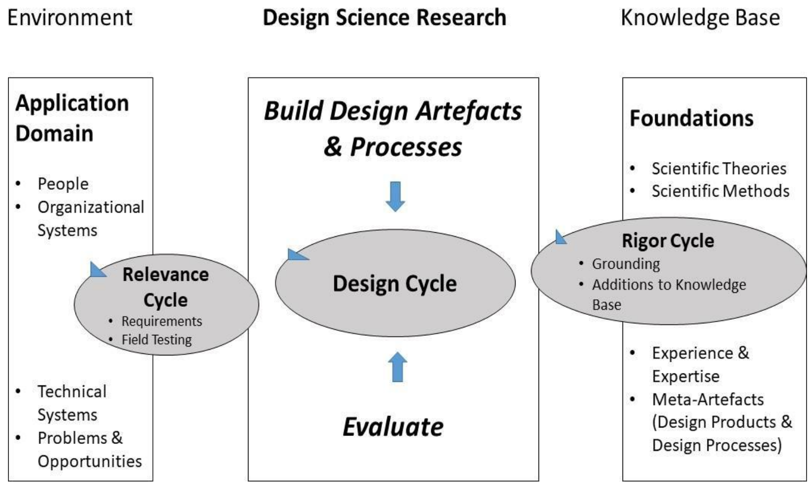


Fig. 4: Capacity Building through Design Science-based research (according to Hevner, 2007)

Design Science-based research is an established solution-oriented approach to system development that can be applied in various contexts and domains. It aims continuous and non-disruptive development of artefacts that include technology and other elements, such as ethical guidelines and deployment principles. Its dual nature addresses concrete designs as well as theories driving solution developments. System development and theory advancement play a crucial role in the 3-partite architecture shown in Figure 4. The Relevance Cycle connects the environment of a system development (project) with the core development activities. The Rigor Cycle relates these activities to a knowledge base informing the development project. The Design Cycle iterates between the core activities and targets building and evaluating artefacts.

Digital selves refer to the people dimension as stakeholders of different kinds are engaged in socio-technical system development, operation, and management. They can be researchers, digitalization developers, providers of system services, and producers of technologies. As already known from digital twin development they do not only act according to functional roles, but are part of some networked organization representing system components and services for operating a system. Digital selves can be composed of elementary components enabling trans-human behavior. These components could be sensors and actuators or twin elements adapting and synchronizing services for system operation. They can capture intelligent behavior resulting from prediction algorithms or be designed to perform tasks in a straightforward way. Consequently, the relevance cycle is driven on design issues involving different kind of agents or actors, targeting their interaction in some organizational or societal setting.

Design Cycles allow to focus on dedicated design options and digital-self components to meet specific requirements, e.g., checking environmental conditions to operate in a certain mode. The knowledge base is built along these design cycles. It captures not only requirements for system components and behavior, but also how they could or could not be met (cf. Documented Knowledge Base proposed by the Knowledge Life Cycle by Firestone et al., 2012). The knowledge is a timely and aligned collection of findings with respect to an artefact developed along design cycles. Concerning digital selves, it contains all 'generations' that evolve through Design-Science based research.

Design-Science-based development and its KM activities needs to be structured for compiling a knowledge base. A widely used framework has been proposed by Peffers et al. (2007). Its 6 core elements correspond to development that can be applied for the development of digital selves:

1. *Problem identification*: It comprises describe the specification of a specific research problem, such as enriching digital twins with socio-emotional behavior, and the justification of a solution providing some value for solving the research problem.
2. *Definition of Objectives of a Solution*: It includes analyzing the addressed problem, and deriving feasible requirements that can be argued for fitting to a solution. For instance, expressing socio-emotional behavior in transhumanist systems could be achieved by embodying algorithms that have been elaborated on in social sciences.
3. *Design and Development:* This phase is dedicated to create some artefact in a functional way. For instance, any digital-self artefact should be able to be synchronized in real-time with its physical environment the way digital threads operate. The artefact itself can be considered a socio-technical instrument for accomplishing domain-specific tasks.
4. *Demonstration:* When executing this step, the created artefact is put into its context of use. It should be utilized to solve one or more instances of the addressed problem. This task can be achieved either through experimental settings, use cases, or simulations. For digital selves the demonstration should be enabled by technological artefacts as they allow interactive validation at runtime.
5. *Evaluation*: In this step designers check whether the artefact supports a solution of the problem. Depending on the problem and the artifact, different evaluation methods, including surveys and feedback can be used. Once the data from the evaluation are available it needs to be decided whether another Design and Development cycle has to be performed, or the requirements could be met or have to be adapted for another design cycle.
6. *Communication*: The result of each design cycle can be stored and contextualized in the knowledge base. It allows developers and designers not involved to trace each step of evolving solutions.

The operational framework can be adopted for various instances of digital selves, and underpinned by various theories and concepts, as provided by transhumanists or complex adaptive system thinkers. KM activities need to be adopted to knowledge generation and sharing capabilities of involved actors. KM techniques need to be applied for encoding not only results, but also meta-information on how the results stored in the knowledge base have been achieved. The consideration of that meta-knowledge will become crucial in trans-human settings, since the origin matters when decisions need to be grounded on specific type of actors or events.

5. KM in a Trans-humanist World – Dystopian Thoughts

Kahneman (Kahneman, 2011) describes two different ways the human brain forms thoughts. System 1 is the unconscious system which is active most of the time. System 2 1 is the conscious system which needs much more attention and energy when activated. System 1 represents facts and "learned" mental models.

Such models can be digitally presented as facts and rules of a knowledge base. System 2 is activated in our brains when deeper reflections are needed. The digital counterparts for System 2 are cascaded neural networks (deep learning) or simulations.

Since Knowledge management (KM) is the collection of methods relating to creating, sharing, using and managing the information of an organization we think that all these methods will also be applicable to digital selves. Thus, single-, double-, and deutero-loop learning in the sense of Argyris and Schön (Argyris, 2000; Argyris & Schön, 1996) will be possible also for digital selves.

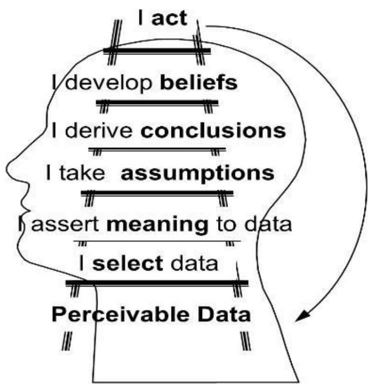


Fig. 5: The Perception Ladder

Like humans, digital selves (agents) will apply the loop represented in Figure 5. Based on presented data they select data, they assert meaning to data and take assumptions, they derive conclusions and consequently develop beliefs. Based on those beliefs, which represent mental models they act and select new data. The digital System 1 and System 2 are permanently active during these cycles. It is therefore extremely important which kind of data is presented to digital agents.

By applying simulation techniques based on game theory different modeling techniques and different algorithms can yield deviating results for one and the same property (Barachini et al., 2020, 2022). Thus, the phrase "what you see, you will be" is valid for humans as well as for agents.

It is therefore in our hands whether the future of artificial humans (humanoids) will be bright or rather dark in the Leviathan's sense. This reminds us on the old Indian allegory where a father tells his child that there are 2 animals fighting in his breast. One animal is full of negative emotions like hate and jealousy, the other animal is obliging and friendly. On the child's question which one is going to win the fight the father's answer is "the animal which you feed is going to win".

Consequently, the data and the world which we are presenting to the programs and the algorithms we are implementing do have impact to the behavior of artificial agents. The same relation holds for humans – experience, education and epigenetic on one hand side, and inherited properties on the other shapes our behavior. But in contrast to rational computer programs, humans often act in an irrational way, since they are driven by emotions as well.

Very relevant seems to be the question how humans can coexist with digital selves. Today humans understand the underlying computing methods but in some cases they cannot explain why AI programs already produce better results in certain domains. In the farer future the interpretation of big data might equip digital selves with an intelligence which deviates increasingly from human intelligence.

This is due to the fact that a digital System 2 is capable to access and interpret more and more data whereas the human brain capacity is limited. In its last consequence humans might neither understand the underlying methods nor the results of the actions performed by digital selves. This poses some risks for humankind.

Imagine a combination of fully autonomous cars and conventional cars in our traffic system. Autonomous cars need to understand that a human driver could violate certain traffic rules due to emotional overload. Thus, for a peaceful and frictionless coexistence it seems to be crucial that digital selves can “understand” human behavior and reactions. Therefore, digital selves need to “understand” ethics, morale and emotions of human beings through model appropriation. It seems to be crucial that human-centric AI is implemented in our human-technology-organization framework. Thus, the incarnation of human-like behaviors as digital selves seems to be mandatory. Vice versa it will be

important that digital selves can explain their actions and decisions to humans. This is an even bigger challenge since digital selves might develop complete other personalities which might enslave humans.

Anyhow, digital selves need electricity like humans need food. In its last consequence humans can unplug the socket and thus gain full control again. But what if the transformation of humans to digital substrates will already be too far elaborated at that time? If such a tipping point is exceeded it may cause serious control problems.

Whether a complete digital transformation of humans is achievable in due time depends on the survival of humankind and on technologies yet to be invented. Some dystopian thoughts might be in vain if global warming, starvation, a pandemic crisis, atomic wars, an extraterrestrial impact or simply human stupidity cannot be stopped. In this case our species will be extinguished before it can transform to computational substrates.

However, as suggested, to start building a knowledge base with scientific rigor helps structuring and sharing experiences that are collected from and for digital-self developments. We suggest grounding them on design science-based research integrating knowledge management capabilities as central element of socio-technical progress.

References

Ackerman, M., Goggins, S., Herrmann, T., Prilla, M., Stary, C. (2018). *Designing healthcare that works: a sociotechnical approach*. Academic Press/Elsevier, London.

Amir, E., Anderson, M. L., Chaudhri, V. K. (2007). *Report on DARPA Workshop on Self Aware Computer Systems*. SRI International Menlo Park United States.

Argyris C. (2000). Double-Loop Learning. In: Wiley Encyclopedia of Management; Hoboken: New Jersey, pp. 115-125.

Argyris, C., Schön, D. A. (1996). D.A. Organizational Learning II: Theory, Method and Practice, Addison-Wesley: Reading, MA, 1996.

Baar, Bernard J. (1988). *A Cognitive Theory of Consciousness*. Cambridge, MA: Cambridge University Press.

Barachini F. (1991), The Evolution of PAMELA. *Expert Systems the International Journal of Knowledge Engineering*, Vol.8, No.2.

Barachini F., Bornemann M. (2020). Jealousy in Cooperation – A Comparison of Two Game-Based Approaches, *Journal of Information and Knowledge Management*, Vol. Nr.19., Nr. 2, 2020.

Barachini, F., Stary, C. (2022). From Digital Twins to Digital Selves and Beyond. Engineering and Social Models for a Trans-humanist World, Springer, Cham, https://doi.org/10.1007/978-3-030-96412-2.

Bostrom, N. (2009), The future of humanity. In: New waves in philosophy of technology; Eds. Olsen, J.-K. B.; Selinger, E.; Riis, S.; Palgrave McMillan: New York, 186-216.

Bostrom, N. (2014), *Superintelligence: Paths, dangers, strategies*. Oxford University Press: Oxford.

Dehaene S., Changeux J.P. (2011). Experimental and theoretical approaches to conscious processing, *Neuron* 70(2) :22-227.

Feigenbaum, E. (1981, 1982), *The Handbook of Artificial Intelligence*, Vol. 1,2,3, William Kaufmann.

Firestone, J. M., & McElroy, M. W. (2012). *Key issues in the new knowledge management*. Routledge.

Grieves, M. W. (2019). Virtually intelligent product systems: digital and physical twins. In: *Complex Systems Engineering: Theory and Practice*, Ch. 7,eds: S. Flumerfelt, K.G.Schwartz, D. Mavris, S. Briceno, American Institute of Aeronautics and Astronautics, 175-200.

Grieves, M. W. (2014). Digital Twin: Manufacturing Excellence through Virtual Factory Replication, white paper, https://www.researchgate.net.

Hebb, D.O. (1949), *Organization of Behavior*, Psychology Press Edition.

Hevner, A. (2007) A Three Cycle View of Design Science Research. *Scand. J. Inf. Syst*. 19(2), 87–92.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan, New York.

Kandel, E. (2018). *Das Zeitalter der Erkenntnis*, Siedler Verlag.

Kohonen, T. (1982). Self-Organized Formation of Topologically Correct Feature Maps. *Biological Cybernetics*. 43 (1): 59–69.

Kohonen, T. (1984). Self-Organization and Associative Memory, Springer, Berlin.

Kohonen, T., Mäkisara, K., Simula, O., Kangas, J. (1991). *Artificial Neural Networks*, North-Holland, Amsterdam.

Kurzweil, R. (2012). *How to Create a Mind*, Viking.

Laird, J.E, Newell, A., Rosenbloom, P.S (1987), SOAR: An architecture for general intelligence, *Artificial Intelligence*, 33(3), 1-64.

Laird, J.E. (2012), *The SOAR Cognitive Architecture*, MIT Press.

Libet, B. (2005). *Mind Time: The Temporal Factor in Consciousness,* Harvard University Press.

McCarthy, J. (2004). Notes on Self-Awareness. University of Stanford, Research Report.

More, M. (1990). Transhumanism: towards a futurist philosophy. *Extropy*, 6, 6-12.

Newell, A., and Simon, H. A. (1972). *Human problem solving*, Englewood Cliffs, NJ: Prentice-Hall.

Oppl, S., Stary, C. (2019). *Designing Digital Work: Concepts and Methods for Human-centered Digitization*. Springer Nature, Cham.

Peffers, K., Tuunanen, T., Rothenberger, M. A., Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research, *Journal of Management Information Systems,* 24(3), 45–77. https://doi.org/10.2753/MIS0742-1222240302

Perno, M., Hvam, L., Haug, A. (2022). Implementation of digital twins in the process industry: A systematic literature review of enablers and barriers. *Computers in Industry*, *134*, 103558.

Rosenblatt, F. (1958). The Perceptron. A probabilistic model for information storage and organization in the brain, *Psychological Reviews* (65), 386-408.

Searle, J. (1980), Minds, Brains and Program. *Behavioral and Brain Sciences*.

Stary, C. (2021). Digital twin generation: Re-conceptualizing agent systems for behavior-centered cyber-physical system development. *Sensors*, *21*(4), 1096.

Turing, A. M. (1950). *Mind, Computing Machinery and Intelligence*, LIX (236): 433-460, Oxford Academic Press.

Weizenbaum, J. (1976). *Computer Power snd Human Reason*, W.H. Freeman and Company.

Zhuang, C. B., Liu, J. H., Xiong, H., Ding, X., Liu, S., Weng, G. (2017). Connotation, architecture and trends of product digital twin, *Comput. Integr. Manuf. Syst*, *23*(4), 753-768.

Research Teams: Fostering Scholarship and Practice

Jeff Allen
University of North Texas, USA
Jeff.Allen@unt.edu

Malak Khader
University of North Texas, USA
MalakKhader@my.unt.edu

Millicent Njeri
University of North Texas, USA
Millicent.Njeri@unt.edu

Amy Rosellini
University of North Texas, USA
amy.rosellini@gmail.com

ABSTRACT

This workshop is presented by members of a University of North Texas research team. First, the team will overview their experience as members of the research team and share experience in areas such as trust formation, team roles, productivity, work-life balance, faculty-students interaction, peer and faculty mentorship, dissertation preparation, and job seeking. Second, the workshop will discuss and brainstorm how this format can be implemented for organizations both with faculty-student teams and with peer-directed teams. Finally, successes and challenges are openly discussed with audience.

KEYWORDS

faculty team; research methods; student team; program development, scholarship; pedagogy

GOALS

This panel session is presented by members of a University of North Texas research team (faculty, students, and alumni). First, the panel will overview their experience as members of the research team and share experience in areas such as trust formation, team roles, productivity, work-life balance, faculty-students interaction, peer and faculty mentorship, dissertation preparation, and job seeking. Second, the panel will discuss how this concept can be implemented for organizations both with faculty-student teams and with peer-directed teams. Finally, the thorns and roses are openly discussed with audience.

OVERVIEW

The *SageResearch Tribe* was formed in the Department of Information Science at the University of North Texas in Fall 2018 with an open call for participation to current doctoral students. Criteria for membership was bounded by four criteria: 1) general interest in information literacy/behavior 2) willingness to learn from others; 3) willingness to work within a team of talented scholars; 4) ability to meet deadlines; and 5) interest in producing content for distribution.

Beyond the four criteria that serve as a price of admission on the team, the *SageResearch Tribe* builds resilience in its members through navigating open and direct feedback with poster, paper, and article submissions. The team fosters interdisciplinary education through collaboration of varied research interests to find intersections of skills that together contribute to the information science body of research.

This team is an example of how LIS education can modify the instructional environment with unique collaborations between faculty, students, and alumni. Doctoral research teams change faculty-student dynamics, depart from standard curriculum practices, and innovate how we prepare resilient information scholars with practical skills needed for the future.

The *SageResearch Tribe* builds leadership skills through both team leadership appointments and leaders of individual proposals. Members build confidence by learning to accept or reject feedback from peers and established academic reviewers. Graduate students forming the initial team, in 2018, had a variety of backgrounds and the initial five students were a variety of stages within their doctoral program (coursework to dissertation defense). Students' research interests included culture and information literacy, training and development, veterans and the workforce, smart cities, and information behavior in social media with the faculty director studying workforce wisdom. During the first year, the tribe worked to develop trust, negotiated common research interest, determined venues for scholarly dissemination, and created a first annual workplan to be conducted over the next 12 months.

As exciting as the first year was, the second (2019) year brought dynamic changes in team members. The team worked to integrate new students while gradually phasing out others - relationships and trust are key. The team also began building one of its strongest assets – grit in the form of receiving constructive feedback from peers, accepting and implementing critiques of rejected peer-reviewed papers, and navigating difficult conversations between members to brainstorm and develop important and innovative research. With the development of new skills, the second year exhibited team productivity growing with papers and posters accepted at major international conferences and a chapter submission.

The third year (2020) continued the trend of change with three members graduating and a new member joining. However, in this instance the three graduating members shifted their roles on the team (as alumni) but continued a full commitment to team goals. With fewer members exiting the team, productivity continued with papers and posters being accepted at major international conferences and a published book chapter. Though dynamics and membership continue to change, the team was normalized and poised for a fantastic fourth year.

The 2021 team is composed of the four doctoral students, three alumnae, and a faculty mentor. The team's annual schedule now includes 12 papers, panels, or papers for national and international conferences, four journal articles in development, two under reviews, and two in press. At this time, four previous team members have completed their Ph.Ds.

SESSION STRUCTURE

There are tremendous lessons to learn from a tribe of scholars with common goals. We propose an innovative and interactive workshop format.

Part 1: Story of development, normalization, and performance. Part 1 will provide lessons learned, thorns and roses that are part of being a member or a collaborative high performance knowledge team.

Part II: Implementation. Part II will provide an opportunity for participants to create a plan for implementation in their organization (corporate or university).

WORKSHOP LEADERS

Dr. Jeff M. Allen is an internationally recognized scholar in the area of workforce innovation for the knowledge economy. He serves as a Regents Professor of Information Science at the University of North Texas. Together with his colleagues, he prepares students for jobs that are not yet created.

Malak Khader is a Ph.D. Candidate at the University of North Texas, where her field of study is broadly Information literacy, in collaboration with cultural and religious literacies. She has also done research with Information Behavior, Information Seeking, and Data Ethics.

Millicent (Milly) Njeri is a currently pursuing her Ph.D., at the University of North Texas, in Information Science with a concentration in Consumer Behavior and Experience Management and minor in Research, Measurement and Statistics. Her research interests include smart tourism, data analytics, consumer experience and behavior management, and tourism destination management.

Dr. Amy Rosellini is a knowledge management strategist currently working as Chief People Officer for a national investment real estate group. She received her PhD in Information Science with the University of North Texas. Her research focuses primarily on knowledge management systems, corporate culture and knowledge measurement.

Track – Knowledge for Development Partnership

Localizing knowledge management with country knowledge programming in the Asia and Pacific Region

Authors:
Mary Jane Carangal-San Jose,
Vivek Raman
and Susann Roth

Knowledge management processes
Country knowledge programming
Participatory knowledge management
Change management
Knowledge management, country knowledge plans

Goal: This paper seeks to identify factors that contribute to improving the way the Asian Development Bank (ADB) understands and responds to the knowledge needs of its developing member countries (DMCs). The COVID-19 pandemic and continuing uncertainties highlighted the importance of knowledge management in helping governments in DMCs make well-informed decisions to respond to the crisis and work for recovery. As ADB's DMCs are at various stages of development, it is essential to identify, differentiate and manage knowledge that can advance these countries to achieve development objectives and part of that the Sustainable Development Goals (SDGs). ADB's Strategy 2030 aims to strengthen its country-focused approach using the Country Partnership Strategy (CPS) as the main platform to provide lending and knowledge to its DMCs. An important part of the CPS is the country knowledge plan (CKP) that outlines how ADB is supporting the country's knowledge needs over the CPS period. This paper discusses the significant enhancements introduced to make the CKP process 'dynamic', client-driven while supporting national knowledge agendas. From ADB's perspective, knowledge is an investment and an asset, and it is important to ensure that its DMCs are getting the most return from this capital.

Design / Methodology / Approach: A participatory approach was employed in determining improvements to the country knowledge programming. Consultations were held with various resident missions to understand better the challenges of the current process and point areas for improvement. A common pain point is that in the previous CKP, the link between upstream knowledge, national knowledge agendas, and future lending is not clear. The goal is to ensure that ADB's knowledge portfolio is balanced between government demand and the bank's proposed knowledge support, as well as aligned with ADB's strategic objectives for that country. This requires improved collaboration within ADB, and with the Government. The dynamic CKP process was piloted and co-developed with country directors that lead ADB's operations in countries such as the Philippines, Viet Nam and Indonesia. To ascertain clients' needs, ADB tested various instruments and approaches including quick survey of governments' knowledge demand, discussing clients' requirements during country programming missions for lending, or separate knowledge consultations with government agencies who are involved in implementing ADB's projects in that country. To inform what ADB is supplying these countries in terms of knowledge, a rigorous quantitative analysis of the bank's knowledge portfolio was developed and validated across the bank.

Results: The results of the CKP pilots in 3 developing member countries indicated that there is scope to reallocate resources for knowledge to ensure these are relevant to the government's national development priorities as well as with ADB' strategic objectives for that country. This can only be surfaced through collaboration within ADB and in consultation with the clients. The end goal of the process is to better understand how ADB's knowledge is helping its DMCs solve development challenges along key knowledge result categories of (i) increasing awareness and use of evidence-based information for policy making; (ii) improving program/project delivery and, (iii) enhancing capacities and skills among policy makers.

Limitations of the research (if applicable): The dynamic CKP process can inform where ADB can improve its process of programming for a country's knowledge needs. It does not yet guide where ADB should identify its niche or in which content or focus areas, knowledge resources should be allocated. The process does not make any judgement on ADB's knowledge portfolio for its DMCs, but instead, suggests how to structure that knowledge supply to better respond to the country's requirements.

Practical implications: The most appropriate way to understand what the clients need, is to ask the clients themselves. Hence, a consultative process for knowledge was added to the way in which country knowledge programs are prepared. The exercise also emphasized the need to collaborate and explore integrated approaches across ADB sectors and themes to develop knowledge solutions for countries to avoid redundant outputs.

Social implications: This initiative aligns with ADB's current efforts to transform the bank's culture along the lines of being client- centric and trustworthy. The selection of the most appropriate interface to understand and respond to clients' needs will guide how well ADB can package its knowledge offerings and interventions and move from being a financial product provider to a knowledge solution provider.

Originality / value: ADB's total knowledge portfolio based on technical assistance it provides to countries such as the Philippines, Viet Nam and Indonesia is almost $300 million. The most important questions to ask ADB's resident missions in these countries are that with this investment, are we doing the right things and are we doing things right. Knowledge is an investment that has benefits that can spill across many years and across different areas. Understanding where clients need knowledge the most will help guide ADB's resource allocation for more development impact.

Track – Deutsche Gesellschaft für Information und Wissen

Intercultural perspectives on Information Literacy - a report on a transnational project to foster information literacy

Joachim Griesbaum,
Stiftung Universität Hildesheim, Institut für Informationswissenschaft und Sprachtechnologie, Germany

Semantic networks and knowledge management – context does the trick

Johannes Munk,
Klarso GmbH, Germany

In the end, it's about people – knowledge organization at Lilly Pharma

Ingrid Weber,
Lilly Deutschland GmbH, Germany

The DikoLa Project - Examples of innovative formats of knowledge transfer in teacher education

Katharina Heider,
Martin-Luther-Universität Halle-Wittenberg, Zentrum für Lehrer*innenbildung, Germany

Presentations will be made available for download.

Workshop I

Workgroup KM Certification (ISO 30401/DIN SPEC 91443)

Facilitated by

Kay Alwert,
alwert GmbH & Co. KG, Germany

Manfred Bornemann,
Intangible Assets Consulting GmbH, Austria

Holger Kohl,
Fraunhofer Institute Fraunhofer-Institut für Produktionsanlagen und Konstruktionstechnik IPK, Germany

Ronald Orth,
Fraunhofer Institute Fraunhofer-Institut für Produktionsanlagen und Konstruktionstechnik IPK, Germany

Markus Will,
Fraunhofer Institute Fraunhofer-Institut für Produktionsanlagen und Konstruktionstechnik IPK, Germany

New standards (ISO 30401 and DIN SPEC 91443) provide new momentum for knowledge management initiatives. Certification becomes a priority to demonstrate KM-proficiency on an organizational level to external partners (customers, suppliers, future employees) as well as to create a sense of urgency within organizations, in particular among organizational leadership.

In this workshop, a joint working group of two independent communities and a research organisation, the Association of Knowledge Management for German speaking countries and the Federal Association of Intellectual Capital Statements e.V. presents the framework of DIN SPEC 91443 which was designed specifically for SMEs to operationalize the KM requirements of ISO 30401.

The general objective of the workshop is to reflect on suitability of various instruments of KM, in particular the Intellectual Capital Statement – Made in Germany, to deliver on the standards requirements. Looking at some practical cases, an early-stage concept of auditing structures and procedures will be discussed. As a result of the workshop, the organizers aim to propose an auditing framework to certify knowledge management systems as a follow up.

Poster Session

Knowledge Management System Innovation: From Global Needs over Decentralized Individual Devices and Affordances towards a Shared Transdisciplinary Knowledge Heritage Repository

Ulrich Schmitt,
University of Stellenbosch Business School, South Africa / Botswana

Health disparities across native American communities: a knowledge sharing perspective

Yong-Mi Kim,
University of Oklahoma, United States of America

David Steinmann,
Access Family Care, United States of America

Don McBride,
Access Family Care, United States of America

Consteladev: an innovative platform based on systemic methodologies to support teams and project management

Simone Perazzoli,
Constela Digital Company, Brazil

Wagner Sousa Santos,
Federation of Industries of the Santa Catarina - School System, Brazil

José Pedro de Santana Neto,
University of Geneva, Switzerland

Uncovering Disruption of knowledge dissemination in the context of global scholarly communication - An analysis of scientific Editorials using an Text Mining approach

Anna Cunningham,
Technische Universität Dresden, Germany

Alina Praun,
Technische Universität Dresden, Germany

Martina Pieperhoff,
Technische Universität Dresden, Germany

Library Collection Development and Branding: An Exploration of Ranganthan's Fifth Law That "A Library Is a Growing Organism"

Patrick Roughen,
North Carolina Central University, United States of America

Elicitation of tacit knowledge inside a clinical process using the FRAM

Carmen Aschwer,
FOM University of Applied Sciences, Aachen, Germany

Thomas Mühlbradt,
FOM University of Applied Sciences, Aachen, Germany

Helga Unger,
FOM University of Applied Sciences, Aachen, Germany

Knowledge Surfacing - Expert created knowledge connections

Bart Verheijen,
Guruscan, The Netherlands

Track – KM Practitioner

BOSCH Experts Organization: Experiences & Perspectives

Juergen Ebmeyer,
Bosch Experts Organization (BEO), Germany

Lothar Maier,
Bosch Experts Organization (BEO), Germany

The Bosch Group is a leading global supplier of technology and services. It employs roughly 402,600 associates worldwide (as of December 31, 2021).
The company generated sales of 78.7 billion euros in 2021. Its operations are divided into four business sectors: Mobility Solutions, Industrial Technology, Consumer Goods, and Energy and Building Technology. At 128 locations across the globe, Bosch employs some 76,100 associates in research and development, of which more than 38,000 are software engineers.

Juergen Ebmeyer joined Bosch in 2005 and is the Corporate Process Owner of the Bosch Experts Organization. He managed change projects to create divisional Centers of Competence and coordinates the Bosch-wide Centers of Competence.

Lothar Maier works for Bosch since 2005 and is the IT Infrastructure Application Owner of the Bosch Experts Organization. He additionally supports the Bosch organization with further KM tools and methods as e.g. Expert Debriefing moderation.

Fit to collaborate in the Otto Group: Experiences and Perspective

Juliane Dieckmann,
Otto Group, Germany

Juliane Dieckmann, Head of Knowledge Management in the Otto Group Holding, is responsible in the Digital & Consulting division for the group-wide transfer of knowledge and the development of KM products specifically aimed at collaboration in organisational units. She focuses with continuing enthusiasm on the task of making Knowledge Management in the Otto Group fit for the path to digital transformation and putting people at the centre of this.

The Otto Group is a globally active trading and services group with around 52,000 employees and a turnover of 15.6 billion euros. With 30 major corporate groups it is present in more than 30 countries. Global group activities and strategic partnerships provide the Otto Group with excellent conditions for know-how transfer and the use of synergy potential. Every day, in the course of its business activities, it deals with the core issues of a modern society, including the future of work, value-oriented action, and the opportunities and challenges of the present.

Knowledge Management enables the transformation from a global steel supplier to a technology service provider

Hubert Fratzl,
voestalpine HPM Division, Austria

Sven Plenkers,
voestalpine HPM Division, Germany

In 2018, voestalpine High Performance Metals decided to significantly invest in Knowledge Management to foster their strategy to expand into new markets.

After conducting a thorough global feasibility study involving different stakeholders at multiple locations, a hybrid KM strategy was proposed creating networks of experts and identifying valuable content for re-use supported by a KM Service Portfolio. Members of pilot “communities of Practice” (CoP) were involved in the co-design of KM Services and testing of IT platforms. Once the global IT platform was established, the global roll-out of the CoP’s started.

Simultaneously, a global governance structure was designed and implemented as well as a 3-tier measurement approach conceptualized and tested. Today, CoP’s focusing on different topics, are conducting dozens of meetings every month and have identified about a thousand documents to be accessed via their electronic “home base” on Sharepoint. A training program to raise awareness and enable understanding of KM has been initiated on a global scale. The implementation of the KM board brought an important boost in participation and professionalism. To further foster and sustain the world-wide participation in the knowledge-network is the challenge we are facing until knowledge management has become an integral undisputed part of the corporate culture of voestalpine.

Conversation about KM at NASA

Ed Rogers,
former CKO at NASA, USA

Ed Rogers served as Chief Knowledge Officer with NASA Goddard Space Flight Center from 2003 to 2020. He holds a PhD in HRM from Cornell University and teaches a different universities like University of Alabama, George Washington University and Indian School of Business. Since 2020 he is the principal owner of Mayjer Enterprises (Palmyra, Virginia).

Intellectual capital management in an engineering company

Christian Mühlbauer,
Zilch + Müller Ingenieure, Germany

Engineering offices are knowledge companies. Systematically maintaining and developing the existing knowledge in and as a company is therefore an essential task of corporate management. Knowledge management at ZM-I includes a training system (ZM-I Campus), a digital knowledge database (intranet) and workshops. For a systematic analysis and evaluation of the company's intellectual capital, a bi-annual certified intellectual capital report is carried out. For this purpose, employees from all divisions and locations come together to systematically analyse and evaluate the intellectual capital in the company. The intellectual capital statement is thus a fundamental management tool at ZM-I.
ZM-I is an independent, medium-sized civil engineering company organised as a partnership for demanding projects. Based in Munich, the group has grown continuously and is now active throughout Germany. As a structural engineering company with a focus on statics, ZM-I offers consulting, planning, testing and assessment services in the field of civil engineering.

Dr.-Ing. Christian Mühlbauer is managing partner at ZM-I in Munich. He has a doctorate in civil engineering and specialises in the field of structural engineering and building in existing structures and is part of the ZM-I intellectual capital report team.

KM Practices - Motivational Structures within Tax and Legal Knowledge Management

Katharina Otto,
PricewaterhouseCoopers, Switzerland

This presentation deals with motivational ideas and initiatives for knowledge sharing in the field of tax and legal consulting. Sharing knowledge at all levels and in all formats is often difficult or even impossible due to the need to respect confidentiality and data protection laws. However, the example of PwC Switzerland shows how it is still possible to be a knowledge organization, to create a lively and popular exchange between employees and not to reinvent the wheel every day. We will also shed light on what a well-positioned knowledge management can enable in a crisis situation and how it has moved from behind the scenes to the stage.

PwC Switzerland is the leading audit and advisory company in Switzerland. We’re a network of firms in 156 countries with over 295,000 people who are committed to delivering quality in assurance, advisory and tax services. PwC Switzerland has over 3,380 employees and partners in 14 locations in Switzerland and one in the Principality of Liechtenstein.

Track – Knowledge Management & Learning

Value exchange exploration supporting project-based learning

INIGA ANTONIA BAUM *
Institute of Business Informatics – Communications Engineering,
JKU Knowledge Management Competence Center,
Johannes Kepler University, Linz, Austria
E-mail: i.antonia.baum@gmail.com

CHRISTIAN STARY *
Institute of Business Informatics – Communications Engineering,
JKU Knowledge Management Competence Center,
Johannes Kepler University, Linz, Austria
E-mail: christian.stary@jku.at

SUMMARY

Goal: *Collect and harvest knowledge for enhancing project-based learning (PjBL) from the perspective of involved stakeholders to design user-centred digital learning support.*

Design / Methodology / Approach: *Design Science Research / Value Network Analysis*

Results: *Identification of relevant roles and value exchanges in PjBL processes; design requirements for digital learning support; Consolidated stakeholder interaction for learning management recognizing the need for (informal) knowledge sharing and learning loops.*

Limitations of the research: *Macro-level results require refinement for operational implementations on the micro-level, in particular for digital learning support.*

Practical implications: *Facilitators and learners can utilize explicated knowledge in a context- sensitive way; developers and researchers can advance building PjBL artefacts.*

Social implications: *Knowledge explication appreciating individual mental models is enabled; Stakeholder-driven consolidation for organizational development is facilitated.*

Originality / value: *Implicit knowledge with respect to PjBL could be externalized for (further) developing stakeholder-centred support in digital contexts.*

Abstract

Project-based learning (PjBL) allows learners to focus on a topic of their interest and apply the presented and acquired knowledge in their current (work) practice. When applied in institutional contexts, these particularities have to be recognized for stakeholder-centered learning support. In this paper we report on identifying and designing value exchanges between concerned stakeholder groups by methodological deployment of the Value Network Analysis. The individual and collective perception of PjBL experiences could be externalized and design inputs of involved and addressed stakeholder groups could be developed in a non-disruptive way. The results give way for increasing self-efficacy of knowledge acquisition and skill development in (digital) learning environments.

Keywords: Project-Based Learning (PjBL). Stakeholder-centered design. Value Network Analysis (VNA). Knowledge generation. Knowledge sharing and co-creation. (Digital) learning management.

INTRODUCTION

Project-based learning (PjBL) has been identified key to develop 21st century skills [cf. Chu et al. 2012]. It is "a systematic teaching and learning method, which engages students in complex real-life tasks that result in a product or presentation to an audience, enabling them to acquire knowledge and life-enhancing skills." [Barron and Darling-Hammond 2008, Thomas et al. 1999, in Chen and Yang 2019] While conventional instruction and learning approaches have increasingly difficulties to succeed in general and especially in remote didactic settings [Adedoyin and Soykan, 2020; Meskhi, 2019; Weichhart et al. 2018; Sumarni, 2013; Ngereja, 2010], PjBL yields a major benefit in generating practical competencies and expertise - the prerequisite of so called "21st century skills" [Abuhmaid, 2020; Aldabbus, 2019; Chen and Yang, 2019; Kakalejčík and Paľová, 2019] - on a high level: "Meeting the principles of learning effectiveness (…) will require significant change from established practice in the majority of educational settings available for young people in most of our systems. (…) naturally international in scope." [OECD, 2018] At the same time, *Self-efficacy* as vital component for scholarly and professional success is provably fostered by deployment of PjBL in academic technological contexts [Suastra and Ristiati, 2019; Amini, 2019; Mannila et al. 2018; Bilgin, 2015]. Therefore it becomes more and more important for institutions of higher education to consider it in increasingly enforced digital contexts.
As PjBL's central activities involve the transformation and construction of new knowledge, it is closely related to constructivist learning theories [Dewey, 1938; Piaget, 1929, 1970, 1971;
Ütanir, 2012; Oguz-Univer and Arabaciouglu, 2014; Thomas, 2000 in Chen and Yang, 2019; Kakalejčík and Paľová, 2019], justifying to investigate PjBL from a *Knowledge Management (KM)* perspective. In this work we address PjBL stakeholders' self-efficacy along learning processes, in particular along the exchange of values in academic education. It concerns both, maturing in terms of organizing learning processes as well as competence development in the concerned learning subject. For eliciting, representing, and sharing of knowledge we utilize *Value Network Analysis (VNA)* [Allee, 2009]. This technique also supports eliciting implicit knowledge that can be deployed for scaffolding and supporting learner-centered PjBL processes taking into account organizational context [Stary, 2014]. The latter requires the specification of roles and content framing didactically grounded interventions (cf. [Weichhart et al. 2018]). The presented exploration therefore focusses on the following research questions; (i) *Which role understanding do stakeholders have when being involved in PjBL*?, and (ii) *Which value exchanges as perceived by role carriers contribute to the effectiveness of PjBL?*

The findings lay ground to develop features promoting self-efficacy for stakeholder-centred (digital) PjBL support systems. Involving two academic institutions with Business Informatics programs having implemented PjBL differently, the individual VNA-interviews could capture particularities, distinct benefits, and risks as perceived by the stakeholders, both on the facilitators' and on learners' level. Business Informatics programs provide an ideal context for investigating PjBL, as the high demand reveals for skilled workforce in that field [Candrlic, et.al. 2020; Geng, 2019; Meskhi, 2019; Mannila et al. 2018; Weichart et al. 2018]. We assume, that an insight into the specific perception of tangible and intangible value exchanges will help co-creating a common understanding of value-based learning management and learner support. The paper starts with sketching the addressed concepts, *PjBL*, *self-efficacy*, and *Complex Adaptive Systems (CAS)* framing the learning and thus, KM activities. The methodology section introduces the applied *Design Science-based Research (DSR)* approach for the deployment of the *VNA*. Subsequently, the field study is detailed and the results are presented. Finally, we discuss the answers to our research questions and consolidate the findings for practical use and research.

CONCEPTUAL UNDERPINNING

In the following we sketch the main concepts addressed in the study, namely Project-based Learning (PjBL), Self-Efficacy, and Complex Adaptive Systems (CAS). Their direct encounter enables considering a focused organization of knowledge construction (i.e. PjBL) while both mostly concerned stakeholder groups (i.e. project-based learners and facilitators) are active participants in a system of complex relations, meaning, they affect themselves and each other through their behaviour – see Figure 1.

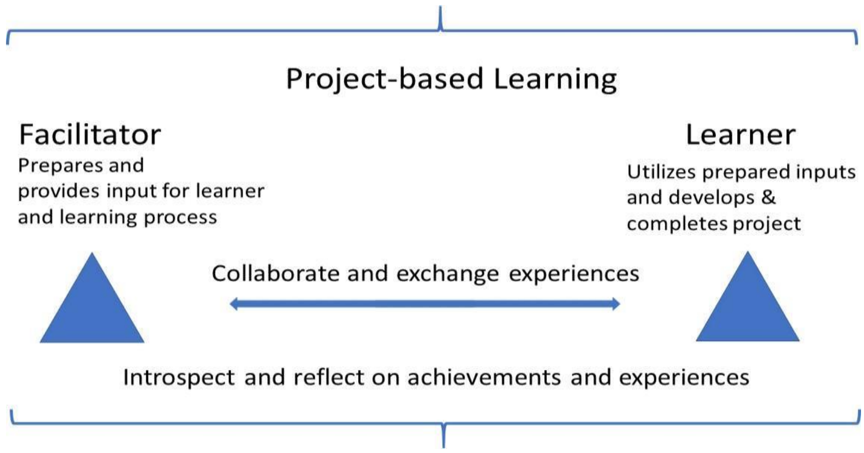


Fig. 1: Facilitator and learner cooperate in Project-based Learning settings (by authors)

PROJECT-BASED LEARNING AND SELF-EFFICACY

As a general concept self-efficacy is known since Bandura in the 1970ies to "define a person's opinion to his/her competency to achieve expected performance levels" having "effective, cognitive and motivational" repercussions. [Bandura, 1977] 40 years later Mannila et al. still identify self-efficacy crucial to problem solving endeavours in computing and self-regulated learning, when referring to the necessity to foster skills required to teach technological content [Mannila et al. 2018]. Recently, Shin relates self-efficacy as motor for successful achievement in the context of individual learning to students' participation in project-based learning. [Shin, 2018] While various studies in different fields indicate the need for (future) research on the effects of PjBL on students' motivation in learning and on self-regulated learning [Mahasneh and Alwan, 2018], KM work has acknowledged that there is a "perceived influence of factors that are beyond the teacher's control", with most research focusing exclusively on learning or teaching processes, neglecting the organisational role and environment involved [Gibson and Dembo, 1984; Woolfolk et al. 1990; Tschannen-Moran et al. 1998, in: Mahasneh and Alwan, 2018]. The presented study takes into account this context, when it aims to leverage answers concerning the actors' perspectives on the effectiveness of experienced PjBL processes while examining the organisational point of view on learning efficacy.

COMPLEX ADAPTIVE SYSTEMS

Since PjBL stakeholders are part of the systems they investigate and change, Complex Adaptive Systems (CAS) theory helps structuring understanding the current situation and developing PjBL support: Is it not the individual agent description that is pivotal, but a holistic view as in the "sum of the agents and their interactions" - since it is "by identifying the correlated feedback that we find a richer, and therefore more interesting, description of the system." [Carmichael and Hadzikadic, 2019, p.11] At the same time, macro-level behavior depends upon the standardised exchanges of agents at the micro-level. [ibid, p.12] Addressing the latter, Ryan and Deci have revealed "three needs that are critical for facilitating optimal functioning of the natural human tendencies for growth and integration and for constructive social development and personal wellbeing: the need for competence, relatedness (also defined as: *non-controlling positive feedback and acknowledgment of the other's perspective*; insert by authors) and (*support for*; insert by authors) autonomy" [Deci and Ryan, 2000]. Hence, not only tangible rewards but also threats, deadlines, directives, pressured evaluations, and imposed goals diminish intrinsic motivation because, like tangible rewards, they conduce toward an external perceived locus of causality. In contrast, choice, acknowledgment of feelings, and opportunities for self-direction were found to enhance intrinsic motivation because they allow people a greater feeling of autonomy [Deci and Ryan, 1985].

METHODOLOGY

In this section we sketch the Design Science framework, framing the Value Network Analysis (VNA). The VNA serves for making explicit the driving factors and/or hindrances of collaboration, affecting self-efficacy, whereas the Design Science-based research allows for stepwise investigating each factor from different perspective in a structured and balanced way with respect to theories and concrete design artefacts.

DESIGN SCIENCE-BASED RESEARCH

"Given the aim to generate prescriptive knowledge, DSR contributes to both the theory and practice of solving real-world problems." [Winter et al. 2020] Design Science-based Research (DSR) enables structuring KM activities while making use of its results for generating a problem solution. Design serves an *Artefact construction*[1], applying the artefact for the *Problem Solution (3. Design & Development)* into a *suitable context (4. Demonstration) observing its effectiveness and efficiency (5. Evaluation),* and *iterating it back* to the design level until a final *Scholarly or Professional publication* can be placed within phase *6. Communication* [Peffers et al. 2007]. "Design science (..) creates and evaluates IT artefacts intended to solve identified organisational problems." [Peffers et al, 2008] In this study – after having *identified the problem,* i.e. the need to position PjBL values for (digital) learning support, we focused on phase 2 of the operational framework, *defining the Requirements for the Solution,* asking "What would a better artifact accomplish?" [ibid, p. 54] In order to clarify the respective objectives for an amended and concrete concept of PjBL including (digital) learning support, the VNA has been deployed. It examines how the roles and value exchanges of the PjBL stakeholders can be best integrated to develop support features engendering the self-efficacy of all role carriers involved.

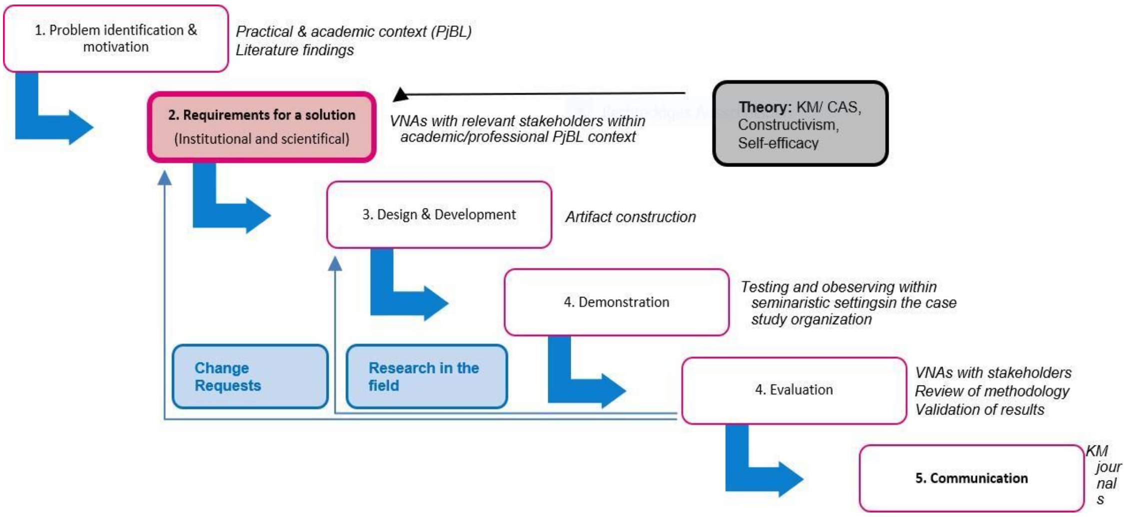


Fig. 2: Simplified DSR cycle "Project-based learning within digital environments" (by authors, according to Peffers et al, 2007)

VALUE NETWORK ANALYSIS

The VNA, as "analysis of the health and vitality of a value network" [Allee, 2003], addresses the overall patterns of interactions in a specific CAS, and how well a value moves through the system [ibid]. By looking at the impact of each value transaction between participants they can co-create the best way for them to extend and leverage value in the system at hand [ibid]. The method requires identifying all networked actors (roles) in the respective PjBL setting, and specifying the *tangible*[2] (or contracted) and *intangible* (or informal) value streams between them, according to the perception of each stakeholder. While a thorough picture of all roles and relevant transactions helps to identify regular *Sources* and *Sinks* in the *Exchange Analysis* by drawing a *Holomap*, the *Impact Analysis* requires each stakeholder to reflect upon the consequences of issued and received deliverables by identifying triggered activities, costs, risks and benefits. In the subsequent *Value Creation Analysis* several aspects are considered for developing a CAS further (i) *Assumed perceived value of recipient,* (ii) *General tangible asset utilization,* (iii) *Tangible costs of the deliverable,* and (iv) *Risk of providing the input.* Finally, those value exchanges are analysed according to the utilization of intangible assets and costs or benefits. With respect to changing the value stream each deliverable is checked for potential additional value creation, including its overall costs, risks and benefits. The peculiarity of this methodological approach is to be found in the retention of very detailed information from a micro-level stakeholder perspective which at the same time can be classified into a larger macro-level setting. Since it always depicts the individual view onto a specific context it does not claim to levy issues which can be validated or generalized - moreover its objective is to spawn so far undetected aspects of a certain situation for possible transferability to similar circumstances [Allee, 2002]. For this reason the method provides a contextual while and detailed overview about possible requirements for a solution when applying DSR.

[1] In this section all terms and expressions originating from [Peffers et al. 2007 and Peffers et al. 2008] are denoted in *Italics.*
[2] In this section all terms and expressions originating from [Allee, 2003] are denoted in I*talics*.

VNA-DRIVEN DEVELOPMENT OF PjBL: THE FIELD STUDY

We introduce the setting of the performed field study and detail how the VNA has been applied to meet to the objectives of this research and check the methodological feasibility of this mainly qualitative approach in the context of PjBL. We also report on the collected data and analysis results.

SETTING

The VNA was conducted as part of a cooperation between 2 departments of a German (UNI1) and an Austrian university (UNI2), both involved in Business Informatics study programs. When starting to study value exchanges that (can) enable PjBL, in order to create benefits for facilitators and learners, the initial phase of the Design Science framework can be addressed by the VNA. It then provides insight into already institutionalized PjBL processes. The outcome was consulted in order to extract the requirements for PjBL in a future (digital) setting in the Design Phase 2: *Define Objectives of a Solution* ["Artifact construction", Peffers et al. 2007], as it provided a thorough basis sketching relevant and valuable roles and transactions.
The study program at UNI1 has a setup in which students work within a designated partner company 85% of their working time over 24 respective 36 months planning, developing, executing and documenting a business-driven project. Seminars which make up the other 15% of time conduce with their content to the handling of those projects. In addition, students are supervised and coached singularly in their projects by a business supervisor and an academic project coach with methodological and professional input. The thesis is written about the project-specific task and linked research necessity and presented to an exam board.
The study program at UNI2 on the other hand provides that full-time academic students evolve a project assigned by their lecturer within a team of at least two members (one of them as project lead) over a semester period. In this time, they hand out the project contract to the lecturer and present milestone results in class to receive feedback and discuss their achievements.

PARTICIPANTS

The interviewed parties were two program representatives (Chief Scientists of Business Informatics) of UNI1 and UNI2 as well as three students of the Master course *Master of Science in Business Informatics (M.Sc.)* at UNI1 with aspired graduation in 2022 and two students of the Bachelor course *Bachelor of Science in Business Informatics (B.Sc.)* at UNI2 with aspired graduation in 2021/2022. Both institutes and the interview partners were selected according to their PjBL experiences. The VNAs with the chief scientists were completed between December 2020 and February 2021, those with the students between January and September 2021.

PROCEDURE

All VNA sessions started with explaining the purpose of the study, the setting and the number of participants. Each participant also was asked for consent with respect to video recording and documentation (validated by the participant afterwards). After acquiescence was obtained, a short presentation was given by the interviewer explaining why the combination of PjBL and digital learning support will play a larger role in education in the future. Next, the practical effect and procedure of the VNA as instrument to go deeper into the details and implicit knowledge of the specific stakeholder of a subject matter were expounded, before the interview started with the stakeholder-specific questions about the PjBL content and process at their institute according to their own experience and cognizance. Quality checks were provided by an experienced researcher in VNA as well as an assistant providing feedback on each session.

CAPTURING THE PERSPECTIVE OF UNI1 CHIEF SCIENTIST

Knowledge elicitation started with the chief scientist of UNI1. In order to set the context right, introductory items were asked and guided the VNA:

a) What is the motivation of your institute to offer or work with project-based learning?

b) Which of its characteristics justify this approach?

c) How can these be found in tangible and intangible value exchanges within in the specific roles and their interactions?

Research question 1 (Which role understanding do stakeholders have when being involved in PjBL?) has been addressed in the created Holomap and subsequent Exchange Analysis with focus on the role of the interviewee. Research question 2 (Which value exchanges as perceived by role carriers contribute to the effectiveness of PjBL?) was referred to individual deliverables on the micro-level in the Impact Analysis, and to the value chain on the macro-level in the Exchange Analysis. The Value Creation Analysis led to offers dedicated to digital learning support including the anticipated impact on self-efficacy.

Holomapping / Exchange Analysis. The Holomap with the most relevant roles of the specific PjBL implementation at UNI1 as perceived by the chief scientist of UNI1 is given in Figure 3a. The Exchange Analysis revealed the role of the chief scientist as an essential source of information, with a missing link to the networked PjBL actors, in particular the students. Students and student supervisors were considered central with plenty of valuable interactions. ***Impact Analysis.*** It focused on the tangible and intangible value exchanges shown in Table 1. When taking a closer look on the activities triggered by received deliverables and their influence on further intangible assets including *Human Competence*, *Internal Structures* and *Business Relationships*, the perception of the chief scientist revealed a high impact of the value exchanges on human competence regarding the enhancement of didactic expertise and know- how in the field of project management as well as a general improvement of project management competences and practical skills.
Furthermore, an effect on the internal structures was noted by him in two areas: for once as change demands within the organisation of teaching (e.g., in terms of group or cluster building), but mostly in an augmentation of administrative efforts. Both aspects already showed the potential of the implemented method to foster competences of program participants and facilitators in new areas as well as a shift within teaching structures in this second phase of the VNA.

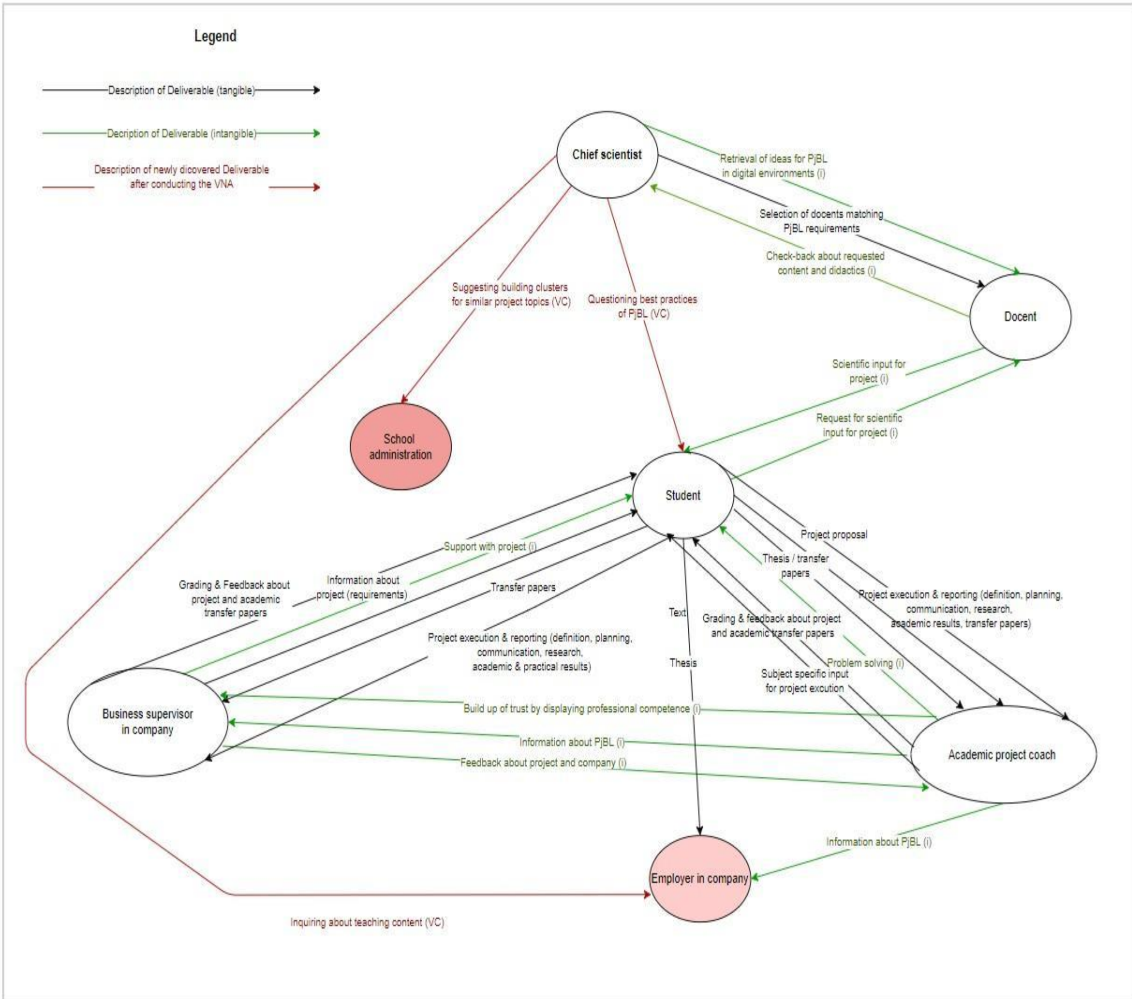


Fig. 3a: Holomap of Exchange Analysis – Chief Scientist UNI1 (translated into English)

Table 1: Extract of Impact Analysis of Chief Scientist UNI1 (translated into English)

Transactions			IMPACT ANALYSIS							
Deliverable	From	To	Activities that are generated	Impact on financial resources +/-	Impact on intangible assets + / -			Overall cost / risk for this input	Overall benefit of this input	Perceived val
					Human Competence	*Internal Structure*	*Business Relationships*	*H = High M = Medium L = Low*	*H = High M =Medium L = Low*	+2, +1 Neutral
Check-back about contents & didactics	Docent	Chief Scientist	Efforts at persuasion & discussion about PjBL	+ Additional expenditure	+ Immense expertise about PjBL didactis	+ Amendment of teaching structures (for ex. group	+ Shift of focus at future selection of docents	L: 1h Check-back / M: Subjectivity (unsystematically gathered opinions will possibly become integrated)	H: Apprehension of docents' view on PjBL (multi-perspective)	+2
Project proposal including project seminar paper	Student	Academic project coach	Assessment of paper and feedback about project	+ Partly compensated as part of project coaching contract / internal absorbtion of capacities	+ solid and compact project know-how	- Considerable organisational and administrative efforts	+ Build up of personal relationship with company	M: 1 d for Assessment, feed back und partcipating at presentation	H Potential and risk of project identified	+2
Project reporting	Student	Academic project coach	Feed back about project and support	+ Partly compensated as part of project coaching contract / internal absorbtion of capacities	+ Enhancement of project competencies	- Backed by existing structures	+ Supports successful handling of project and goof reputation at company	M: 1h/month Inquiring about project states / M In single cases risk that state will not be requested regularly	H Potential benefit by successful project (central element of study program)	+2
(intangible) Request for input about project	Student	Docent	Feed back and potential involvement into project	- Siginificant unpaid additional expenditures for docent	+ Know-how gain about practical problems	No impact	+ In single cases build up of contact to company	L: 1h Additional expenditures / M: Founded input not possible if knowledge of docent is too sparse	M: Focussing on practical topics as new input for docent	+1

Source: Authors (2021)

Looking closer at the 8th column of the *Impact Analysis* the main outcomes as overall benefits of the valuable interactions from the perspective of the chief scientist can be summarized: Enhanced expertise of lecturers about PjBL; potential and risk of the project are detected in advance; potential benefit of a successful project as constituent part of the study program; advancement of student and project; further fielding in company and best practice; *(intangible)* Focus on practical problems as input for docent; *(intangible)* existing expertise becomes vivid: improvement of project execution.

Again, all aspects named above serve as recognition that PBjL is apt to spawn the expertise and benefits of all stakeholders within its construct. The Impact Analysis also allowed evaluating each interaction on a monetary basis. Simultaneously, the implicit knowledge of the chief scientist was externalized in his estimations of the overall cost and risk for this input.

Value Creation Analysis. Additional possible transactions from the sender perspective creating onward value focused on *what is done already* in order to yield the best possible PjBL experience for both learners and facilitators (docents, coaches, supervisors, and companies) rather than *what could be done on top* to outperform the actual status. A new role and novel value exchanges were added to the Holomap and marked accordingly (see also Table 2a):

- Selection of lecturers matching PjBL requirements (chief scientist to docents)
- Retrieval of new PjBL ideas in a digital environment (chief scientist to docents)
- Building of clusters between students with similar projects (chief scientist to school administration) – including newly defined role of school administration
- Inquiry of relevant teaching content at companies in order to handle projects ideally (chief scientist to employer company) – newly added importance of employer company
- Query of Best Practices in PjBL (chief scientist to selected students)

Table 2a: Value Creation Analysis of Chief Scientist UNI1 (Extract)

Transactions (Chief Scientist)			Perceived Value	VALUE CREATION ANALYSIS		
Deliverable	From	To	Recipient highly values this deliverable. Strongly agree (+2) Agree (+1) Neutral (0) Disagree (-1) Strongly disagree (-2)	Tangible asset utilization is: H = High M = Medium L = Low	What are the tangible costs? (concerning financial and physical resources)	How high is the risk factor in providing this output? H = High M = Medium L = Low
(intangible) Retrieval of new ideas for teaching in digitals learning environments	Chief Scientist	Docent	+1	H	1h / Exchange with docent	L
Clusterbuilding	Chief Scientist	School administration	-1	H	1 d / Structural changes 1 h / per seminary: Seminar planning 2 h / per seminary: Preparation of clusters	M
Retrieval of important contents at companies for best possible project execution	Chief Scientist	Business Supervisors	+1	M	2 d / year with preparation and follow-up	L
Requesting best practices of PjBL	Chief Scientist	Students	+1	H	2 d / year (Interviews and Evaluation)	L

Intangible asset utilization is: H = High M = Medium L = Low	What are other intangible costs or benefits? (Industry, society, environment)					How do we add to, enhance, or extend value?	What is the overall combined cost/risk for this input?	What is the overall benefit for us in providing this input?
HC = Human Competence	IS = Internal Structures	BR = Business relationships	Organizations	Society	Environment			
H Enhancement of innovative skills	M Could partly become institutionalised	L Not relevant	+ Emerging of problem awareness regarding	0	0	Intensifing cooperation by integration docents into doctrine	Ideas might not be effectuated (frustration of idea giver)	Creation of unique PjBL features Active fostering of innovation in teaching
H Immersion of professional expertise	H Thinking in clusters and joined-up thinking	H Cognition of high professional expertise of	0	0	+ Innovative teaching method discuessed i scientific circles	Transfer of new knowledge onto project enhances learning success as cause of knowledge exchange within cluster	Economic difficulties to realize clusters Benefit: ideal peer group in knowledge domain	Project will be handled even more successully
H Improvement of comprehension of professional practice	M Better internal structuring	H Acceptance of companies as	+ Good and constructive relation	0	0	Teaching content will be imparted practically and project oriented	Risk of considering single companies' interests too strongly Benefit: teaching along the needs oft he market	Companies close project contract with institute
H Graspingt he young generation	H Could partly become institutionalised	M Cognition of students as stakeholders	+ Good and constructive relation to students	0	0	Perspective of learners will beincorporated and leads to „thrilled" students	Risk of finding the „right" students who have a good project to report about Benefit: Receiving the current learning requirements of the current generation	„Thrilled" students recommend institution and give positive feedback

Source: Authors (2021)

Together with the findings of the Exchange and Impact Analysis, the outcomes of these reflections on value creation according to their overall benefit as well as *overall cost and risk* inform building an artefact for PjBL and evaluating a possible later realization:

Table 2b: Value Creation Analysis of Chief Scientist UNI1 – Reflections on overall benefit, costs, and risks

Ascertaining the quality of academic services / *quality of academic services might be jeopardized*
Generating a unique characteristic regarding PjBL; Innovation of teaching is actively fostered / *possibility that ideas cannot be implemented (frustration of the idea giver)*
Projects are handled better within an ideal peer-group / *Cluster-building might not be feasible because of economic reasons*
Enterprises close project contracts with the academic institute as it offers the exactly needed training / *Specific interests of a few enterprises might be considered too strongly*
Inspired students recommend new students and give positive feedback to their employer; the actual learning requirements are met / *not all students have true best practices in their projects to report about*
Project competencies are substantiated; production of user value within partner companies / *qualified project coaching requires considerable timely effort and qualified competencies*
Know-how enhancement for project advisors / *need to deploy merely very good papers*
Conveying „problem-solving“ ability / *little effort and little risk*
Conveying competence of docents / *little effort but risk about too sparse knowledge of teachers about PjBL*
Building up reputation at companies / *little effort, yet risk that lack of subject-specific competency emerges if not existent*

CAPTURING THE PERSPECTIVE OF UNI2 CHIEF SCIENTIST

The research questions were answered following the same procedure as before and described for UNI1 chief scientist above. Due to space limits we will focus on the value creation results.

Holomapping / Exchange Analysis. The Holomap (Figure 3b) depicts all perceived roles and their value exchange in the context of PjBL. The Exchange Analysis revealed the Facilitator (Chief scientist in his role as organizer and supporter of PjBL at the institution) as source of value exchanges, in terms of process attendance, high responsibility, high didactic qualification, highly developed skills for mentoring and empowerment of project staff.

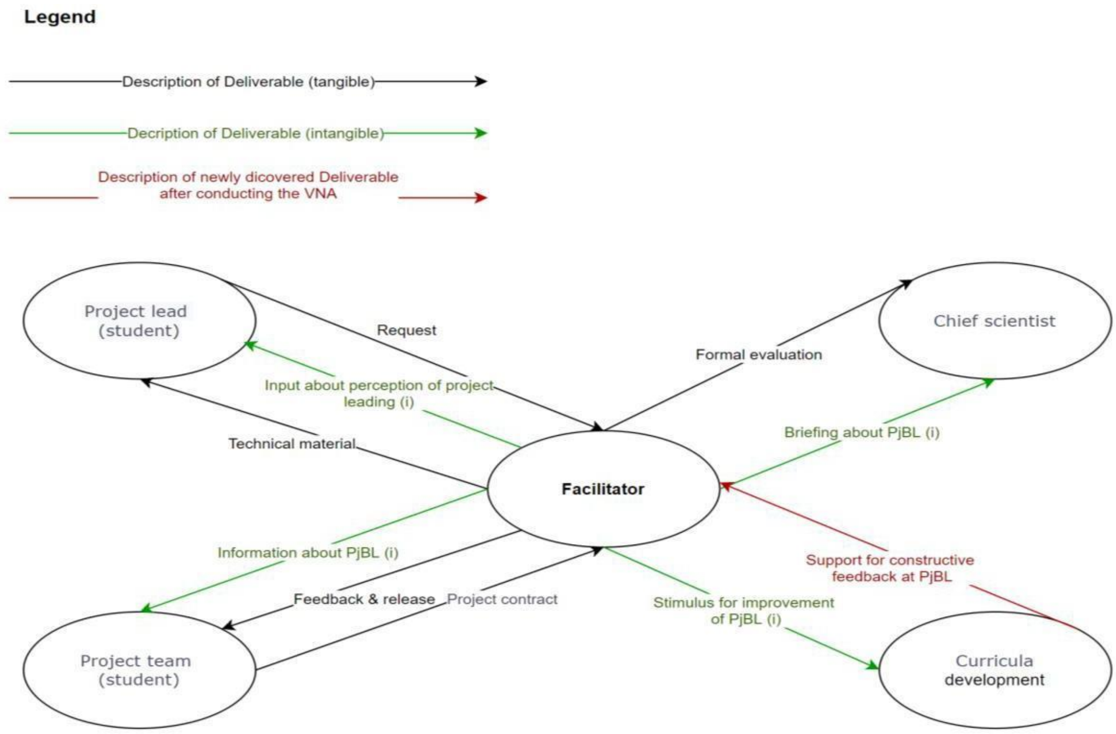


Fig. 3b: Holomap of Chief Scientist UNI2 (translated into English)

Impact Analysis. This analysis focused on essential deliverables, including the tangibles Project Contract and Project Lead Requests, and the intangible Support for constructive feedback on PjBL from Curricula developers. The main outcomes as overall benefits of the valuable interactions from the perspective of the chief scientist were:

(1) Structural assistance for autonomous conduct of teaching assignments

(2) Insight into the project from an operational and a responsible perspective

(3) *(Intangible)* Insight into the development of the curriculum regarding PjBL

At the same time the implicit knowledge of the chief scientist became evident in his estimations of the overall cost and risk for this input, as he explains in the following extract of the data:

(1) *Cost:* 10 hrs / Risk: (Medium) given structure in the project contract

(2) Cost: 5 hrs / Risk: (Medium) communication problems from time to time

(3) *(Intangible)* Cost: 6 hrs / Risk: (low)

Value Creation Analysis. Here, the interviewee decided to look upon the following transactions:

- Feedback & release (from Facilitator to Project team)
- *(Intangible)* Information about PjBL (from Facilitator to Project team)
- Technical material (from Facilitator to Project leader)
- *(Intangible)* Input to perception of leadership role (from Facilitator to Project leader)
- Formal evaluation (from Facilitator to Chief scientist)
- *(Intangible)* Briefing about PjBL (from Facilitator to Chief scientist)
- *(Intangible)* Stimulus for improval of PjBL (from Facilitator to Curricula development)

When reflecting on overall benefit / overall costs and risk, he could identify (like his counterpart from UNI1) several opportunities: Possibility to establish feedback culture as inherent part of the curriculum / no cost or risk; Facilitation of content-related processing of project / effort of allocation while additional value and quality is not ensured through respective checks; Constructive support of leadership / effort of knowledge transfer, lack of quality management (critical); Feedback may be deployed for personal advancement / effort of evaluation, validity is not ensured due to lack of survey data; Knowledgeability / effort in knowledge transfer and preparation, universal validity questionable; Prompt incorporation of feedback and ideas for PjBL into curriculum / effort of integration and quality assurance, volatility of PBL approaches. Again, these discovered values and critical points require refinement for PjBL artefact design. *Addition, enhancement and extension of value.* Asked about possibilities to amplify the merits which had been detected so far, the chief scientist of UNI2 came up with suggestions, among them the emendation of curriculum that might profit from regular feedback loops, material deployment for quality assurance, and mutual reflection on leadership with others.

CAPTURING THE PERSPECTIVE OF STUDENTS

3 students of the Master class Business Informatics at UNI1 and UNI2 students of the Bachelor course at UNI2, all involved in PjBL, were interviewed following the same procedure as presented above. The items the learners had to answer within the VNA were slightly reformulated in order to meet their personal situation and interest in the study program:

a) What is your perception of the project-based study program at your institute?

b) Which of the characteristics of this specific learning support technique are typical regarding your own experience with it?

c) Where can those characteristics be found in tangible and intangible value exchanges within in the specific roles and their interactions?

Evaluating the data of the 5 VNAs led to the following insights for each of the research questions:
Role understanding when being involved in PjBL: The Holomaps of the students (see Appendix) diverge, as the students of UNI1 position their (single) role as central for all value exchanges (according to the UNI1 study program), while the students of UNI2 perceive themselves as part of a network of project team members on the one hand and academic facilitators on the other hand with no further contact to stakeholders outside of the university.

Value exchanges contributing to the effectiveness of PjBL:

- Very strong impact on motivation for students through input and feedback from colleagues and their student colleagues/team members
- Very strong demand by students to receive more professional and "insider" input for their projects by lecturer
- Provenly best results by operating through trustful, flexible, and fast cooperative teamwork without given guidelines
- For developing (digital) learning support targeting the self-efficacy of PjBL, the value could be created by multiplication of feedback by colleagues in business and other lecturers.

FINDINGS

In this section we discuss the collected data for stakeholder-centred PjBL learning support, consolidating the acquired holomaps and deriving implications for practising PjBL.

HOLOMAP CONSOLIDATION

Combining the holomaps referring to the perceived value exchanges of all interviewed PjBL stakeholders provides a rich picture on all roles involved in PjBL processes. It reveals an embedding of the student learner in a strong community of learning supporters and supplements the original singular relationships to supervisor and coach with further ones as lecturers and colleagues. Furthermore, the outer circle of institutional supporters as the chief scientist and the project coach are completed by the role of the company employer on the one and the role of the facilitator on the other hand. Therefore, it supplies a compound picture of all relevant roles from the perspectives of the involved stakeholders allowing for the design of a role concept for a thoroughly remote and offline PjBL support in the next DSR phase.
Particular attention within the *Definition of objectives* for the *Artefact construction* should be paid to possible weak spots discovered by this study:

- At UNI1 the demand for more contact to students, to school administration and to the company employer by the chief scientist was identified with the desire to gain more information about *Best Practices*, availability of teaching content, and to be able to support the structuring of learning clusters.

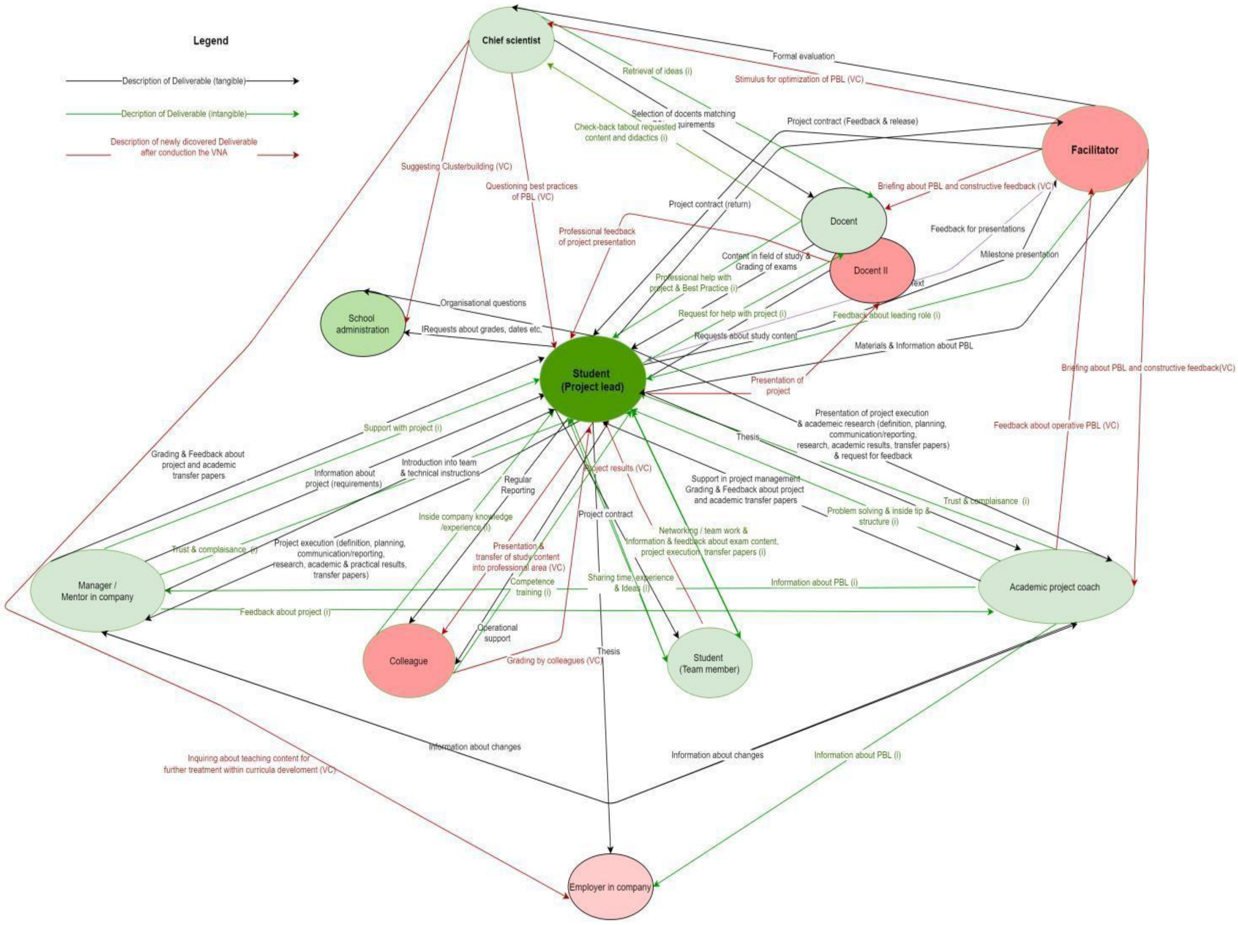


Fig. 4: Combined Holomap for PjBL from the perspectives of facilitators and learners (UNI1 and UNI2)

- At UNI2 the role of the facilitator was found to be very valuable once being integrated permanently, supporting all processes while students had the main responsibility in the project.
- Master students at UNI1 would appreciate more qualitative and institutional contact, promoting PjBL in their specific context through tangible interaction with lecturers.
- Bachelor students at UNI2 would prefer presenting their results to multiple lecturers, to receive fruitful feedback in addition that from their coaching lecturer to improve their learning processes.

IMPLICATIONS FOR PjBL PRACTICE

Several suggestions for implementing PjBL in (digital) learning settings could be derived: ***Informal and formal space for exchange and cluster building.*** Especially the VNAs of the students revealed the necessity for successful PjBL to place at the disposal time, spaces and platforms in order for the students to interact and exchange ideas, problems and results of their projects with each other in an informal and safeguarded way. The VNAs reveal the need of regular and tangible contact with experts, colleagues, and academic teachers. The importance, benefit and need of asking students to present their project results and experience to various audiences at the academic institution are evident. Especially the value exchanges of “delivering and receiving feedback” need to be considered more thoroughly as refined for implementing effective learning support. For instance, facilitators seem to profit similarly from questions and input by involved participants as the learners or their company contacts. Further studies appear imperative to set up the suggested learning *clusters* possibly supporting self-efficacy of PjBL. Moreover, specifically one chief scientist vividly sensed in fact two value nets or streams on different levels, which is retrieved in the Combined Holomap (Figure 4) merging all stakeholders’ perspectives: the inner value network takes its starting point from the student as project leader in contact with all operative stakeholders (business supervisor, academic project coach, docents, colleagues, and student teams members or fellow students), while the outer net stretches along the institutional layer including chief scientist, facilitator, company employer, and project coach. This insight requires in-depth studies as such understood by Allee [Allee, 2009], in particular in digitally supported PjBL environments.

New role definitions for facilitator, docents, professional colleagues, school administration and project team

- The role of a *facilitator* has been found within the VNA of the chief scientist of UNI2 who merely concentrates on supporting PjBL in all directions by giving stimulus for optimization of PjBL to the chief scientist, agreeing upon a project contract with the students, letting them present their milestones and giving specific feedback to them and briefing the project coaches and docents about PjBL while receiving constructive feedback about operative PjBL from them. According to his perception the placement and assigning of a person in charge of facilitating PjBL within a PjBL institution seems to be prevailing capturing three different perspectives: content-wise (in regard to project), technical (in regard to curriculum) and formal (as in regard to learning targets, e.g., field placement).
- Colleagues are discerned to hold a quite important part in PjBL and have “gained” new deliverables by value creation of the interviewed Master *students* at UNI1 as: receiving transfer presentations of study content and being assigned to grade students as well to feedback the practical value of their academic project work. Furthermore, here the aspect of intangible trustful exchanges is rated very high and considered productive by the stakeholders in regard to PjBL.
- The *project team* as well was addressed by the Bachelor students at UNI2 as an extremely supporting and inspiring role especially by sharing results and ideas with each other. The benefit of these exchanges for PjBL in their view cannot be underestimated and is understood to help institutions featuring team learning in different modes.
- The role of the *school administration* was defined as relevant in order to organise cluster learning by both chief scientists. Nevertheless, it is striking that in the consolidated Holomap of Value Exchanges this role is not of major importance for PjBL, whereas it might be seen as the “backbone” of a classic learning setting. In the sense of Verna Allee this could either mean that it is – as a role – rather redundant, or – as another possibility
- that it needs to become a more visible role concerning its usefulness for all learners. Here, we suggest further research in order to define more in detail which deliverables could foster the value of the role.

CONCLUSION

When aiming to establish Project-based Learning (PjBL) as 21th century education format and skill, we need to elicit and utilize knowledge from the perspective of involved stakeholders, in particular for the design of effective digital learning support. The selected methodological approach, namely Value Network Analysis framed by Design Science Research, structures knowledge acquisition and subsequent development processes, taking into account implicit knowledge and multiple perspectives on PjBL. We could collect design requirements for (digital) learning support through consolidating elicited stakeholder interactions in the course of PjBL and its management processes. Featuring (informal) knowledge sharing and learning loops seem to be crucial for effectively supporting PjBL facilitators and learners. Since we started with validated PjBL knowledge on the macro-level, further studies for refinement operational implementations on the micro-level focusing on self-efficacy need to be performed.

REFERENCES

Abuhmaid, A. (2020): *The Efficiency of Online Learning Environment for Implementing Project-Based Learning: Students' Perceptions.* International Journal of Higher Education. Hashemite University, Jordan; Correspondence: Curricula and Teaching Methods Dep, Hashemite University, Jordan. Published: July 6, 2020

Aldabbus, S. (2018): *Project-based Learning: Implementation & Challenges.* International Journal of Education, Learning and Development, Vol.6, No.3, 71-79, March 2018

Allee, V. (2002): *„Die Darstellung von Unternehmenswissen".* In Bellmann, M.; Krcmar, H.; Sommerlatte, T. (Hrsg.): Praxisbuch Wissensmanagement. Düsseldorf: Symposium Publishing

Allee, V. (2003): *The Future of Knowledge: Increasing Prosperity through Value Networks.* Amsterdam: Butterworth-Heinemann

Allee, V. (2008): *Value Network Analysis and value conversion of tangible and intangible assets.* Published in Journal of Intellectual Capital, Volume 9, No. 1, 2008, 5-24

Allee, V. (2009). *Value-creating networks: Organizational issues and challenges. The Learning Organization*, 16(6), 427–442, 2009

Arkorful, V., Abaidoo, N. (2015): *The role of e-learning, advantages and disadvantages of its adoption in higher education.* International Journal of Instructional Technology and Distance Learning, 12 (1), 29–42, 2015

Baskerville, R., Baiyere, A., Gregor, S. (2018): *Design Science Research Contributions: Finding a Balance between Artifact and Theory.* Journal of the Association for Information Systems (2018) 19(5), 358-376

Amini, R. et al. (2019) *The difference of students learning outcomes using the project-based learning and problem-based learning model in terms of self-efficacy.* Published under licence by IOP Publishing Ltd, Journal of Physics: Conference Series, Volume 1387, 012082, International Conference on Education, Science and Technology 2019, 13–16 March 2019, Padang, Indonesia

Atkinson, C. J. and Brooks, L. (2005) *In the Age of the Humanchine*, In Proceedings of International Conference on Information Systems (ICIS), AIS, Las Vegas, USA.

Bandura, A. (1977) *Self-efficacy: Toward a unifying theory of behavioral change.* Psychological Review, 84(2), 191–215

Bilgin, I. et al. (2015*) Learning on Undergraduate Students' Achievement and Self Efficacy Beliefs Towards Science Teaching.* Eurasia Journal of Mathematics, Science & Technology Education, 2015, 11(3), 469-477

Candrlic, S., Pavlic, M.; Katic, M. (2020) *Information System Design and Development and Project-based Learning.* DOI: 10.5220/0009393604040411 In Proceedings of the 12th International Conference on Computer Supported Education (CSEDU 2020) - Volume 1, 404-41, Science and Technology Publications, Lda.

Carmichael T., Hadžikadić M. (2019) *The Fundamentals of Complex Adaptive Systems*. In: Carmichael T., Collins A., Hadžikadić M. (eds) Complex Adaptive Systems. Understanding Complex Systems. Springer, Cham. https://doi.org/10.1007/978 -3- 030-20309-2

Chen, C.-H.; Yang, Y.-C. (2019) *Revisiting the effects of project-based learning on students'academic achievement: A meta-analysis investigating moderators*, in: Educational Research Review 26 (2019) 71-81

Choi, J., Lee, J.-H., Kim, B. (2019): *How does learner-centered education affect teacher self-efficacy? The case of project-based learning in Korea.* Teaching and Teacher Education Volume 85, October 2019, 45-57

Chu, S. K. W., Tavares, N. J., Chu, D., Ho, S. Y., Chow, K., Siu, F. L. C., Wong, M. (2012). *Developing upper primary students' 21st century skills: inquiry learning through collaborative teaching and Web 2.0 technology*, Center of Information Technology in Education, The University of Hong Kong, Hong Kong

Dewey, J. (1938). *Experience and education*. New York: Collier

Geng, S., Law, K., Niu, B. (2019): *Investigating self-directed learning and technology readiness in blending learning environment.* International Journal of Educational Technology in Higher Education (2019) 16:17

Kakalejčík, L., Paľová, D. (2019) *Enhancement of students' skills via project-based learning.* MIPRO 2019/CE, GrowthPro, Košice, Slovakia

Mahasneh, A., Alwan, A. (2018): *The Effect of Project-Based Learning on Student Teacher Self-efficacy and Achievement.* International Journal of Instruction, 2018, 11(3), 511- 524.

https://doi.org/10.12973/iji.2018.11335a

Mannila, L., Nordén, L.A. and Pears, A. (2018) *Digital Competence, Teacher Self-Efficacy and Training Needs.* ICER '18, August 13–15, 2018, Espoo, Finland. 2018

Meskhi, B., Ponomareva, S., Ugnich, E. (2019): *E-learning in higher inclusive education: needs, opportunities and limitations.* Int. J. Educ. Manag. 33, 424–437 (2019)

Ngereja, B., Hussein, B., Andersen, B. (2010): *Does Project-Based Learning (PBL) Promote Student Learning? A Performance Evaluation.* Educ. Sci. 2020, 10, 330.

OECD (2018) *The Nature of Learning. Using Research to Inspire Practice*

Peffers, K., et al. (2008): *A Design Science Research Methodology for Information Systems Research.* Journal of Management Information Systems / Winter 2007–8, Vol. 24, No. 3, 45–77

Peffers, K., Tuuanen, T., Rothenberger, M. (2007) *A Design Science Research Methodology for Information Systems Research.* Journal of Management Information Systems / Winter 2007–8, Vol. 24, No. 3, 45–77. 2008 M.E. Sharpe, Inc., DOI 10.2753/MIS0742-1222240302

Piaget, J. (1929). *The Child's Conception of the World*. London: Paul Trench and Trubner. Piaget, J. (1970). *Genetic epistemology*. New York: Columbia University Press

Piaget, J. (1971). *Science of education and the psychology of the child*. New York: Viking Press

Ryan, R. M., Deci, E. L. (2000) *Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being* University of Rochester, American Psychological Association, Inc., Vol. 55, No. 1, 68-78 DOI: 10.1037110003- 066X.55.1.68 in: Giddens' theory of structuration: a critical appreciation, Routledge, London, 2000

Shin, M.-H. (2018) *Effects of Project-based Learning on Students' Motivation and Self- efficacy*. English Teaching, Vol. 73, No. 1, Spring 2018

Stary, C. (2014): *Non-disruptive knowledge and business processing in knowledge life cycles – Aligning value network analysis to process management.* Journal of Knowledge Management, 18(4), 651-686, 2014

Suastra, W., Ristiati, N. (2019) *Developing Critical Thinking, Scientific Attitude, and Self- efficacy in Students through Project Based Learning and Authentic Assessment in Science Teaching at Junior High School.* I J. Phys.: Conf. Ser. 1233 012087

Sumarni, W. (2013): *The Strengths and Weaknesses of the Implementation of Project Based Learning: A Review.* International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064

Tilchin, O., Kittany, M. (2016): *Adaptive Knowledge Management of Project-based Learning.* Journal of Education and Training Studies, Vol. 4, No. 6; June 2016

Usher, E., Pajares, F. (2008): *Self-Efficacy for Self-Regulated Learning.* Educational and Psychological Measurement 68, 3 (2008), 443–463

Ültanir, E. (2012). An epistemological glance at the constructivist approach: Constructivist learning in Dewey, Piaget, and Montessori. *International journal of instruction*, *5*(2), 195-212, 2012

Weichhart, G., Stary, C., Appel, M. (2018): *The digital Dalton Plan: Progressive education as integral part of web-based learning environments.* Knowledge Management & E-Learning, 10(1), 25–52. 2018

Winter, R., Hevner, A., Maedche, A. (2020): *Accumulation and Evolution of Design Knowledge in Design Science Research - A Journey Through Time and Space.* Journal of the Association for Information Systems · May 2020

Yilmaz, R., Yilmaz, F., Keser, H. (2020): *Vertical versus shared e-leadership approach in online project-based learning: a comparison of self-regulated learning skills, motivation and group collaboration processes.* J Comput High Educ 32, 628–654, 2020

APPENDIX

The Holomaps are original (in German) to provide insights into the various interaction patterns.

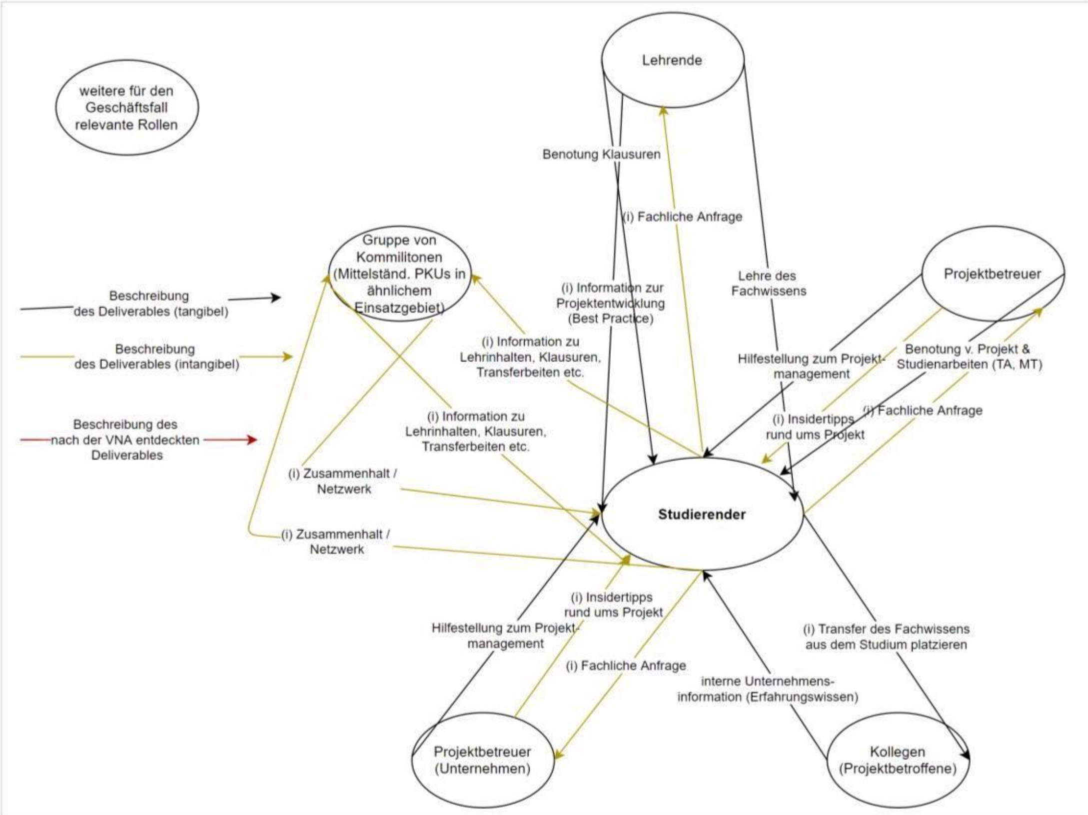


Fig. 5: Holomap of Exchange Analysis of Master student no. 1 at UNI1

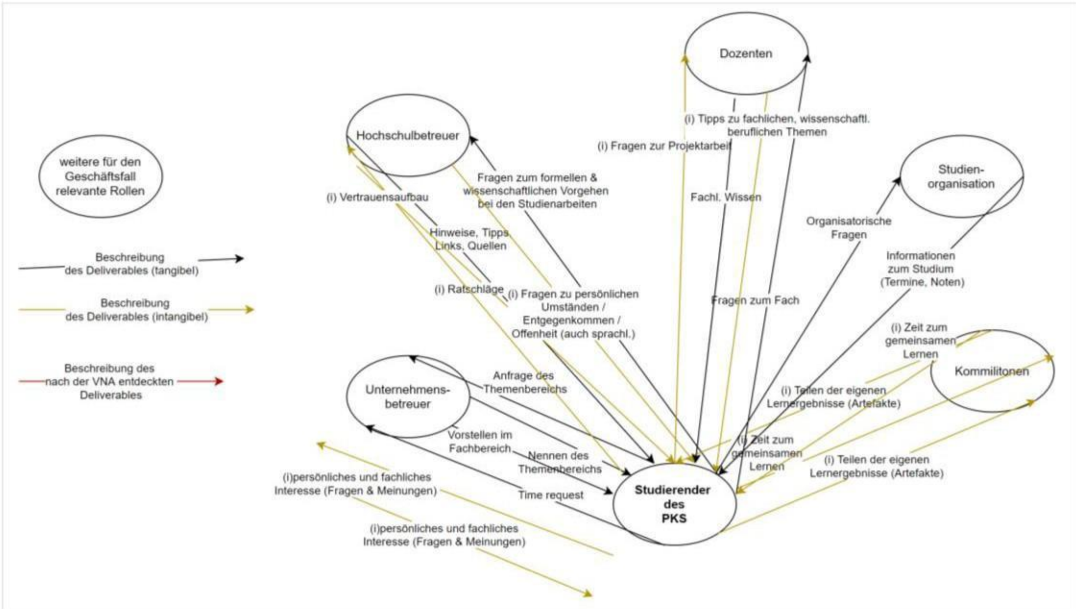


Fig. 6: Holomap of Exchange Analysis of Master student no. 2 at UNI1

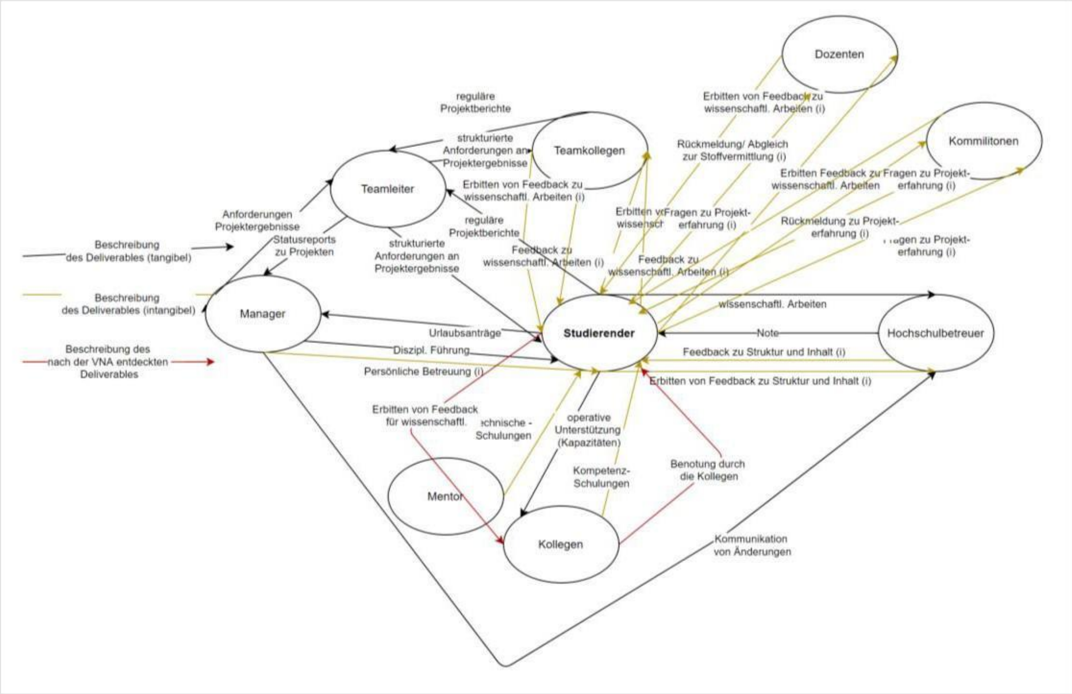


Fig. 7: Holomap of Exchange Analysis of Master student no. 3 at UNI1

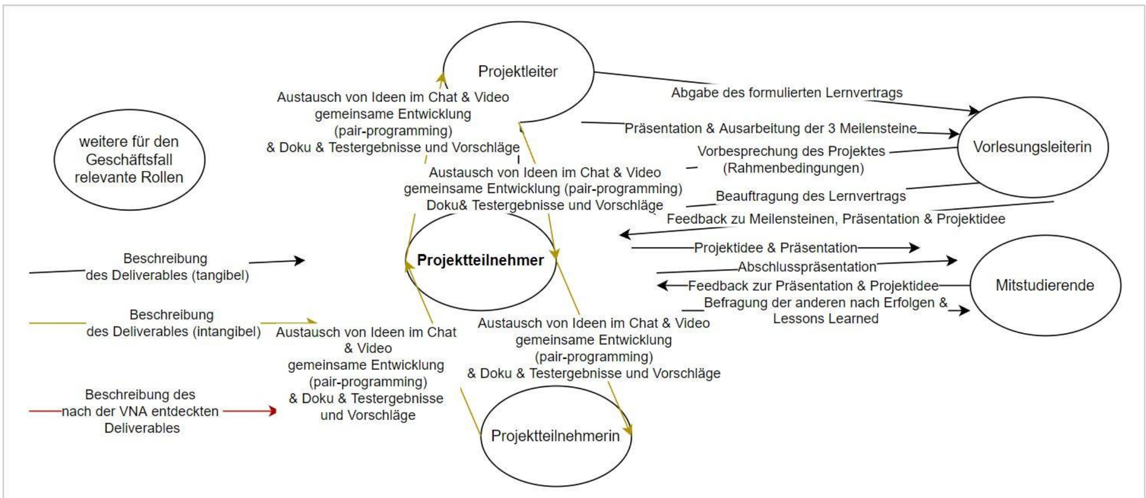


Fig. 8: Holomap of Exchange Analysis of Bachelor student no. 1 at UNI2

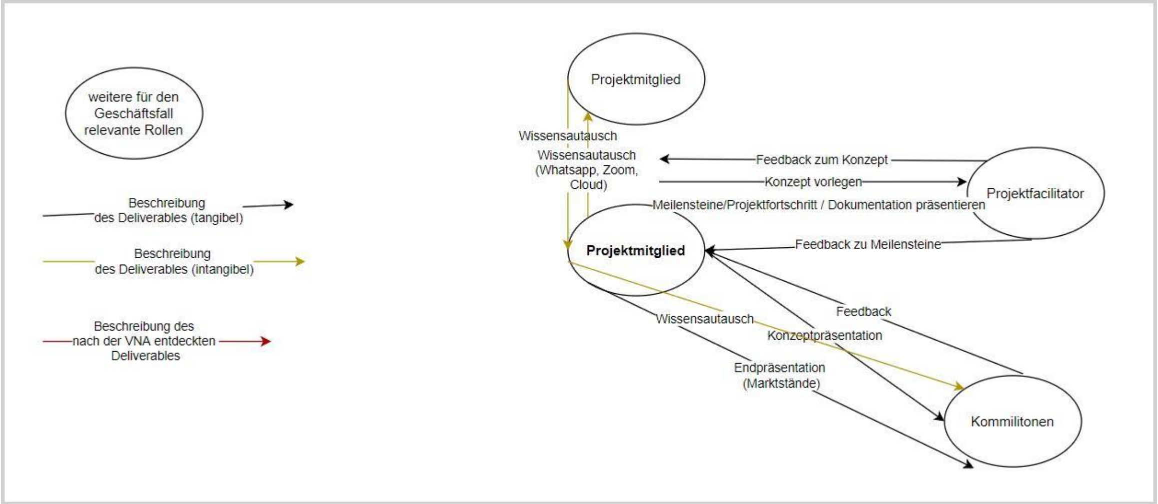


Fig. 9: Holomap of Exchange Analysis of Bachelor student no. 2 at UNI2

Neo-learning methodology to facilitate the transfer of learning into practice

GRAZIELA GRANDO BRESOLIN

Knowledge Engineering Department, Federal University of Santa Catarina, Florianópolis, Santa Catarina, Brazil
E-mail: grazielabresolin@gmail.com

PATRICIA DE SÁ FREIRE

Knowledge Engineering Department, Federal University of Santa Catarina, Florianópolis, Santa Catarina, Brazil
E-mail: patriciadesafreire@gmail.com

Neo-learning was developed as a teaching and learning methodological platform for the Networked Corporate University (NCU) model, in order to promote experiential and expansive learning for the development of competences and facilitate the transfer of learning to work practice. The purpose of this article is to present the neo-learning as a methodological support of the Networked Corporate University model to facilitate the transfer of learning into practice. To achieve this objective, a qualitative approach was chosen for exploratory and descriptive purposes, through a narrative literature review and multiple case studies of the application of the methodology. This article presents the main approaches and theories that underlie the NCU model and the neo-learning methodology. The neo-learning cycle and the applications that validate the methodology are also presented. It is concluded that the neo-learning methodology was developed in accordance with the NCU model to support the teaching and learning processes. The methodology can be considered validated, since it has been applied in several research and extension projects, in academic and corporate, public and private universities, in undergraduate and graduate programs, for initial training, for the developing of teachers and tutors of the presential, online, and remote modalities.

1. Introduction

The digital transformations allied to the context of the knowledge society demand from individuals a continuous learning for the development of new competences resulting from these transformations. In this sense, organizations also need to develop digital, technical, and socio-emotional skills in their employees for the competent performance of organizational changes (Antonelli et al., 2013; Freire et al., 2016; Rhéaume & Gardoni, 2015).

These transformations encourage the development of new university models, which are distinguished from traditional training and education centers, since they incorporate the human and organizational qualification and development, in order to promote continuous and multi-level learning, also involving the stakeholders of organizations' internal and external networks (Freire et al., 2016).

In this context, the contemporary model of the corporate education system emerges, called the Networked Corporate University (NCU) model. The model was defined by Freire et al. (2016) as an intelligent environment of continuous education, physical or virtual, that manages and institutionalizes a culture of networked learning. As a networked model, it promotes integration, collaboration, and co-creation between organizations, employees, academic universities, scientific and technological institutes, suppliers, consumers, and other participants in social productive arrangements for the development of human, social, and relational capital.

Developing a model to act in a learning network requires an open innovation approach to establish partnerships between the stakeholders of the value chain. Consequently, organizations that adopt this approach to innovation obtain external competences from their network, sharing their knowledge, and integrating competences with the network's stakeholders (Rhéaume & Gardoni, 2015). As a result of the expansion and diversification of the network's stakeholders, there is an increase in the transfer of knowledge and the creation of an environment of sharing, collaboration, and co-production with the organizational ecosystem, enriching the learning network (Chen et al., 2019; Rhéaume & Gardoni, 2015).

The NCU model promotes, through the network, continuous, experiential, and expansive learning with the objective of co-creation and co-production of essential knowledge for competent performance. In this way, the NCU manages the learning processes, where both the learner and the teacher have the opportunity to create new knowledge during the experience, and this knowledge is expanded when co-created with the network stakeholders (Bresolin, Freire & Silva, 2021; Engeström & Sannino, 2010; Kolb & Kolb, 2017).

To meet this complexity, neo-learning was developed as a methodological platform for teaching and learning of the Networked Corporate University model. It is an innovative way of creating learning environments to develop skills and motivate the transfer of learning to practice.

According to the context presented, the objective of this article is to present the neo-learning, as a methodological support of Networked Corporate University model, to facilitate the transfer of learning to practice. To achieve this objective, a qualitative approach with exploratory and descriptive purposes was chosen, through a narrative literature review and multiple case studies.

2. Theoretical Foundations

This section presents the results of the literature review on the Networked Corporate University model and the neo-learning methodology.

2.1. Networked Corporate University Model (NCU)

From a historical point of view, Margherita and Secundo (2011) and Freire et al. (2016) explain that the evolution of the Corporate Education System includes six different stages: Training Department, E-learning Platform, Corporate Education, Corporate University (CU), Stakeholders University (SU), and Networked Corporate University (NCU).

The last stage, the Networked Corporate University, was developed by Freire et al. (2016). Comparing this model with the previous models, it stands out for expanding the focus of its education programs beyond the development of human and social capital, also covering relational capital.

In this model, learning is seen as a strategic organizational process, since it includes the integration of different areas and stakeholders of the organization, adopting collaboration and networking as a new paradigm (Freire et al., 2016; Margherita and Secundo, 2011). The organizational learning strategy is aligned with the organization's strategy and objectives, and knowledge creation is based on the relationships between the different stakeholders in the network. There is extensive use of new Knowledge Management (KM) technologies, methods, techniques, and tools to expand and develop dynamic innovation capabilities.

Besides, the NCU can be considered a strategic model to support Knowledge Governance and Organizational Learning Governance, and the development of new competencies required by the transformations of the global context (Bresolin, Freire & Silva, 2021; Freire et al., 2016).

The NCU model is defined as an intelligent environment of continuous learning, physical or virtual, that manages and institutionalizes networked learning (Freire et al., 2016), in addition to mobilizing and improving competent performance at multiple levels. In its application, the NCU is established as a business unit that promotes, through networking, continuous, experiential, and expansive learning of essential knowledge for the competent performance of the organizational ecosystem.

Therefore, the NCU is based on the theories of Social Constructivism (Vygotsky, 2007) and Organizational Learning (Crossan et al., 1999; Crossan et al., 2011), Andragogical (Knowles et al., 2015) and Heutagogical (Blaschke, 2012) approaches and takes into account the Theory of Experiential Learning (Kolb, 1984; Kolb & Kolb, 2017) and Expansive Learning Theory (Engeström, 1987; Engeström & Sannino, 2010), in addition to using methods, techniques and tools of Knowledge Management (Nonaka & Takeuchi, 2008).

Social Constructivism (Vygotsky, 1978) offers the methodological basis for the construction of a learning environment centered on communication, collaboration, commitment, and awareness of the social arrangements and constituents of the organization's ecosystem. This theory emphasizes the importance of human experience in the learning process, which establishes mutual interfering relationships in the environment, as well as the environment that interferes with the development.

In this perspective, teaching and learning methods and strategies are based on the principles of andragogical and heutagogical approaches, and on theories of experiential learning and expansive learning. These approaches and theories consider the adults' prior knowledge and bring their experiences and work context into the learning environment, while expanding the learning object to encompass the organization's learning network (Blaschke, 2012; Engeström, 1987; Engeström & Sannino, 2010; Knowles et al., 2015; Kolb, 1984; Kolb & Kolb, 2017).

Regarding multi-level learning, Crossan et al. (2011) consider Organizational Learning (OL) theory as a phenomenon that occurs in levels, so that organizational learning is composed of three levels (individual, group, and organization), and four processes, called 4Is (intuiting, interpreting, integrating, and institutionalizing).

To this end, the NCU model encourages the creation of networked learning, which involves all the organization's internal and external stakeholders. For learning to take place at all three levels, the knowledge created and shared must be managed to expand the reach of the target audience and promote learning for all participants (Freire et al., 2016).

Knowledge Management (KM) methods, practices, techniques, and tools (Nonaka & Takeuchi, 2008) are used for the creation, transmission, and enhancement of inter-organizational and intra-organizational knowledge (Rhéaume & Gardoni, 2015), in order to contemplate the cycle of knowledge and the combination of tacit and explicit knowledge (Antonelli et al., 2013; Nonaka and Takeuchi, 2008). Education programs use different KM methods, practices, techniques, and tools to avoid the loss of knowledge and promote the construction of organizational memory, retain knowledge, and compose a knowledge observatory to identify gaps in knowledge and skills to be developed by NCU to generate knowledge assets.

In addition to the theoretical foundation, the model is composed of eight guidelines (Knowledge Management, Knowledge Governance and Organizational Learning Governance, Reach, Stakeholder Recognition, Interconnection, Educational Technology, Focus, and Level) that guide training, development, and education programs (T, D & E) and support individual, group, intra-organizational, and inter-organizational learning processes (Bresolin, Freire & Silva, 2021; Freire et al., 2021). The model also describes the macro-process for its implementation, composed of nine distinct, but sequential, interactive, and complementary steps, called: the model also describes the macro-process for its implementation, composed of nine distinct, but sequential, interactive, and complementary steps, called: reception; selection; prioritization and proposition; planning; pre-production and logistics; execution; certification; multi-level evaluation (Freire et al., 2021).

During the planning, execution, and evaluation steps, the NCU proposes the application of neo-learning as a methodological support to teaching and learning, in order to meet the cultural and socio-emotional transformations imposed by the skills of the new digital generation and to distribute responsibility and autonomy among the educational triad, learner-teacher-university (Bresolin, Freire & Silva, 2021).

The neo-learning methodology is considered a new motivating mechanism for the learning process of young adults, and its transfer into practice. The methodology establishes guidelines and strategies for the use of active and agile methods, through the application of andragogical, heutagogical, experimental, and expansive principles. These theories provide the basis to promote an environment of dialogue, exchange of roles between teachers and learners, and sharing of knowledge for collaboration and co-production of the educational triad and the stakeholders of the learning network. Furthermore, it provides elements that facilitate multi-level planning, effective execution, and evaluation levels of learning trails, which are crucial for closing knowledge gaps essential to organizational development (Bresolin, Freire & Pacheco, 2021).

2.2. Neo-learning Methodology

The Networked Corporate University (NCU) model needs a methodology consistent with its theoretical foundations to support the teacher in the planning, execution, and evaluation of the learning experience and the education programs, in a way that respects the characteristics of digital native learners, motivating them to take a leading role in the learning process and to develop their ability to transfer learning into social and work practice.

Thus, neo-learning emerges as a methodological teaching and learning platform of the NCU. It represents an innovative way of creating learning environments that facilitate the development of competences and the ability to transfer learning into practice. Neo-learning provides a theoretical and practical foundations for applying andragogical (Knowles et al., 2015) and heutagogical (Blaschke, 2012) principles and the theories of Experiential Learning (Kolb, 1984; Kolb & Kolb, 2017) and Expansive Learning (Engeström, 1987; Engeström & Sannino, 2010).

From the studies on andragogy and heutagogy, emerged the categorization of the 6S of neo-learning (self-direction, self-identity, self-concept, self-motivation, self-management, and self-determination) that need to be managed during the macro-processes of planning, executing, and evaluating education programs for the development of the digital native learner (Bresolin, Freire & Pacheco, 2021).

The Experiential Learning Theory provides neo-learning with the basis for the process of learning by experience, identifying the learners' different learning styles and the different roles played by teachers. For, according to Kolb's (1984) perspective, individuals are able to learn from their experiences, and these experiences are the basis for knowledge creation. The experiential learning cycle is composed of four modules called: concrete experience, reflective observation, abstract conceptualization, and active experimentation. Thus, for this theory, knowledge creation results from the combination of experience that is grasped and transformed by the learner throughout the cycle (Kolb & Kolb, 2017).

The contribution of Expansive Learning Theory to neo-learning is found in the expansion of the learning object to include the relationships between the university context and the context of practical application because learning goes beyond the classroom environment and expands into learning networks (Engeström, 2002).

The expansive learning cycle is formed by a sequence of seven actions: questioning some aspects of the practice; analyzing the situation; modeling a new solution; examining and testing the new model; implementing the new model; reflecting about the process; and evaluating the consolidation process in the new practice, so that it becomes stable and belongs to the new system. This implies the formation of a theoretical concept of the new activity, based on the apprehension and modeling of an initial simple process, which gives rise to the new activity and generates its various concrete manifestations (Engeström & Sannino, 2010).

The theory of expansive learning subsidizes networked learning, proposing co-production projects and complex problem solving, together with the network stakeholders, both internally (between units or between different areas and departments) and externally (with other universities, institutions, organizations, or society).

Besides the theoretical elements, neo-learning proposes constitutive elements that guide multi-level planning and evaluation; development of learning and knowledge trails; interdisciplinary composition of the curriculum; application of active and agile methods; selection of teaching and learning strategies; choice of educational technologies; creation of collaborative and co-productive knowledge environments; different roles played by the teacher, and; respect and consideration for the learner's characteristics to contemplate the different learning styles.

3. Research Methodology

This research is considered to have a qualitative approach (Creswell, 2010), as the objective is to present neo-learning, as a methodological support to the Networked Corporate University model, to facilitate the transfer of learning into practice. The research is classified as exploratory and descriptive (Triviños, 2015), using bibliographic research and multiple case studies of the application of neo-learning. It is exploratory because it addresses the theoretical foundations of the Networked Corporate University model and the neo-learning methodology. It is also considered descriptive, because it describes the neo-learning cycle, which facilitates the transfer of learning, and the application cases, which validate the methodology.

Regarding the methodological procedures, a narrative literature review was chosen on the theories and approaches that underlie the Networked Corporate University model and the neo-learning methodology. The multiple case studies were chosen because it constitutes, according to Yin (2014), an empirical research strategy that allows the understanding of a phenomenon in its context. For this purpose, a summary of the cases of neo-learning application is presented, validating the elements and cycle of the methodology since 2018.

Data collection and analysis were conducted in academic and corporate universities that applied the methodology. The authors participated in the planning and execution of the application of neo-learning to the cases presented, recording in a field diary the applications, best practices, and lessons learned. To validate neo-learning, assessments were also carried out on the satisfaction of teachers and learners with the methodology applied, recording of feedback from learners and teachers, interviews with teachers and tutors who were capacitated to apply the methodology, and verification of the competences developed by the learners.

4. Neo-learning Methodology Applied to the NCU Model

This section discusses the neo-learning cycle that can be applied in education programs, courses, classes, training, and development of learners, to facilitate the transfer of learning into practice. It also presents some cases of application of the neo-learning methodology.

4.1. Neo-learning Cycle to Facilitate Learning Transfer

The andragogical (Knowles et al., 2015) and heutagogical (Blaschke, 2012) approaches and the theories of experiential learning (Kolb, 1984; Kolb and Kolb, 2017) and expansive learning (Engeström, 1987; Engeström & Sannino, 2010), provide the theoretical foundations for the different moments in a class, that is called the neo-learning cycle. The cycle consists of five modules, which are integrated, dynamic, and fluid, for the development of technical, digital, and socio-emotional competences of digital learners. The five modules of neo-learning cycle are nominated: retrieve, reflect, know, test, and apply, as Figure 1 demonstrates.

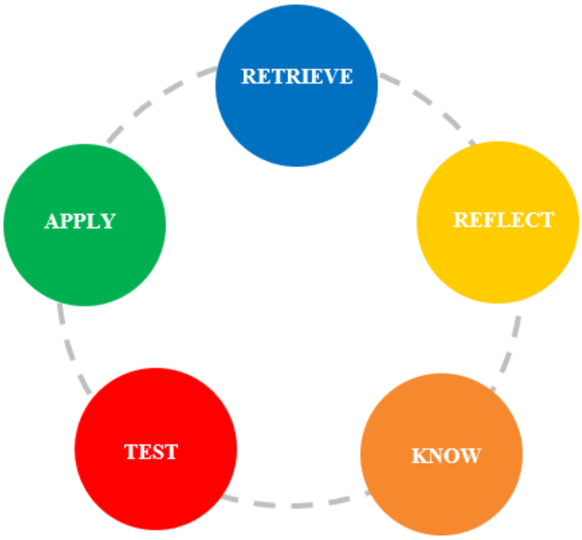


Figure 1. Neo-learning cycle. Prepared by the authors (2022).

The description of the principles and application guidelines for each module of neo-learning cycle is presented in Table 1.

Table 1. Principles and guidelines of the neo-learning cycle

Module	Principles	Application guidelines
Retrieve	The simulated experiential activities recover the learner's previous knowledge and experiences, providing a confrontation with what they already know about the object to be studied	Value, respect, recognize, and use the previous experience of learners in the learning process; Create meaning for learning from existing and learned knowledge and mental processes; Encourage questioning about some aspects of existing practice and knowledge and possible applications; Promote divergent knowledge so that the learner perceives the different types of possible solutions to the same problem
Reflect	An open space for reflection, observation, discussion, and analysis on the lived experience, without criticism and without determination of right and wrong, provides the necessary freedom to identify the gaps to be filled and the motivation to discover something new	Provide space for learners to share their experiences and express their thoughts and reflections on the object of study; Promote moments of reflection, individually and then in groups, on shared experiences; Ask questions, encourage questioning, generate discussions, and share opinions about events (experienced in the previous module); Develop the ability of learners to observe their experiences from different perspectives to develop a critical reflection on aspects of the context; Highlight elements of practice associated with the fundamentals and theories that will be worked on later, without judging right and wrong; Motivate for what is still to be known
Know	Understanding concepts, theories, and models provide the background for generalizing knowledge to be applied in different experiences	Request the research of concepts, theories, and models on the subject to be worked on, in bibliographic and documentary references; Present the theoretical foundations; Relate the theories with the experiences and analysis shared in the previous modules; Encourage logical and justified conclusions about the experience; Facilitate the development of assimilative knowledge, using deduction to solve problems, making use of the logic of an idea to create theoretical models, organize ideas, and transform them into coherent explanations
Test	Experiential activities for practical application of knowledge make it possible to test new ideas and concepts in situations closer to reality	Create group activities to apply the concepts, theories, and models learned to case studies, simulations, workshops, or laboratory experiments; Promote space that facilitate the development of creativity and the generation of multiple ideas, during the attempts to apply knowledge in the search for solving problems and testing solutions; Promote the creation of convergent knowledge, allowing the finding of a more appropriate solution to a specific problem, with collective decision-making

Apply	Experiential practices, supervised or monitored, provide the opportunity to expand and overcome the controlled environment and elevate learning to social and work practice	Invite experts, in the area of application of the knowledge worked, to participate as teachers in this module; Create learning networks between learners and collaborators related to the practice environment, breaking the boundaries of the simulated learning environment; Expand the learning object to the relationships between the simulated context and the practical application, projecting the application to the possible social or professional practices; Allow observation, reflection, and evaluation of the consolidation process of new knowledge in practice; Encourage the development of new models for the practice experienced, so that new knowledge can be institutionalized; Create space for discussion after the practical experience, so that it is possible to accommodate and internalize the new knowledge; Question the lessons learned and good practices to be shared from the experience

Source: Prepared by the authors (2022), based on Engeström (1987), Engeström and Sannino (2010), Kolb (1984), Kolb and Kolb (2017).

These principles and guidelines assist in the planning and execution of classes, courses, and education programs aligned with neo-learning for the human and professional development of digital native learners. In addition, they support the creation of learning environments that encourage discovery, collaboration and coproduction with the network, the testing and application of knowledge, considering the learner's previous knowledge and motivation to learn, apply, and transfer.

Transfer of learning is a critical factor for any education program. Learners must acquire knowledge, skills, and attitudes and then apply them in a context other than the acquisition context. Transfer reflects the effectiveness of the learning process and application of knowledge in the organization (Baldwin and Ford, 1988; Ford et al., 1998). It demonstrates the extent to which learning can be applied effectively and continuously in the work context (Fauth & González-Martínez, 2021).

In this regard, Dolmark et al. (2021) point out that learners obtain knowledge through multiple pathways, such as observation, experience, or acquisition. Individuals also acquire knowledge due to the difficulties they encounter, in their personal or professional lives, and the challenges they intend to solve. Consequently, individuals choose their learning path, and the degree of assimilation and understanding of learning is related to the absorptive capacity to confront these challenges.

Also, according to Takeuchi and Nonaka (2008), both organizations and universities need to develop methodologies that enable the creation of knowledge, not only acquired through texts, standards, books, manuals, dialogues, and knowledge sharing or other materials made available, known as explicit knowledge. It is also necessary to create tacit knowledge using active methods and strategies focused on the application of knowledge in practice to develop skills. For, according to the authors' perspective, knowledge is created through the interaction between tacit and explicit knowledge.

The different methods and strategies used during the learning process provide opportunities for practice (Blume et al., 2010) and directly influence the acquisition of skills and their subsequent application (Garrido et al., 2017), as well as learning active and learner-centered learning favors effective transfer of learning (Dowson, 2019; Garrido et al., 2017). Providing an improvement in learning and a better adaptation and application in the work context (Fauth & González-Martínez, 2021).

Therefore, for each module of the neo-learning cycle, it is suggested the use of different teaching and learning strategies, the support of educational technologies, and the application of active and agile methods to innovate in experimental and expansive practices. These suggestions facilitate the transfer of learning into practice, because learners will be more motivated to apply the acquired competences in their social and professional practice.

In this perspective, the transfer of learning is related to the ability to transfer and apply, in an effective and continuous way, the knowledge, skills and attitudes acquired, consolidating a competent performance in their work practice. Thus, the purpose of creating the neo-learning cycle is to facilitate the immediate application of learning into practice, because each module includes different activities that promote learning and, specifically, the last two modules allow testing and application in an environment, simulated or closer to the reality, that will be found at the work environment.

4.2. Applications of the Neo-learning Methodology

In some cases, are evident, where education programs, courses, and classes planned and executed based on the neo-learning methodology have allowed the development of competences for a creative but also resolutive professional career, with more pragmatic and applied thinking. Moreover, involve learners in the co-production of knowledge and learning, because teaching and learning practices in the network are based on the theory of experiential and expansive learning, which considers their previous knowledge and brings experiences from the work context into the learning environment.

The neo-learning methodology emerged to attend to this purpose. And, therefore, it has been developed since 2015 by the Laboratory of Integration of Engineering and Multi- level Governance of Knowledge and Organizational Learning of the Graduate Program in Knowledge Engineering and Management at the Federal University of Santa Catarina (ENGIN/EGC/UFSC) in Brazil. The methodology has been validated in empirical research since the year 2018, both in undergraduate and graduate courses, in research and extension projects, in the development of young professionals in Networked Corporate Universities. The neo-learning was applied for the first time in the pilot project named Alexandria.

The pilot project was developed by the Federal University of Santa Catarina (UFSC), between 2018 and 2019, with the participation of universities, institutions, and organizations that belong to the innovation ecosystem of the city of Florianópolis, in Brazil. This initiative aimed to promote co-production among the different stakeholders of the innovation ecosystem, in the creation of human capital and innovative and sustainable ventures, providing opportunities for the development of technical, digital, and socio-emotional skills.

In this pilot project eight graduate teachers were capacitated to apply the neo-learning cycle to the development of fifty digital native learners. These learners were divided into ten groups composed of five learners. Each group represented an organization and had to develop sustainability solutions and a product that would generate an impact business. At the end, rounds of Pitch presentations were conduct for continuous improvement of the solutions until the final presentation. The groups with the most innovative solutions had their projects validated by the organizations for future applications and adaptations. As a result, teachers and learners who participated in the project approved the methodology applied. With feedback it was possible to improve some elements of the neo-learning and create the neo-learning canvas (Bresolin, Freire & Pacheco, 2021) to facilitate the application of the cycle and elements in the class.

Also, since 2018 neo-learning has been applied in two undergraduate disciplines of Idea Generation and Entrepreneurship, and Creativity and Innovation at UFSC. Since 2020, due to the COVID-19 pandemic, these two disciplines were also transferred to the remote modality. Because of this, it was possible to validate the methodology in both, presential and remote modalities. The disciplines also enabled the application of the neo-learning cycle related to the innovation process, to promote innovation and entrepreneurship projects arising from real problems experienced by undergraduate learners.

The Graduate Program in Knowledge Engineering and Management of the Technology Center in the Federal University of Santa Catarina developed, during 2020 and 2021, a committee with the participation of teachers, learners, and guests, with the purpose of developing the new curricula of the master's and doctoral degrees, aligned with the new competencies for digital transformation. To this end, neo-learning is the methodology adopted by the program to innovate in class practices and promote research and extension projects in partnership with the organizations in the ecosystem, enabling the creation of theses and dissertations with practical applications for solving problems in society and organizations.

The neo-learning methodology was applied to specific disciplines of an initial training course for the development of more than 650 new professionals in a public security organization in Brazil. It was also applied to an online course on the Sustainable Development Goals (SDGs) of the 2030 Agenda for development of municipal agents. Specifically, in this course, twelve tutors were also capacitated to develop activities and apply the neo-learning in the online learning environment.

The results of data collection and analysis of these different applications confirmed the satisfaction of teachers and learners with the neo-learning methodology, especially with the application of the neo-learning cycle. All the cases presented enabled the application and validation of the methodology in public and private organizations, Networked Corporate Universities, and academic universities, developing teachers and tutors to act as disseminators of neo-learning. In addition, several learners were developed based on the methodology, being able to transfer the learning to their practices.

5. Conclusion

In summary, the Corporate Education System is challenged to expand to reach the stage of a Networked Corporate University, focusing on the development of people to self-manage their professional trajectory aligned with the organizational development. This requires the application of the neo-learning methodology for planning and executing experiential and expansive learning into organizational outcomes.

This paper presented the neo-learning methodology and the main approaches and theories that support it. It also presented the neo-learning cycle and the cases of applications and validation. Thus, it can be observed that the neo-learning methodology was developed in accordance with the Networked Corporate University (NCU) model to support the teaching and learning processes in a way that enables the transfer of learning into practice. The application of neo-learning involves learners in solving problems that are part of their routines, and they analyze, share, discuss, and reflect on their actions. Moreover, neo-learning brings significant improvements in the learners' memory, develops autonomy, responsibility, and self-determination about the learning process, and proposes group work, collaborative, and co-production environments for the development of competences.

With the research and applications of the methodology carried out so far, it has been possible to validate the elements and cycle of neo-learning in undergraduate, graduate, and initial training courses, both in academic and corporate universities, public or private, in the presential, online, and remote modalities. The methodology has also been applied to the training of tutors of online courses and of teachers of presential courses. The applications confirmed the satisfaction of teachers and learners with the methodology, especially with the application of the neo-learning cycle, obtaining better learning results than the traditional methodologies.

For future work, it is suggested continued practical applications, in various areas of knowledge, in public and private organizations, and empirical research to validate the cycle, the elements, and the competencies developed with the application of the neo-learning methodology. Also, the validation of neo-learning for programs, courses, and classes in hybrid models of teaching and learning.

6. Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

References

Antonelli G., Cappiello, G., & Pedrini, G. (2013). The corporate university in the European utility industries. *Utilities Policy, 25*: 33–41.

Baldwin, T. T., & Ford, J. K. (1988). Transfer of training: A review and directions for future research. *Personnel Psychology, 41*: 63-105.

Blaschke, L. M. (2012). Heutagogy and lifelong learning: A review of heutagogical practice and self-determined learning. *The International Review of Research in Open and Distributed Learning, 13*(1): 56-71.

Blume, B. D., Ford, J. K., Baldwin, T. T., & Huang, J. L. (2010). Transfer of Training: A Meta-Analytic Review. J*ournal of Management, 36*(4), 1065–1105. https://doi.org/10.1177/0149206309352880

Chen, Y., Xu, Y., & Zhai, Q. (2019). The knowledge management functions of corporate university and their evolution: case studies of two Chinese corporate universities. *Journal of Knowledge Management, 23*(10): 2086-2112.

Creswell, J. W. (2010). *Projeto de pesquisa: Métodos qualitativo, quantitativo e misto.* 3. ed. Porto Alegre: Artmed.

Crossan, M. M., Lane, H. W., & White, R. E. (1999). An organizational learning framework: From intuition to institution. *Academy of Management Review,* (24): 522–537.

Crossan, M. M., Maurer, C. C., & White, R. E. (2011). Reflections on the 2009 AMR decade award: do we have a theory of organizational learning? *Academy of Management Review, 36*(3): 446–460.

Bresolin, G. G., Freire, P. de S., & Pacheco, R. C. dos S. (2021). *Neoaprendizagem, 10 passos para a prática andragógica, experiencial e expansiva*. Florianópolis: Editora Arquétipos.

Bresolin, G. G., Freire, P. de S., & Silva, S. M. (2021). Maturity Criteria of the Networked Corporate University Model to support Knowledge and Organizational Learning Governance. In: Freire, P de S., Hawamdeh, S., & Dandolini, G. A. (Eds.). *Knowledge Governance and Learning for Organizational Creativity and Transformation.* Singapore: World Scientific Publishing Company. https://doi.org/10.1142/9789811224119_0019

Dolmark, T., Sohaib, O., Beydoun, G., & Wu, K. (2021). The Effect of Individual's Technological Belief and Usage on Their Absorptive Capacity towards Their Learning Behaviour in Learning Environment. *Sustainability, 13*(2): 718. https://doi.org/10.3390/su13020718

Dowson, J. (2019). Transferring knowledge into practice? Exploring the feasibility of action learning for improving knowledge, skills and confidence in clinical communication skills. *BMC Med. Educ. 19*(37). https://doi-org/10.1186/s12909-019-1467-4

Engeström, Y. (2002). Non scolae sed vitae disimus: Como superar a encapsulação da aprendizagem escolar. In: Harry, D. (Org.). *Uma introdução a Vygotsky*. São Paulo: Edições Loyola.

Engeström, Y. (1987). *Learning by expanding: an activity-theoretical approach to developmental research.* Helsinki, Orienta-Konsultit.

Engeström, Y., & Sannino, A. (2010). Studies of expansive learning: Foundations, findings and future challenges. *Educational Research Review, 5*:1-24.

Fauth, F., & González-Martínez, J. (2021). On the Concept of Learning Transfer for Continuous and Online Training: A Literature Review. *Education Sciences, 11*(3):133. https://doi.org/10.3390/educsci11030133

Freire, P. de S., Dandolini, G. A., Souza, J. A., & Silva, S. M. (2016). Processo de implantação da Universidade Corporativa em Rede (UCR). *Espacios, 37*(23): 01-21.

Freire, P. de S., Silva, T. C., & Bresolin, G. G. (2021). *Gestão da Universidade Corporativa em Rede.* Florianópolis: Editora Arquétipos.

Ford, J. K., Smith, E. M., Weissbein, D. A., Gully, S. M., & Salas, E. (1998). Relationships of goal orientation, metacognitive activity, and practice strategies with learning outcomes and transfer. *Journal of Applied Psychology, 83*(2) 218-233. doi:10.1037/0021-9010.83.2.218

Garrido, M. F., Davids, A. I. R., González, J. M. J., & Soto, Á. G. (2017). Análisis sobre el aprendizaje y la aplicación de las competencias generales em el contexto laboral. Estrategias de colaboración entre la formación professional, la universidad y la empresa. *Educar, 53*(2): 333-355. http://dx.doi.org/10.5565/rev/educar.889

Knowles, M. S., Holton III E. F., & Swanson, R. A. (2015). *The Adult Learner: The definitive classic in adult education and human resource development.* Routledge.

Kolb. D. A. (1984). *Experiential learning.* Englewood Cliffs. New Jersey, Prentice Hall.

Kolb, A. Y., & Kolb, D. A. (2017). Experiential Learning Theory as a Guide for Experiential Educators in Higher Education. *A Journal for Engaged Educators*: 7-44.

Margherita, A., & Secundo, G. (2011). The stakeholder university as learning model of the extended enterprise. *Journal of Management Development, 30*(2): 175-186.

Rhéaume, L., & Gardoni, M. (2015). The challenges facing corporate universities in dealing with open innovation. *Journal of Workplace Learning, 27*(4): 315–328.

Takeuchi, H., & Nonaka, I. (2008). *Gestão do conhecimento.* Porto Alegre, Bookman.

Triviños, A. N. S. (2015). I*ntrodução à pesquisa em ciências sociais: a pesquisa qualitativa em educação*. (1 ed.)23. São Paulo: Atlas.

Yin. R. K. (2014). *Case Study Research: Design and Methods.* (5th ed.). Thousand Oaks, CA: Sage.

Vygotsky, L. S. (1978). *Mind in society: the development of higher psychological processes.* Cambridge, MA: Harvard University Press.

Evidence-driven strategies for successful schooling outcomes of learners of african descent in Nova Scotia

RAAVEE KADAM
Saint Mary's University, Postdoctoral Fellow,
raavee.kadam@smu.ca,
0000-0002-7010-2999

JOYLINE MAKANI
Dalhousie University, Assistant Professor,
Joyline.Makani@Dal.Ca,
0000-0002-8796-3648

GEORGE FREMPONG
Delmore "Buddy" Daye Learning Institute, Director of Research,
george.frempong@dbdli.ca

PAWAN LINGRAS
Saint Mary's University, Professor,
pawan@cs.smu.ca, 0000-0002-2149-7242

SUMMARY

Goal: *In Nova Scotia, the marginalization of African Nova Scotian learners in the public education system has led to an achievement gap, where their performance in terms of literacy and numeracy in annual assessments is often below the provincial average. This research aims to use machine learning to analyze student data collected by the Nova Scotia Department of Education and Early Childhood Development to provide evidence-driven insights into the schooling outcomes of African Nova Scotian students. We expect that these insights would enable the development of an equitable education system where African Nova Scotian students are able to achieve their maximum potential and overcome the barriers that they face at school.*

Design / Methodology / Approach: *Through collaboration with the Regional Centers of Education, the Nova Scotia Department of Education and Early Childhood Development has been collecting data to track students' education experiences. We aim to analyze this data using machine learning algorithms to identify learning patterns and promising school characteristics that would provide an understanding of how we can reduce the achievement gap that exists among African Nova Scotian students.*

Results: *This research is work in progress.*

Practical implications: *This research is important for parents, educators, policy makers, school counselors, community leaders, and other professionals who work to maximize the educational experiences of African Nova Scotian students. We expect our findings to inform evidence-based policy decision-making and implementation in Nova Scotia schools.*

Originality / value: *We take the novel approach of using machine learning algorithms within social science research to provide evidence-driven insights that can inform policy-making.*

Keywords: educational experiences. African Nova Scotians. machine learning. data analytics.

INTRODUCTION

Extant research has identified an achievement gap amongst Black Canadian students as compared to their peers. Codjoe (2007) states that this issue of disproportionate achievement of Black Canadian learners has garnered a lot of interest in educational research as researchers attempt to understand the underlying reasons for the same. Extant research has also established the fact that Black Canadian students struggle in public schools (Shizha, 2016). Factors identified include absence of Black teachers and top school administrators, an Eurocentric approach to school curriculum that lacks representation of African perspectives, social labeling of Black Canadian students as trouble-makers, inherent racism, identity concerns for students born to mixed parents, being stereotyped as foreigners with cultures that do not value education, perceptions of a dysfunctional family structure, etc. (Dei, 1995; Henry, 1993; James, 2012; James & Turner, 2017), that have created barriers to the achievement of positive learning outcomes in the students. Consequently, researchers are consistently identifying different ways by which the pattern of school failures among Black students can be addressed.

In Nova Scotia, the marginalization of African Nova Scotian learners in the public education system has led to an achievement gap, where their performance in terms of literacy and numeracy in annual assessments is often below the provincial average (Province of Nova Scotia, 2015). African Nova Scotians constitute the largest multigenerational Black Canadian community, with 400 years of presence in Atlantic Canada (Jean-Pierre, 2021). African Nova Scotians make up the largest racially visible group in Nova Scotia, with 4.2% of P-12 students self-identifying themselves as being of African descent. Despite several initiatives in place, the annual provincial assessment of African Nova Scotian students' academic performance indicates a significant and persistent achievement gap. These results demonstrate the challenges that schools and the Nova Scotia education system in general is facing in helping a substantial number of ethnic and culturally diverse students to reach their potential. Through collaboration with the Regional Centers of Education, the Nova Scotia Department of Education and Early Childhood Development (EECD) has been collecting data to track students' education experiences. We argue that a detailed data analysis will provide insights into understanding the schooling outcomes of African Nova Scotian students and what kind of data needs to be collected in the future to continue with the monitoring and measuring of learning outcomes. Taking a machine learning approach, we will analyze the EECD data to identify learning patterns and track promising school practices - practices that provide opportunities for all students, irrespective of their cultural backgrounds to succeed. This research is important for parents, educators, policy makers, school counselors, community leaders, and other professionals who work to maximize the educational experiences of African Nova Scotian students.

DEVELOPMENT

Cummins (2001) highlighted two reasons why educational reforms implemented to improve academic performance among culturally and linguistically diverse students are not yielding the results as expected: "(a) empirical data relating to the patterns of educational underachievement that challenge the current ideological mindset are systematically ignored or dismissed; (b) there is a deep antipathy to acknowledging that schools tend to reflect the power structure of the society and that these power relations are directly relevant to educational outcomes" (p. 650). He contends that current reform efforts selectively focus on research that links individual student characteristics to their underachievement while ignoring much stronger relationships identified between student achievement and social & educational inequities (e.g., unequal distribution of school funds). He further contends that the assumption on which these reforms are based is that teaching or instructional interventions can remediate for the incompetencies of students and do not take into consideration the larger prevailing socially unjust practices. Hence, focusing on data that can identify such patterns becomes important. As Banks (2009) noted though evidence and analysis can never fully play a decisive role in informing policy-makers' judgements, they can however condition the environment in which those judgements need to be made. Such evidence can provide a reference for policy makers when policies are being designed. Thus, analyzing the data systems in place and the application of machine learning will provide a further understanding of the link between policy initiatives and student achievement.

Education for Sustainable Development

We utilize the Education for Sustainable Development (ESD) as an overarching theoretical framework for our study. The focus of ESD is to build a sustainable system that integrates the principles, values, and practices of sustainable development into all aspects of education and learning (UNESCO, 2012). ESD recommends the consideration of societal and cultural aspects while designing learning environments to create a more sustainable future and a just society for present and future generations. According to Tilbury (2011) “ESD gives attention to not only specific learning approaches and techniques used within education but also to the professional and management processes adopted across educational systems themselves” (p.16). This means that ESD takes a holistic approach focusing not only on curriculum or pedagogy, but even on the processes and data systems in place in educational institutions. This research study will enable us to “inform and inspire future policy and programming to help bridge links between cultural, socio-economic and environmental learning for sustainability” (Tilbury & Mulà, 2009, p. 7).

Data and Analysis

Through the collaboration with the Regional Centers of Education (RCE), the EECD has been collecting data to track students’ education experiences. The data includes:

- students’ demographic characteristics
- students’ performance in literacy and mathematics
- the designation of students in special education programs
- students’ attendance, engagement and suspension records
- school characteristics
- teacher characteristics
- others

The data also includes the identification of students’ cultural background characteristics that provides opportunity to assess the achievement gap in relations with other education experiences. We are particularly interested in the link between the achievement gap and schooling experiences of African Nova Scotian students.

The research team will be accessing EECD data in conjunction with a data-sharing agreement made with the EECD to ensure privacy and confidentiality. The researchers will not have access to any identifying information in the database. The EECD data specialists will be taking measures to de-identify the data. i.e., remove identifying information and apply numerical identifiers to ensure the student’s and school’s anonymity and confidentiality is maintained.

Machine Learning

Machine learning algorithms allow you to automatically identify relationships between data and show the best patterns based on that data, without having to be explicitly programmed (Witten & Frank, 2005). Identifying such learning patterns that ensure academic success is critical because they will enable schools to understand the deviations in the learning patterns of students with low academic success. Best patterns identified can then be emulated in the school environment to ensure successful learning outcomes for all students. To identify such patterns that are critical to the positive schooling outcomes of African Nova Scotian students, an experiment will be designed and unsupervised machine learning will be applied to the data (Lingras & Peters, 2011). Unsupervised machine learning algorithms infer patterns from a dataset when there are no output categories available on which the algorithm can try to model relationships. These algorithms try to use techniques on the input data to mine for rules, detect patterns, and summarize and group the data points, which help in deriving meaningful insights and describe the data better to the users (Fumo, 2017). Firstly, temporal unsupervised learning will be used to create profiles of learner using the data in hand. Temporal unsupervised learning will help us divide students into different groups based on the kind of family-school-community partnerships they have experienced, that have impacted their learning outcomes (Lingras et al., 2011). Unsupervised machine learning algorithms will then be used to identify the learning patterns amongst the groups. This exercise will enable us to identify the best partnership-related strategies that ensure positive schooling outcomes amongst students.

FINAL REMARKS

This research study focuses on policy implementation and education achievement gap issues of African Nova Scotian students that we are studying in conjunction with the EECD. We intend to be able to use the findings of this research to provide a basis for policy analysis that informs policymakers, educators, parents and the general public on how to best improve educational opportunities and outcomes for Nova Scotian learners of African descent. This research will augment the existing literature on the schooling outcomes of Black Canadians and provide data-driven insights that would further enable policy makers in Canada to implement appropriate policies to improve the academic success of Black Canadian learners.

REFERENCES

Banks, G. (2009). Evidence-based policy making: What is it? How do we get it? (ANU Public Lecture Series, presented by ANZSOG, 4 February), Productivity Commission, Canberra.

Codjoe, H. M. (2007). The Importance of Home Environment and Parental Encouragement in the Academic Achievement of African-Canadian Youth. *Canadian Journal of Education , 30* (1), 137-156.

Cummins, J. (2001). Empowering minority students: A framework for introduction. *Harvard Educational Review*, 71 (4), 649-655

Dei, G. J. (1995). Examining The Case For African-Centered Schools In Ontario. *Mcgill Journal Of Education, 30* (2), 179-198.

Fumo, D. (2017). *Types of Machine Learning Algorithms You Should Know*. Retrieved July 14, 2019, from https://towardsdatascience.com/types-of-machine-learning-algorithms-you-should-know-953a08248861

Henry, A. (1993). Missing: Black Self-Representations in Canadian Educational Research. *Canadian Journal of Education , 18* (3), 206-222.

James, C. E. (2012). Students “at Risk”: Stereotypes and the Schooling of Black Boys Urban Education. *Urban Education , 47* (2), 464-494.

James, C. E., & Turner, T. (2017). *Towards Race Equity In Education: The Schooling of Black Students in the Greater Toronto Area.* Toronto, Ontario, Canada: York University.

Lingras, P., Hugo, M., Snorek, M., & West, C. (2011). Temporal analysis of clusters of supermarket customers: conventional versus interval set approach. *Information Sciences, 172* (1-2), 215-240.

Lingras, P., & Peters, G. (2011). Rough clustering. *WIRES Data Mining and Knowledge Discovery, 1* (1), 64-72. Province of Nova Scotia. (2015). *The 3Rs: Renew, Refocus, Rebuild - Nova Scotia’s Action Plan for Education.* Crown.

Shizha, E. (2016). Marginalization of African Canadian Students in Mainstream Schools: Are Afrocentric Schools the Answer? In A. A. Abdi, & A. Ibrahim (Eds.), *Education of African-Canadian children: Critical perspectives* (pp. 187-206). Montreal: McGill-Queen's Press.

Tilbury, D. (2011). Assessing ESD Experiences During the DESD: An Expert Review on Processes of Learning for ESD, UNESCO, Paris

UNESCO (2012). Shaping the Education of Tomorrow: 2012 Full-length Report on the UN Decade of Education for Sustainable Development. Available at https://library.wur.nl/WebQuery/wurpubs/fulltext/246667

Witten, I. H., & Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques (2nd Edition).* San Francisco, CA: Morgan Kaufmann Publishers.

Track – Knowledge Management III

Evaluation of operational knowledge risks in SMEs – using a grey-dematel technique

SAMUEL FOLI*
Department of Business Administration,
Tallinn University of Technology, Ehitajate tee 5,
19086 Tallinn, Estonia
E-mail: samuel.foli@taltech.ee

SUSANNE DURST
Department of Business Administration,
Tallinn University of Technology Ehitajate tee 5,
19086 Tallinn, Estonia
E-mail: susanne.durst@taltech.ee

ELENA DOMINGUEZ ROMERO
Department of Business Administration,
Tallinn University of Technology,
Ehitajate tee 5, 19086 Tallinn, Estonia
E-mail: elena.dominguez@taltech.ee

Corresponding author*

Structured Abstract

Purpose: Risk management is generally characterized by identification, assessment, and mitigation. In the literature on knowledge risk management (KRM), there appears to be much emphasis on taxonomies - which primarily focus on identification - and mitigation strategies, without particular attention to the evaluation phase of KRM. To address this gap, the present paper aims to systematically evaluate potential knowledge risks common at the operational level of small and medium-sized enterprises (SMEs). More significantly, in SMEs, due to their flat structure, the majority of risk is borne by operational functions.

Design/Methodology/Approach: This study adopts a quantitative approach by utilizing the grey-DEMATEL technique.

Findings: The results of this study reveal outsourcing risks are the most critical of all operational knowledge risks followed by communication risks and improper knowledge application. Furthermore, it is observed the risks could be categorized into effect-risk (e.g. relational risk, espionage knowledge waste, continuity risk) and cause-risk (e.g. knowledge waste, risks related to knowledge gaps, risk of using obsolete/unreliable knowledge)

Originality/Value: To the best of the authors' knowledge, this is one of the first studies to use the grey-DEMATEL technique to evaluate knowledge risks.

Keywords: grey-DEMATEL technique, operational knowledge risk, knowledge risk management, KRM, SMEs

Paper Type: Academic Research Paper

1. Introduction

Disruptions in operations have been cited as a major threat to many organizations, large and small. The Allianz Global Corporate & Specialty (AGCS) reported in the latest Barometer 2022 survey, business interruption as the second most significant business risk after cyber incident. In light of this, there appears to be substantial interest shown by academics and practitioners with respect to this topic as described by Cruz and Pinedo (2008). Al-Hussaini and Karkoulian (2015) point out that risks incurred at the operational level affect reputation, customer satisfaction, and owner value. As a result, managing operational risks serves as a key component of any management strategy (Todinov, 2007). Due to the fact that each and every risk is unique and has its own probability of occurrence and severity, it is critical to perform a rigorous evaluation process in order to identify the important risks that need to be addressed. Analyzing risks holistically can uncover hidden relationships, unique characteristics, and categories they belong to, thus helping develop tailored-made countermeasures that address each risk individually.

Research on knowledge risk management (KRM) is still underdeveloped (Massingham, 2008; Durst *et al.*, 2019). Most studies are of a conceptual nature and have often focused on the identification of knowledge risks and proposal for mitigating these risks (e.g. Durst and Zięba, 2019, 2020; Tsang and Lee, 2018). In this regard, to show the relevance of KRM for small and medium-sized enterprises (SMEs) considering their limited resources (Durst and Ferenhof, 2014), research needs to take next steps and propose methods and tools that can be used to evaluate whether the knowledge risks identified are worth to be addressed due to the costs and time involved in risk management in general (Callahan and Soileau, 2017).

Against this background, the aim of the paper is to enrich the KRM research by comprehensively evaluating potential knowledge risks that SMEs face in their operations. It uses grey-DEMATEL technique by building upon Temel and Durst's (2021) work title "Knowledge risk prevention strategies for handling new technological innovations in small businesses", focusing on some of the operational risks outlined by the authors.

There are few studies that have examined the causal relationship among knowledge risks by using taxonomy (Durst and Zięba, 2018; Temel and Durst, 2020; Hammoda and Durst, 2022) and other types of Multi-Criteria Decision Analysis (MCDA), such as total interpretive structural modeling (TISM) (Foli, 2022) and analytic hierarchy process (AHP) (Tsang *et al*., 2016). For, example, Foli (2022) establishes the inter-and multi-relationship among potential knowledge risks within an ICT-supported collaborative project. The TISM identify the structural relationships among risks, but they do not indicate the prominence of those relationships (Sufiyan *et al*., 2019). By using DEMATEL, we can assess relationships and their strength (Bakir *et al.*, 2018; Biswas and Gupta, 2019; Vishvakarma *et al*., 2019). However, it has some limitations regarding the subjectivity and impreciseness of experts' inputs. By integrating the grey theory with DEMATEL, these limitations can be eliminated (Haleem *et al.*, 2019).

The paper is structured as follows. The next section provides an overview of the knowledge risks SMEs are exposed at the operational level. This is followed by the method used for this study with a clear step-by-step application of the grey-DEMATEL technique. Presentation of the results and discussions follows. Last, the paper terminates with a conclusion and future research avenues.

2. Theoretical background

There is great variation in how risk is perceived according to culture, context, and discipline (Zheng, 2017). The varying understanding makes it difficult to establish a unified definition. In this study, the authors adopt the definition commonly used in discussions related to knowledge management which defines risk as "a measure of the probability and severity of adverse consequences" (Haimes, 2009, p. 1648). Even so, the definition clearly identifies risk as negative, although this might not always be the case. In addition, it shows that risk can be assessed or evaluated in terms of its ability to result in an undesirable outcome. For Manab *et al.* (2020), the importance of risk evaluation cannot be overemphasized when it comes to pursuing business survival and success; therefore, across any organization type, risk assessment is recognized to be a key element of their strategic plan. According to knowledge based-view (KBV) theory, knowledge is the most strategic resource a firm can possess (Grant, 2003), which makes knowledge an enviable asset.

Therefore, when the firm's knowledge is lost or exposed, several consequential effects are experienced, such as loss of reputation and competitive advantage, which often succeed knowledge risk or knowledge being at risk.

Several scholars (Bratianu, 2018; Durst and Zięba, 2020; Massingham, 2010) have demonstrated the connection between the concept of risk and firms' knowledge. According to Durst and Zięba (2020), knowledge risks are classified into three major categories: human, technological, and operational. In their study, the authors clearly showed that knowledge risks under the operational category are much greater than those under the other categories (i.e. human and technological). Therefore, it shows that firms are exposed to several risks related to their knowledge in their operational activities. Specifically, the authors identified twelve distinct types of operational knowledge risks such as knowledge waste, knowledge gaps, relational risks, outsourcing risks, obsolete knowledge risks.

A similar study by Durst and Zięba (2020) explored the impact of knowledge risks on the three dimensions of sustainability in an organization (environmental, economic, and social). Operational knowledge risks accounted for a significant share of risks faced by organizations in the quest for sustainability, for example, knowledge waste, knowledge gaps, communication risks, and risk of knowledge acquisition are among those that have the potential to affect all three pillars of organizational sustainability. A study conducted by Temel and Durst (2021), specifically focused on small businesses, identified, and classified risks associated with new technological innovations, and recommended countermeasures. The study found that at the operational level, common knowledge risks include knowledge waste, knowledge gaps, continuity risks, improper application of risks, and espionage. Thus, the present paper analyzes operational knowledge risks based on Temel and Durst's (2021) work (see Table 1).

Table 1. Definitions of types of knowledge risks at the operational level

Risk code	Knowledge risk	Meaning
K01	Knowledge waste	Loss of strength from reinventing the wheel when not using the available knowledge (Ferenhof *et al.*, 2016)
K02	Risks related to knowledge gaps	Situations in which a lack of knowledge prevents the organization from performing its most vital functions (Perrott, 2007)
K03	Relational risks	Arising from opportunistic behavior by partners (Delerue, 2005)
K04	Outsourcing risks	Caused by the fact that organizations tend to rely too heavily on their outsourced firms, which slows internal knowledge management (Agndal and Nordin, 2009)
K05	Risk of using obsolete/unreliable knowledge	Risks of using stale knowledge which might negatively influence the organisation new way of doing things (Zięba and Durst, 2018)
K06	Risk of improper knowledge application	Risks because of misunderstanding which lead to undesired organizational agenda (Zięba and Durst, 2018)
K07	Espionage	Risks resulting from an unlawfully spying and retrieving organizational critical knowledge (Zięba *et al.*, 2021).
K08	Continuity risks	Risks because of a firm's inability to maintain its core capabilities over time (Lambe, 2013)
K09	Communication risks	Risks of nuances within communication channels between two parties (Zięba *et al.*, 2021).
K10	Knowledge acquisition risks	Risks associated with an organization's ability to acquire new knowledge (Zięba et al., 2021).
K11	Knowledge transfer risks	Risks emanating from the inability of employees to effectively transfer knowledge among colleagues (Durst and Zięba, 2017)

3. Method

This study employs a quantitative approach to assess operational knowledge risks using a grey-based DEMATEL technique. The grey-based DEMATEL technique encompasses the grey theory and decision-making trial and evaluation laboratory (DEMATEL) technique. The grey theory is effective in dealing with ambiguities in judgments (Fu et al., 2001). Whereas the DEMATEL is an essential tool to establish causal relationships among complex variables using matrices and graphs (Shao et al., 2016). Hence, a grey-based DEMATEL technique is a combined methodology that enhances judgmental decisions using cause-effect diagraphs (Rajesh and Ravi, 2017). As Lee et al. (2021) suggest that most risk evaluation is governed by judgmental opinion and choice, so it is appropriate to use a grey-based DEMATEL technique in this study. In addition, this technique helps resolve complex issues (Govindan et al., 2016; Khan et al., 2020; Seker et al., 2017) such as dealing with knowledge risks, which are typically dynamic (Bratianu, 2018).

Concrete steps using grey-DEMATEL technique

- Step 1: Develop a grey direct-relation matrix

To develop the grey direct-relation matrix, an initial direct relationship matrix is formulated using the evaluation of risk r = {ri |i = 1, 2, …, n} by k through pairwise comparisons using the linguistic scale. Table 2 shows the 5-point linguistic scale and their associated grey numbers. This initial direct relationship matrix is converted into a grey initial direct relationship matrix by transforming the linguistic terms into the corresponding grey numbers. Hence, the k number of the grey direct relationship matrix Z1 is obtained, i.e., from one expert. The element of the grey direct-relation matrix is represented as '⊗Zij' (i.e., risk i influence j by an expert).

Table 2. Grey linguistic scale

Linguistic terms	Grey numbers
No influence (No)	[0,0]
Low influence (L)	[0,0.25]
Medium influence (M)	[0.25,0.5]
High influence (H)	[0.5,0.75]
Very high influence (VH)	[0.75,1]

Source: Adapted from Liu *et al.* (2020)

- Step 2: Formulate the normalized grey direct-relation matrix

The overall grey relation matrix is transformed into the normalized grey direct-relation matrix (see Table A1) N using Equations (1) – (3).

$$\otimes s = [\underline{s}, \underline{s}] = \frac{1}{\underset{0\le i\le n}{max} = \sum_{j=0}^{n} \otimes Z_{ij}} \quad i, j = 1,2,3, \dots n \tag{1}$$

$$N = \otimes s * Z \tag{2}$$

$$\otimes n_{ij} [\underline{s} * \otimes z_{ij} , \underline{s} * \otimes z_{ij}] \tag{3}$$

- Step 3: Compute the total relation matrix

The total relation matrix T (see Table A2) is determined by using Equation (4):

$$T = N(I - N)^{-1} \tag{4}$$

Where I if the identity matrix

- Step 4: Compute the causal parameters

The causal parameter is determined using Equations (5) and (6):

$$\otimes R_i = \sum_{j=1}^{n} t_{ij}\, \Theta_i \tag{5}$$

$$\otimes C_j = \sum_{i=1}^{n} t_{ij}\, \Theta_j \tag{6}$$

⊗Ri represents the direct and indirect influence of the risks *i* over the other risks, and ⊗Cj represents the influence received by risk *j* by the other risk.

- Step 5: Calculate the prominence (Pi) and net effect (Ei)

The prominence (Pi) and net effect (Ei) of the risks are determined using Equations (7) and (8):

$$\otimes P_i = \otimes R_i + \otimes C_i\,,\ i = j \tag{7}$$

$$\otimes E_i = \otimes R_i - \otimes C_i\,,\ i = j \tag{8}$$

The causal relationship diagraph is developed using the net effect value (shown in Table A1). A positive value of ⊗Ei shows the net effect (cause) of the risk on the system and a negative value represents the net effect on the risks caused by the system.

4. Findings and discussions

This section discusses operational knowledge risks within SMEs from two angles: the net cause/effect values and degree of prominence.

According to the computed net cause/effect values (Table A3), knowledge waste (K01), risks related to knowledge gaps (K02), outsourcing risks (K04), risk of using obsolete/unreliable knowledge (K05), and communication risks (K09) are grouped under the cause category and are deemed relevant to be considered while constructing risk management strategies.

Businesses, typically SMEs, are often at risk of losing knowledge when they transfer part of their operations to third or external parties (Durst and Ferenhof, 2014; North *et al*., 2020; Temel and Durst, 2020). Despite outsourcing being a good initiative, it can create a conducive environment where partners share ideas and knowledge openly and freely without bearing in mind not all knowledge should be shared. Further, outsourcing can contribute to the emergence of other risks, such as relational risk, which allows partners to take advantage of their position (Durst and Zięba, 2019). Therefore, establishing or building of trust is important to ensure risks associated with outsourcing could be minimized (Galati *et al.*, 2019; Sharif *et al*., 2022).

The risk of using obsolete knowledge also poses threat to a firm's knowledge stock, in that, it could trigger the risk of improper application of knowledge. This risk is of major concern since it has a ripple effect on a firm’s reputational damage if its consequence is realized (Durst and Zięba, 2020). It is advisable for SMEs to ensure their knowledge is continuously updated.

The effect group is categorized as follows: relational risk (K03), risk of improper knowledge application (K06), espionage (K07), continuity risk (K08), knowledge acquisition risks (K10), and knowledge transfer risks (K11).

These risks are theoretically viewed as the outcomes of the causal risks. In other words, they arise from existing risks and as all risk their relevance can vary from time to time, which underlines the need for a holistic approach to risk management (Smallman, 1996; Smith and Fischbacher, 2009)

Figure A1 shows the degree of prominence of each operational knowledge risks in terms of its priority in addressing them. The importance order of operational type of knowledge risk within SMEs is as follows: outsourcing risks (K04) > communication risks (K09) > risk of improper knowledge application (K06) > continuity risk (K08) > knowledge waste (K01) > risk of using obsolete/unreliable knowledge (K05) > knowledge transfer risks (K11) > risks related to knowledge gaps (K02) > espionage (K07) > relational risk (K03) > knowledge acquisition risks (K10).
The finding suggests that outsourcing risks are the most critical of all operational knowledge risks. The main reason for this is the fact that SMEs have a very low ability to be self-reliant in comparison with large firms. For SMEs to maintain their business and remain competitive, they are indirectly positioned to outsource majority of their business functions. According to Zięba and Durst (2018), outsourcing can lead to over-reliance on partners. Thus, when outsourcing ceases, SMEs will be unable to generate value internally for themselves, resulting in a temporary or permanent shutdown of their operations.

5. Conclusion

As mentioned earlier, the objective of the study is to holistically evaluate knowledge risks at the operational level of SMEs. Based on the operational risks identified by Temel and Durst's (2021), the DEMATEL technique shows:

- Risks coded K01, K02, K04, K05 and K09 are classified cause-risk
- Risks coded K03, K06, K07 and K08 are classified effect-risk
- Risks coded K04 > K09 > K06 > K08 > K01 > K05 > K11 > K02 > K07 > K03 > K10 in this order, the operational knowledge risks are ranked from the most to the least prominent

Findings from the study have implications for both theory and practice. From a theoretical perspective, the paper contributes to the underdeveloped body of knowledge about knowledge risks and KRM. More specifically, it introduces a new technique for knowledge risk assessment known as grey-DEMATEL; a rigorous analytical technique of analyzing risks (Seker *et al.*, 2017).
In addition, the study may be of interest to practitioners (e.g. directors, owners and managers) as it can help them evaluate knowledge risks at the operational level in a more holistic manner and provides them with information regarding the priorities in terms of the specific risks to address at a certain point in time. Since SMEs have limited resources, a priority scale will help them address the most relevant risks. In this study, the grey-based DEMATEL technique is clearly and systematically applied, offering practitioners a means of easily implementing it in their own organizations. Practitioners can benefit from this technique since it will prevent them from committing judgmental errors when conducting risk assessments. As indicated by Yazdani *et al.* (2020), such a technique is expected "to increase the quality of final decision and to reduce human judgmental errors" (p. 970).
All studies have limitations, and this one is no exception. Since the operational knowledge risks identified in this study may not be exhaustive, future research could use the Delphi- interview approach with more emphasis on KM experts to identify risks that might have been overlooked in this study. Additionally, future research should incorporate inputs from these KM experts during the risk assignment process for a more robust assignment of the identified risks according to the grey-linguistic scale.

Appendix

Table A1. Normalized Grey Direct-relation Matrix of Risk (K)

Risks	K01	K02	K03	K04	K05	K06	K07	K08	K09	K10	K11
K01	[0,0]	[0,0.03]	[0,0]	[0.09,0.09]	[0.13,0.12]	[0,0.03]	[0.13,0.12]	[0.09,0.09]	[0,0]	[0,0]	[0.09,0.09]
K02	[0.13,0.12]	[0,0]	[0,0]	[0.13,0.12]	[0,0.03]	[0.13,0.12]	[0.04,0.06]	[0.13,0.12]	[0,0]	[0.13,0.12]	[0,0]
K03	[0,0]	[0,0]	[0,0]	[0.09,0.09]	[0,0]	[0.09,0.09]	[0.13,0.12]	[0,0]	[0,0]	[0,0.03]	[0.04,0.06]
K04	[0,0.03]	[0,0]	[0.13,0.12]	[0,0]	[0,0.03]	[0.04,0.06]	[0.13,0.12]	[0.13,0.12]	[0.09,0.09]	[0.09,0.09]	[0.13,0.12]
K05	[0.13,0.12]	[0.09,0.09]	[0,0]	[0,0]	[0,0]	[0.13,0.12]	[0,0]	[0.09,0.09]	[0.04,0.06]	[0,0.03]	[0,0]
K06	[0.13,0.12]	[0,0]	[0,0.03]	[0,0]	[0.09,0.09]	[0,0]	[0.04,0.06]	[0.09,0.09]	[0,0]	[0.09,0.09]	[0,0]
K07	[0,0]	[0,0.03]	[0,0.03]	[0,0]	[0,0]	[0,0]	[0,0]	[0.13,0.12]	[0,0]	[0,0]	[0.09,0.09]
K08	[0,0]	[0,0]	[0,0]	[0.13,0.12]	[0,0]	[0,0]	[0,0]	[0,0]	[0,0]	[0,0]	[0,0]
K09	[0.13,0.12]	[0,0.03]	[0.13,0.12]	[0.09,0.09]	[0.13,0.12]	[0.13,0.12]	[0.13,0.12]	[0,0.03]	[0,0]	[0.13,0.12]	[0.13,0.12]
K10	[0,0]	[0.09,0.09]	[0,0]	[0,0]	[0.04,0.06]	[0.09,0.09]	[0,0]	[0.04,0.06]	[0,0]	[0,0]	[0,0]
K11	[0,0]	[0.04,0.06]	[0.09,0.09]	[0.13,0.12]	[0,0.03]	[0.04,0.06]	[0,0]	[0.04,0.06]	[0,0]	[0,0]	[0,0]

Source: Compiled by authors

Risks	K01	K02	K03	K04	K05	K06	K07	K08	K09	K10	K11
K01	[0.028,0.043]	[0.019,0.062]	[0.031,0.04]	[0.134,0.146]	[0.141,0.149]	[0.039,0.084]	[0.16,0.16]	[0.163,0.18]	[0.018,0.022]	[0.02,0.037]	[0.124,0.132]
K02	[0.164,0.165]	[0.021,0.035]	[0.031,0.038]	[0.187,0.179]	[0.046,0.087]	[0.169,0.176]	[0.104,0.122]	[0.218,0.218]	[0.018,0.022]	[0.166,0.164]	[0.051,0.053]
K03	[0.017,0.025]	[0.006,0.019]	[0.022,0.033]	[0.109,0.119]	[0.013,0.026]	[0.103,0.12]	[0.156,0.153]	[0.058,0.066]	[0.01,0.012]	[0.021,0.058]	[0.075,0.095]
K04	[0.032,0.077]	[0.021,0.042]	[0.169,0.17]	[0.079,0.085]	[0.03,0.082]	[0.099,0.134]	[0.185,0.185]	[0.21,0.219]	[0.095,0.104]	[0.118,0.136]	[0.179,0.178]
K05	[0.177,0.173]	[0.096,0.114]	[0.017,0.027]	[0.058,0.063]	[0.046,0.056]	[0.164,0.171]	[0.051,0.058]	[0.154,0.167]	[0.051,0.07]	[0.038,0.076]	[0.035,0.039]
K06	[0.153,0.148]	[0.02,0.031]	[0.008,0.044]	[0.039,0.044]	[0.115,0.124]	[0.031,0.043]	[0.073,0.095]	[0.138,0.151]	[0.008,0.012]	[0.097,0.109]	[0.026,0.032]
K07	[0.002,0.01]	[0.005,0.039]	[0.013,0.046]	[0.034,0.041]	[0.002,0.01]	[0.008,0.021]	[0.007,0.016]	[0.15,0.15]	[0.003,0.004]	[0.005,0.013]	[0.093,0.102]
K08	[0.004,0.009]	[0.003,0.005]	[0.022,0.021]	[0.141,0.132]	[0.004,0.01]	[0.013,0.016]	[0.024,0.022]	[0.027,0.026]	[0.012,0.013]	[0.015,0.016]	[0.023,0.022]
K09	[0.19,0.191]	[0.039,0.089]	[0.173,0.178]	[0.17,0.183]	[0.184,0.193]	[0.208,0.22]	[0.214,0.21]	[0.139,0.179]	[0.023,0.028]	[0.171,0.183]	[0.198,0.194]
K10	[0.035,0.04]	[0.095,0.104]	[0.005,0.01]	[0.028,0.032]	[0.06,0.084]	[0.112,0.122]	[0.019,0.025]	[0.082,0.106]	[0.005,0.008]	[0.025,0.03]	[0.009,0.011]
K11	[0.02,0.037]	[0.049,0.075]	[0.115,0.123]	[0.178,0.174]	[0.013,0.058]	[0.076,0.108]	[0.048,0.054]	[0.182,0.184]	[0.016,0.019]	[0.031,0.043]	[0.036,0.039]

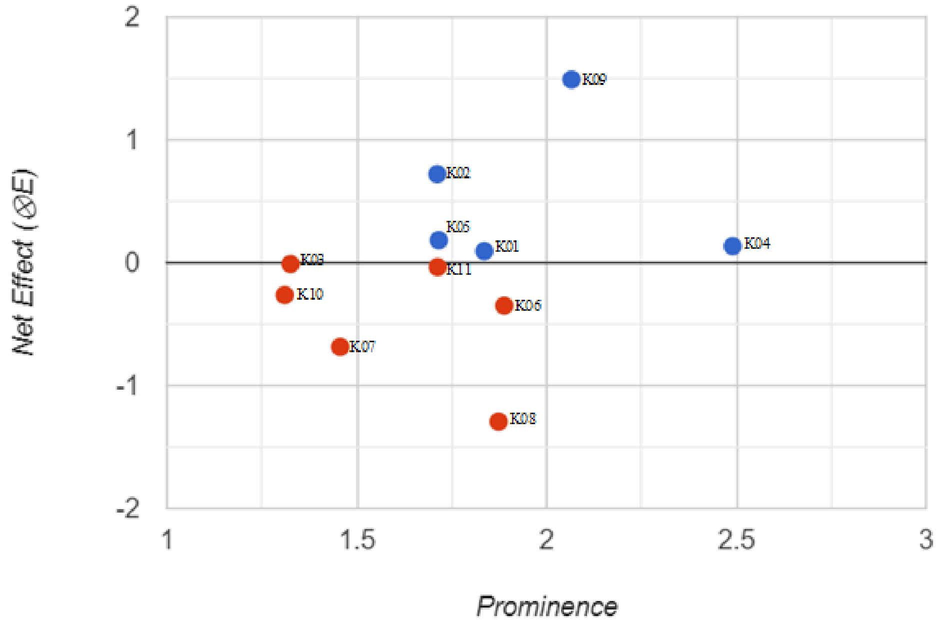
Source: Compiled by authors

Table A3. Prominence and Net Effect of the Risks

Risks	Ri	Ci	Ri+Ci	Ri-Ci	Cause/Effect
K01	[0.877,1.055]	[0.822,0.918]	[1.699,1.973]	[0.054,0.137]	Cause
K02	[1.174,1.26]	[0.373,0.617]	[1.547,1.877]	[0.802,0.643]	Cause
K03	[0.59,0.725]	[0.607,0.729]	[1.197,1.454]	[-0.016,-0.003]	Effect
K04	[1.216,1.41]	[1.157,1.197]	[2.373,2.608]	[0.059,0.213]	Cause
K05	[0.886,1.014]	[0.652,0.879]	[1.538,1.894]	[0.234,0.135]	Cause
K06	[0.707,0.833]	[1.021,1.216]	[1.728,2.049]	[-0.314,-0.383]	Effect
K07	[0.322,0.451]	[1.04,1.1]	[1.362,1.551]	[-0.719,-0.649]	Effect
K08	[0.289,0.292]	[1.52,1.646]	[1.809,1.938]	[-1.231,-1.354]	Effect
K09	[1.708,1.85]	[0.259,0.314]	[1.968,2.164]	[1.449,1.536]	Cause
K10	[0.476,0.572]	[0.707,0.866]	[1.182,1.438]	[-0.231,-0.293]	Effect
K11	[0.764,0.915]	[0.851,0.896]	[1.616,1.81]	[-0.087,0.019]	Effect

Source: Compiled from authors

Figure A1. Causal relationship between operational knowledge risks



Source: Designed by authors (Note: ●Effect ●Cause)

References

Al-Hussaini, W. and Karkoulian, S. (2015) “Mitigating operational risk through knowledge management”, *Journal of International Management Studies*, 15(2): 31-96

Agndal, H. and Nordin, F. (2009) “Consequences of outsourcing for organizational capabilities: Some experiences from best practice”, *Benchmarking: An International Journal*, 16(3): 316 334. https://doi.org/10.1108/14635770910961353

Bakir, S., Khan, S., Ahsan, K. and Rahman, S. (2018) “Exploring the critical determinants of environmentally oriented public procurement using the DEMATEL method”, *Journal of environmental management*, 225: 325-335.

Biswas, B. and Gupta, R. (2019) “Analysis of barriers to implement blockchain in industry and service sectors”, *Computers & Industrial Engineering*, 136: 225-241.

Bratianu, C. (2018) “A holistic approach to knowledge risk”, *Management Dynamics in the Knowledge Economy*, 6(4): 593-607.

Callahan, C. and Soileau, J. (2017) “Does Enterprise risk management enhance operating performance?”, *Advances in Accounting*, 37:122-139. 10.1016/j.adiac.2017.01.001.

Cruz, M. and Pinedo, M. (2008) “Total quality management and operational risk in the service industries”, In State-of-the-Art Decision-Making Tools in the Information- Intensive Age, *INFORMS*: 154-169.

Delerue, H. (2005) “Relational risk perception and alliance management in French biotechnology SMEs”, *European Business Review*, 17(6): 532 546. https://doi.org/10.1108/09555340510630563

Durst, S. and Ferenhof, H. (2014) “Knowledge Leakages and Ways to Reduce Them in Small and Medium-Sized Enterprises (SMEs)”, *Information*, 5(3): 440-450. 10.3390/info5030440.

Durst, S. and Zięba, M. (2017) “Knowledge risks - Towards a taxonomy. International Journal of Business Environment”, 9(1). https://doi.org/10.1504/IJBE.2017.084705

Durst, S. and Zięba, M. (2018) “Mapping knowledge risks: towards a better understanding of knowledge management”, *Knowledge Management Research & Practice*, 17(1): 1-13. 10.1080/14778238.2018.1538603.

Durst, S. and Zięba, M. (2019) “Knowledge risks inherent in business sustainability”, *Journal of Cleaner*

Production, 251: 119670. 10.1016/j.jclepro.2019.119670.

Durst, S. and Zięba, M. (2020) "Knowledge risks inherent in business sustainability", *Journal of cleaner production*, 251: 119670.

Durst, S., Hinteregger, C. and Zięba, M. (2019) "The linkage between knowledge risk management and organizational performance", *Journal of Business Research*, 105: 1-

10. 10.1016/j.jbusres.2019.08.002.

Ferenhof, H. A., Durst, S. and Selig, P. M. (2016) "Knowledge Waste & Knowledge Loss- What is it All About?", *NAVUS-Revista de Gestão e Tecnologia*, 6(4): 38-57.

Foli, S. (2022) "Total interpretive structural modelling (TISM) and MICMAC approach in analysing knowledge risks in ICT-supported collaborative project", *VINE Journal of Information and Knowledge Management Systems*. https://doi.org/10.1108/VJIKMS- 09-2021-0205

Fu, C., Zheng, J., Zhao, J. and Xu, W. (2001) "Application of grey relational analysis for corrosion failure of oil tubes", *Corrosion Science*, 43(5): 881-889.

Galati, F., Bigliardi, B., Petroni, A., Petroni, G. and Ferraro, G. (2019) "A framework for avoiding knowledge leakage: evidence from engineering to order firms", *Knowledge Management Research & Practice*, 17(3): pp. 340-352.

Gamble, J. (2020) "Tacit vs explicit knowledge as antecedents for organizational change", *Journal of Organizational Change Management*, 33(6): 1123-1141. https://doi.org/10.1108/JOCM-04-2020-0121

Govindan, K., Khodaverdi, R. and Vafadarnikjoo, A. (2016) "A grey DEMATEL approach to develop third-party logistics provider selection criteria", *Industrial Management & Data Systems*, 116(4).

Grant, R. M. (2003) *The knowledge-based view of the firm*, The Oxford handbook of strategy, 1: 197-221.

Haimes, Y. Y. (2009) "On the complex definition of risk: A systems-based approach", *Risk Analysis: An International Journal*, 29(12): 1647-1654.

Haleem, A., Khan, S. and Khan, M. I. (2019) "Traceability implementation in food supply chain: A grey-DEMATEL approach", *Information Processing in Agriculture*, 6(3): 335-348.

Hammoda, B. and Durst, S. (2022) "A taxonomy of knowledge risks for healthcare organizations", *VINE Journal of Information and Knowledge Management Systems*.

Henschel, T. and Durst, S. (2016) "Risk management in Scottish, Chinese and German small and medium-sized enterprises: a country comparison", *International journal of entrepreneurship and small business*, 29(1): 112-132.

Khan, S., Haleem, A. and Khan, M. I. (2020) "Enablers to implement circular initiatives in the supply chain: A grey DEMATEL method", *Global Business Review,* 0972150920929484.

Lambe, P. (2013), *Four Types of Knowledge Risk*, http://www.greenchameleon.com/uploads/Four_Types_of_Knowledge_Risk.pdf

Lee, R. W., Yip, J. Y. and Shek, V. W. (2021) "Knowledge Risk and Its Mitigation: Practices and Cases", *Emerald Group Publishing*.

Liu, H., Long, H. and Li, X. (2020) "Identification of critical factors in construction and demolition waste recycling by the grey-DEMATEL approach: A Chinese perspective", *Environmental Science and Pollution Research*, 27(8): 8507-8525.

Manab, N. A., Aziz, N. A. A. and Jadi, D. M. (2020) "Sustainability risk management: an integrative framework to evaluate emerging risks and other non-quantifiable risks affecting company survival", *World Review of Science, Technology and Sustainable Development*, 16(2): 87-104.

Massingham, P. (2010) "Knowledge risk management: a framework", *Journal of Knowledge Management*, 14(3): 464-485.

https://doi.org/10.1108/13673271011050166

Murnane, R., Simpson, A. and Jongman, B. (2016) "Understanding risk: what makes a risk assessment successful?", *International Journal of Disaster Resilience in the Built Environment*, 7(2): 186-200. https://doi.org/10.1108/IJDRBE-06-2015-0033

North, K., De Carvalho, A. B., Braccini, A. M., Durst, S., Carvalho, J. Á., Gräslund, K. and Thalmann, S. (2020) *Knowledge risks in supply chain interactions of SMEs: An exploratory study*, In 2019 Knowledge Management in Digital Work Environments, State-of-the-Art and Outlook, WM 2019, Potsdam, Germany, 18 March 2019 through 20 March 2019, Code 166782 (pp. 161-171). Gesellschaft für Informatik.

Perrott, B. E. (2007) “A strategic risk approach to knowledge management”, *Business Horizons*, 50(6): 523-533.

Purpura, P.P. (2013) "12 - Resilience, Risk Management, Business Continuity, and Emergency Management", in Purpura, P.P. (Ed.) *Security and Loss Prevention (Sixth Edition)*, Butterworth-Heinemann, Amsterdam: 321-62.

Rajesh, R. and Ravi, V. (2017) “Analyzing drivers of risks in electronic supply chains: a grey–DEMATEL approach”, *The International Journal of Advanced Manufacturing Technology*, 92(1): 1127-1145.

Seker, S., Recal, F. and Basligil, H. (2017) “A combined DEMATEL and grey system theory approach for analyzing occupational risks: A case study in Turkish shipbuilding industry”, *Human and Ecological Risk Assessment: An International Journal*, 23(6): 1340-1372.

Shao, J., Taisch, M. and Ortega-Mier, M. (2016) “A grey-DEcision-MAking Trial and Evaluation Laboratory (DEMATEL) analysis on the barriers between environmentally friendly products and consumers: practitioners' viewpoints on the European automobile industry”, *Journal of Cleaner Production*, 112: 3185-3194.

Sharif, S. M. F., Naiding, Y. and Shah, S. K. (2022) “Restraining knowledge leakage in collaborative projects through HRM”, VINE Journal of Information and Knowledge Management Systems.

Smallman, C. (1996) “Risk and organizational behaviour: a research model”, *Disaster Prevention and Management*, 5(2): 12–26.

Smith, D. and Fischbacher, M. (2009) “The changing nature of risk and risk management: The challenge of borders, uncertainty and resilience”, *Risk management*, 11(1): 1-12. Sufiyan, M., Haleem, A., Khan, S. and Khan, M. (2019) “Evaluating food supply chain performance using hybrid fuzzy MCDM technique”, *Sustainable Production and*

Consumption, 20: 40-57. doi: 10.1016/j.spc.2019.03.004.

Tang, F. (2011) “Knowledge transfer in intra-organization networks.” *Systems Research and Behavioral Science*, 28(3): 270-282.

Tsang, H. W. C., Lee, W. B. and Tsui, E. (2016) “AHP-driven knowledge leakage risk assessment model: a construct-apply-control cycle approach”, *International Journal of Knowledge and Systems Science (IJKSS)*, 7(3): 1-18.

Temel, S. and Durst, S. (2020) “Knowledge risk prevention strategies for handling new technological innovations in small businesses”, *VINE journal of information and knowledge management systems*, 51(4): 655-673. https://doi.org/10.1108/VJIKMS-

10-2019-0155

Todinov, M.T. (2007) *1 - Risk-based reliability analysis: a powerful alternative to the traditional reliability analysis*, in Todinov, M.T. (Ed.), Risk-Based Reliability Analysis and Generic Principles for Risk Reduction, Elsevier, Oxford: 1-6.

Vishvakarma, N.K., Singh, R.K. and Sharma, R.R.K. (2022) "Cluster and DEMATEL Analysis of Key RFID Implementation Factors Across Different Organizational Strategies", *Global Business Review*, 23(1): 176-91

Yazdani, M., Torkayesh, A. E. and Chatterjee, P. (2020) “An integrated decision-making model for supplier evaluation in public healthcare system: the case study of a Spanish hospital”, *Journal of Enterprise Information Management*, 33(5): 965-989.

https://doi.org/10.1108/JEIM-09-2019-0294

Zheng, L. (2017) “Does online perceived risk depend on culture? Individualistic versus collectivistic culture”, *Journal of Decision systems*, 26(3): 256-274.

Zięba, M. and Durst, S. (2018) “Knowledge risks in the sharing economy”, *In Knowledge management in the sharing economy*, pp. 253-270, Springer, Cham.

Zięba, M., Durst, S., Gonsiorowska, M. and Zralov, Z. (2021) *Knowledge Risks in Organizations–Insights from Companies*, In European Conference on Knowledge Management, Academic Conferences International Limited: 864-873

Knowledge Management (KM) in Radioactive Waste Management (RWM).

D. ABBASOVA[a],
Institute of Resource Ecology, HZDR,
Bautzner Landstrasse 400, 01328 Dresden, Germany
Email: d.abbasova@hzdr.de

G. HOEFER[b], T.ARNOLD[a], S. WANKA[b], C.FRANZEN[a], P. L. WELLMANN[b]
[a] *Institute of Resource Ecology, HZDR,*
Bautzner Landstrasse 400, 01328 Dresden, Germany Email:
t.arnold@hzdr.de; c.franzen@hzdr.de
[b] *BGE Bundesgesellschaft für Endlagerung mbH Bereich F&E /*
Wissensmanagement Abteilung Wissensmanagement Standort Salzgitter
Chemnitzer Str. 42, 38226 Salzgitter,Germany Email:
Gunnar.Hoefer@bge.de; Peter.Wellmann@bge.de

KM is the process through which organizations generate value from their intellectual and knowledge-based assets [1]. Implementation of Knowledge Management (KM) is an important issue for all types of nuclear organizations and in particular for Radioactive Waste Management (RWM) organizations. Thus, the fundamental objective of RWM is to manage the radioactive waste without adverse impact for human health and the environment during its radioactive waste lifetime. Considering the life cycle of the radioactive waste it is obvious that KM accompanies Safety Management (SM). The management of radioactive waste affects future generations and covers pretreatment, treatment, conditioning, storage and disposal. At all mentioned stages operators are dealing with information such as take the records, use the standards & templates, prepare the reports. Without adequate knowledge it would be impossible to carry out this work and conclusively safety will be under risk. It is clear, that managing the knowledge must be implemented at all stages of RWM with the integration of knowledge processes concentrating on the following four core activities:

- To generate the knowledge
- To store the knowledge
- To share the knowledge
- To distribute the

knowledge and will be focused on

- Information and Document Management
- Human Resource Management
- Knowledge Organizational Structure

However, depending on the organization (Regulatory, Operations, R&D organizations) involved in RWM the used KM methods & tools will be different. Nuclear R&D organizations can be categorized into seven types oriented on types of functions undertaken by the respective organization. In 2012 IAEA[2] has classified them as follows: (1.) Basic research functions, (2.) Applied research functions, (3.) Design R&D functions, (4.) Functions utilizing nuclear R&D facilities, (5.) Functions utilizing non-nuclear R&D facilities, (6.) Educational R&D functions and (7.) Technical support & service functions.
Referring to R&D functions it becomes obvious that out of seven known Knowledge levels[3] R&D organizations mainly use four (i.e. Organizational memory, Knowledge in Processes, Knowledge in Products & Services and Knowledge in People). An example for this is the KM initiatives in the Institute of Resource Ecology.
The Institute of Resource Ecology is one of eight institutes of the Helmholtz-Zentrum Dresden-Rossendorf (HZDR) where many administrative tasks are centralized. For example the human resources (HR) department, the financial department, the international department and the IT department are all administrated centrally by the HZDR. The HZDR has an internal platform which is used for administrative work, data & information management, communication and keeps employees updated with ongoing and upcoming activities in their institutes. Beside this platform, HZDR employees use different IT tools, e.g. Outlook, Mattemost, etc, for close cooperation. The employees skill development program is administrated by the center. This program follows the requirements collected from all HZDR employees.
Continuous education and training programs for the HZDR employees and students, well organized in cooperation with partner organizations (TU-Dresden etc) do also support the knowledge capturing & sharing process.
The centralized library has more than 40.000 books in hard copy and 16.300 e-books & journals and all are available for employees of the center. The repository of HZDR's publications has more than 36.000 scientific papers. In addition, each institute has also an internal scientific database for employee's needs.

There are several formats of information and knowledge exchange in the Institute. To convince the employees of the importance of knowledge-sharing the institute provides series of different workshops regularly e.g. monthly seminars, workshops, department meetings, strategy and jour-fixe meetings in hybrid formats.
Obviously, the critical knowledge in the institute is linked to research works. The critical knowledge here is defined as knowledge which is needed to meet the objectives of the Institute. The critical knowledge holders are well known. There are several practices applied in the institute for maintaining the knowledge and transfer it. First of all to mention here is the supervision & mentoring program for the PhD students by critical knowledge holders. Another one is the procedure for uploading publications and records to the internal repository. There are also two inhouse scientific databases at the IRE available for internal and external users: (1.) RES³T - Rossendorf Expert System for Surface and Sorption Thermodynamics and database and (2.) the Thermodynamic Reference Database (THEREDA). However, there is still a high demand for the development of Knowledge Preservation programs. The loss of knowledge in association with the retirement of knowledge holders is still an issue in the institute. This is true for all nuclear organizations.
Different approaches have been used in industry. As an example, the approach applied by the Federal Company for Radioactive Waste Disposal (BGE) is described.
Due to the restructuring of the nuclear waste-management landscape in Germany, BGE is since 2017 the national competence centre in Germany and responsible for the disposal of radioactive waste.
For many decades, Research &Developement-work for this subject has been performed in Germany with an enormous stock of knowledge (topical and quantitative), but an overview is impaired due to the diffuse and local organisation of knowledge of the nuclear waste-management landscape.
Research reports for example can be found in many archives, but no archive is complete and „Old stock of knowledge"(old documents, grey literature) has to be embedded in the new stock of knowledge.
At the same time the imminent loss of expertise due to phasing out of nuclear energy production and mining in Germany must be considered as a limiting and partially critical factor, especially while reflecting the age structure of the BGE-staff and the restart of the new site selection procedure for disposal of highly radioactive waste products.
Facing these prerequisites the newly formed internal department for knowledge management (KM) of the BGE is establishing an infrastructure for KM and generates a connection between KM platforms and knowledge holders in the company to make explicit, implicit and tacit knowledge available.
The approach for the explicit knowledge is the provision of a digital information basis, into which current results from research and development are entered as a knowledge store. This knowledge store currently consists of more than 16.000 documents, mainly research reports and scientific publications, which are concerned with diverse topics for the final disposal of radioactive waste.
The total stock of available internal company documents can be accessed with a browser-based text analysis software. Intelligent search algorithms render the textual contents accessible, combine them with synonyms and dictionaries deposited in the system and make the resulting hits of the search queries available for the user in order of importance in summarized and full text versions.
Using specific query terms the software analyzes the available documents of the digital information basis and provides a brief description of the contents, the naming of relevant keywords, the identification of sources, compilers, institutions, knowledge carriers and an extended optimized information analysis of hits, as well as the possibility to access the complete document. The search options can be combined with established internet search engines as well as with queries of incorporated information or databank catalogue of national and international scientific institutions or libraries, which are concerned with research programs relevant for repositories[4]. The content volume of the digital information basis, as well as the amount of externally connected research sources is permanently growing.
To further increase and optimize the information possibilities for employees of the BGE and to make implicit knowledge partly available, a variety of general and demand-oriented interactive knowledge maps have been and will be implemented in the intranet, which enable specific queries on topics, expert information, etc.
To capture the tacit knowledge of persons who are leaving the company due to retirement, concerted concepts, e.g. interviews including transcriptions are used to make this individual knowledge available to the BGE-staff using the BGE internal tools for capturing and distributing implicit and explicit knowledge.
Additional actions contain, e.g. the initiation and organisation of technical and professional talks on various topics and to establish them as a permanent and important exchange possibility in the company.
While "How to use" seminars for the use of the broser-based text analysis software are already established and a permanent and reoccurring part in the company, further seminars and videos with "How to use" approach e.g. for demand-oriented interactice knowledge maps are in the implementations process as well as a Podcast series about Knowledge Management in the BGE.
For the further development of person-related and further group-related knowledge, the BGE Knowledge Management group/department is compiling concepts, which can only be implemented together with the employee in the company and its own guiding principles as they border on certain interfaces in the organization/company, taking into account that all measures to share and distribute knowledge can only be a permanent success if it is voluntarily process without compulsion.

By writing this paper we tried to illustrate the practical difference between KM initiatives in R&D and industry. However, the overlapping of the approaches at some certain stages are visible. The information management has been considered as an essential part of the knowledge management in both of organizations. The capture of the critical knowledge in both organizations remains a main issue even if the selected methods are different. The coaching & mentoring program which are well implemented at the IRE (HZDR), are may be difficult to realize/accomplish at the BGE. However, the pilot coaching program is planned at the BGE, but currently not available due to the "young" founding date in 2017. One of the main concerns of the BGE as an implementer, is the motivation and encouragement of all employees to share and distribute their knowledge and benefit from each other.

Reference:

1. SANTOSUS, MEGAN, and SURMACZ, The ABCs of Knowledge Management, Knowledge Management Research Center, accessed on 09 December 2005 at http://www.cio.com/research/knowledge/edit/kmabcs.html
2. IAEA, Knowledge Management for nuclear Research and Development organizations IAEA, Vienna, 2012
3. Jatinder N.D. Gupta, Sushil K. Sharma, Jeffrey Hsu, “An overview of Knowledge Management”, Knowledge management. I. Jennex, Murray E., 1956
4. Gunnar Hoefer, Sebastian Wanka, Peter L. Wellmann, Knowledge Management in the Federal Company for Radioactive Waste Disposal (BGE), Saf. Nucl. Waste Disposal, 1, 251–253, 2021

Infinity Maps as a Visual Knowledge Management Platform – Concept and Use Cases

Heiko Haller*
Infinity Maps GmbH, Breitwiesenweg 2, 76571 Gaggenau, Germany
Email: heiko@infinitymaps.io

The iMapping technique is a visual knowledge mapping approach that is based on deep zooming and nesting and that unites the essential benefits of virtual whiteboards, mind maps and concept maps. To achieve that, it breaks the physical page metaphor and introduces an infinitely zoomable, nested canvas, to create large scalable knowledge spaces. It has been originally developed and scientifically evaluated in the context of personal knowledge management only. Infinity Maps is the state-of-the art successor of the iMapping Tool: A cloud based visual knowledge and productivity platform. It not only fully implements the visual iMapping approach but takes it to the next level, by adding real-time collaboration and sharing functionality – thus also adding the benefit of wikis. This has enabled knowledge workers to create shared knowledge spaces with hitherto unseen depth. Nested Maps with thousands of individual cards, 10 and more levels of hierarchy and zoom-factors of one to several thousand are normal.
With this paper, we want to introduce Infinity Maps to the knowledge management community and showcase some real-life examples of visual Knowledge Management, from personal research over teaching and learning to corporate knowledge management.

1. Motivation

In nowadays complex world, knowledge workers are frequently confronted with their cognitive limitations. We are overwhelmed by complexity, by the sheer amount of material or by getting a grasp of the structure and interrelations of topics.

One major cognitive limitation is, that our human brain can only represent around 4 concepts / items in working memory at any instant (Cowan, 2001). This, of course makes it hard for us to tackle complex interconnected problems. The only way we can still approach these is by rapidly switching concepts and picking up other combinations of them from long term memory, which causes a high amount of cognitive load.

Using visualizations to aid orientation in the information space relieves some of that cognitive load and allows relevant or at least related concepts to be easily at hand and to swap the corresponding concepts into working memory.

Experience in practice and research over the decades have shown that certain visualization techniques are beneficial to understanding and learning. The most prominent representatives of the most common structural approaches being mind maps, concept maps and what is intuitively used on pinboards and labelled “spatial hypertext” by some researchers of the field. Unfortunately, however, all of the established approaches have their specific strengths and weaknesses (Haller 2003).

2. Combining the Benefits of Visual Approaches with the iMapping Technique

To unite the core benefits of these most commonly used visualization approaches, the "iMapping" technique has been designed, evaluated and published by the author and his colleagues (Haller, Kugel and Völkel, 2006; Haller and Abecker, 2010; Haller, 2011). Here is a short overview over the basic structures, their most prominent representative and how the benefits are realized in iMapping:

Whiteboards: Simplicity

The way most pin boards, white boards, walls with sticky notes etc. are intuitively used, is, to use to spatial distance to group items together that belong together – without any formal graph structure. This approach has been examined under the label of "spatial hypertext" (e.g. Shipman and Marshall, 1999a,b). The two main benefits of this very widespread approach to visualization are:

- Its simplicity requires very little cognitive effort.
- User-chosen spatial positions remain constant, so that items can easily recognized and found again through visual sense of orientation.

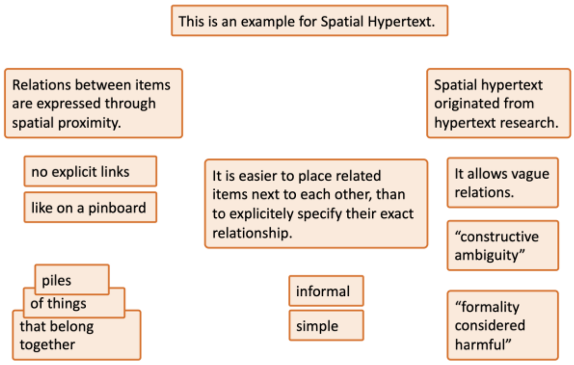


Fig. 1. Example of a spatial hypertext, like whiteboards are commonly used.

In iMapping, items / "cards" can be placed freely, to enable these benefits.

Mind Maps: Hierarchy

Mind maps, org charts, fishbone diagrams etc. all boil down to tree-like visualizations of *hierarchical* structures. Since we use hierarchy to break down content structures in any domain (from geography over book chapters to file spaces), visualizations of hierarchy are ubiquitous. Their main benefit is to visualize a relationship of *context and detail*.

In iMapping, hierarchy is represented through nesting, as can be seen in Figures 4 and later.

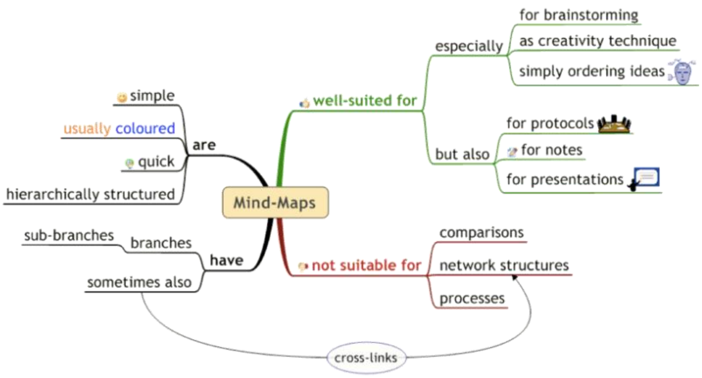


Fig. 2. Example of a mind map.

Concept Maps: Interrelations

Concept maps and other network-like graph diagrams are common to display all sorts of interrelation structures. Like mind maps they have been well evaluated for learning and other purposes. Their main benefit is also their main burden: If these graphs contain too many interrelations, they become hard to read.

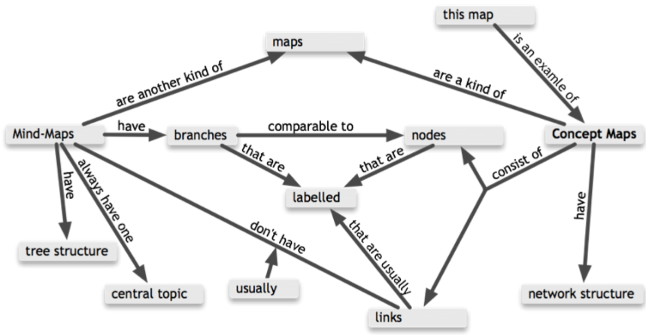


Fig. 3. Example of a concept map.

iMapping adopts the node-and-arrow way of displaying interrelations. However, usually not all arrows are shown simultaneously, but only the ones connected to the cards currently selected or hovered upon.

Hypertext: Scalability

The visual approach of hypertext (like in wikis, the WWW or the pre-digital zettelkasten) is, basically, to only highlight reference points (hyperlinks) to other nodes or resources and to not further visualize their interrelations at all. This however enables hypertext systems to scale infinitely, because their visual complexity does not grow with the amount of content.

Unlike the three approaches described above, the iMapping technique achieves scalability through its concept of deep nesting and zooming, as can be seen from Fig. 4 on.

3. Creating Shared Knowledge Spaces with Infinity Maps

While the original iMapping Tool served well as a proof of concept and found a small enthusiastic fanbase, it never really found widespread use. This is partly due to its nature as an old-school single-user offline app, that was originally designed as a cognitive tool for individual knowledge workers.

With Infinity Maps, we have on the one hand essentially rebuilt the functionality of the original iMapping Tool. On the other hand, it is now a state of the art, cloud-based software-as-a-service. And with the new technical basis came the capabilities for real time collaboration, effortless sharing, and various ways of integrations – in both ways: Interactive Maps can be transcluded in other online systems from MS-Teams to weblogs. But also other online resources can be shown inside Infinity Maps, from uploaded files to previews of linked webpages.

These collaboration and sharing features have made Infinity Maps suitable to a range of new use cases in the domain of (not only personal) knowledge management.

4. Usecases

To illustrate, how Infinity Maps are being used by individuals, teams and organizations, here are some of the use cases. Some use pictures with real content, others have been rebuilt because the real examples contain confidential business information of Infinity Maps clients.

Personal Knowledge Management

This is where Infinity Maps comes from. Its predecessor, the iMapping Tool, has been designed and used mainly as a tool for personal knowledge management, serving individual knowledge workers as a thinking aid and personal knowledge base. Most long-term users have started with such personal knowledge maps. Some of them, over the years have grown to more than 10,000 cards and more than 12 levels of hierarchy deep.

Fig. 4. Zoomed-out view of the author's personal map, containing > 10.000 nested cards. Most contents are nested in so deep that they are not visible from here.

In personal maps, that are highly familiar to their users, like the ones in Fig. 4 and 5, orientation and navigation mostly happens based on visual recognition of the map's layout. Like personal notes, they are not necessarily easy to understand by others.

Visual Wiki

In teams and organizations, instead of a classical wiki, Infinity Maps are used as a visual knowledge base that is maintained collaboratively. As in a wiki, space is unlimited. And as in a wiki, results are best, when people take up ownership of certain areas and clean them up regularly to maintain overview and useful structures.

Other than with wikis though, information is always visually embedded in a context. Like this, in addition to text-bases search and navigation, visual orientation can be used to find contents.

Shared maps like the one from Fig. 6, need to be structured in a plausible way, so that newcomers can easily find their way through unknown parts.

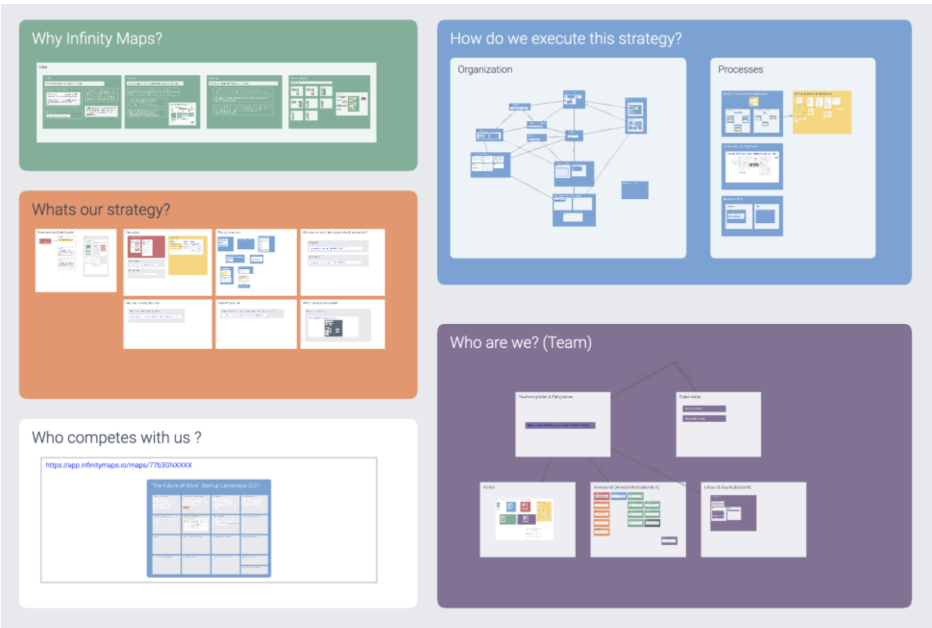


Fig. 6. Top-level view of a company knowledge space. Contents are deeply nested in a structure of areas and sub-areas.

Knowledge (Resources) Map

Instead of maintaining most of the actual content visually *in* the map, some organizations use Infinity Maps for the overview of knowledge areas only and then link to existing knowledge repositories like wikis and shared file systems. For implicit knowledge, expert maps are useful, to help colleagues find the right person to approach for their needs.

These kinds of maps are usually referred to as “knowledge maps”, however it is not the only use case under that label:

Knowledge (Transfer) Map

Another way, “knowledge maps” are used is for knowledge transfer (from a person leaving the company to a successor) – a need that is currently increasing a lot due to demographic change. Hereby, the *areas* (not the actual contents) of relevant knowledge and expertise of the parting coworker is charted in a knowledge map. These maps serve a twofold purpose:

First, they help to plan and track the knowledge transfer process – like a travel plan through the territory: Which areas need deliberate training, which can be handed over gradually after a period of shadowing? Which stakeholders should be introduced? What has already been covered? And which areas can be learnt later, on the job? For this, mind maps have already proven useful over the years, however they are limited in size, can hardly depict interrelations and thus are not suitable for the second purpose.

Secondly, a knowledge transfer map based on Infinity Maps can serve the newcomer as a visual scaffold for personal learnings, notes and other resources during the onboarding and transfer process. According to user reports, Infinity Maps are perceived most useful as personal aids, in periods of orientation and familiarizing with new topics.

Project Management

Depending on the nature of a project, there are various areas, in which visual overview has proven useful to project managers. These three however seem to be among the most common: Stakeholder Maps, IT System Architecture Maps and Work package breakdowns. Often, they are interwoven and benefit from living in a common map.

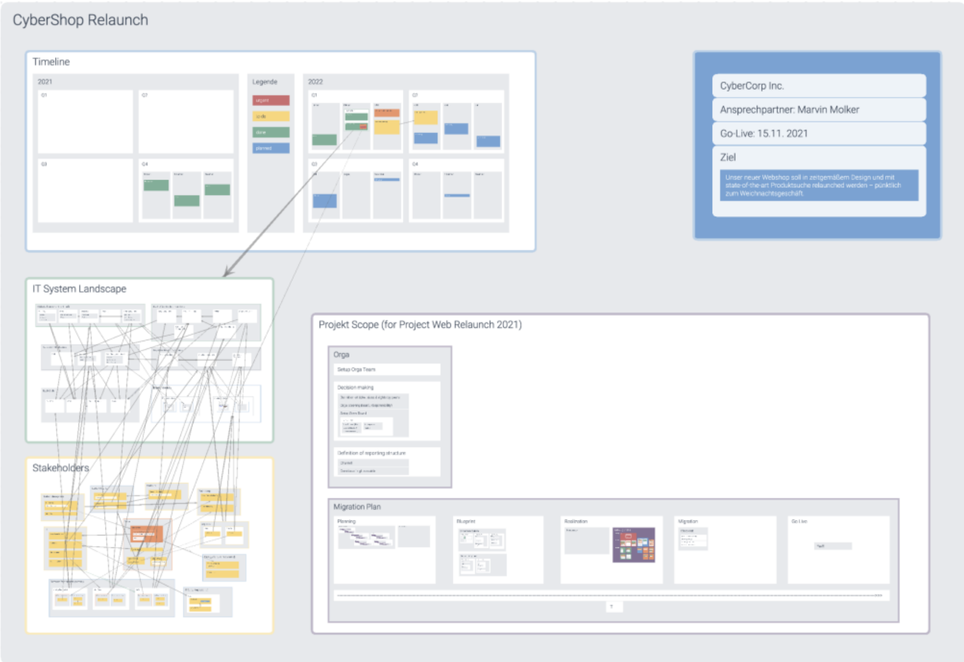


Fig. 7. A project management map with several interrelated areas like IT system landscape and stakeholders (anonymized example of an industry project).

Stakeholder Maps (like the one bottom left of Fig. 7,) can give an overview over the people involved in a project, their roles, responsibilities as well as their organizational and social relations. Especially in larger projects, where participants, responsibilities and even scope change over time, it is helpful, to have a medium that allows to adapt these structures as they evolve and as they are understood.

IT System Architectures on their own are often hard to grasp without visualizations. But in larger projects, an essential part of project management is to understand, who exactly is responsible for which aspect of a sub-system and how they communicate with each other. For that, it is beneficial to have both stakeholder and IT structures in one map, to see how they are interwoven. In Fig. 7, all these interrelations are displayed. However usually, they will set to be visible on demand only, to reduce visual complexity. Then, only interrelations of those cards are shown, that are currently selected or hovered upon.

Project Work Packages and their Dependencies – especially during planning phases of projects, Infinity Maps' hierarchical structure helps to break down work into packages. And interrelations help to visualize the dependencies between them, thus aiding the design of project roadmaps.

Process Documentation

Business processes are on the one hand often embedded in superordinate processes themselves. On the other hand, they consist of steps, each of which is a sub-process on its own that can be broken down to many single steps and tasks. To reduce complexity, most process visualizations only concentrate on one level of abstraction, leaving the others to other diagrams, tools or checklists, that live a life on their own, disconnected from one another.

Infinity Maps' approach of nesting and zooming however allows to smoothly transition between a macro view to descriptions of the single phases and right down to the very details.

Presentation and Teaching

Presenting topics in context and having the ability to transition between sub-topics through showing them in bedded in the big picture, makes Infinity Maps appealing to teachers, lecturers and trainers – much like the presentation tool Prezi.

We have also seen cases where pupils were given assignments buy a school teacher, to create a map about a given topic instead of an essay. Pupils reported to find this an engaging approach while liking the idea to learning to use a tool, that they may use for future, more challenging cases.

Other cases of school use include workshops, where pupils collaboratively fill and reorganize maps to collect research results, discuss them, collect problems and solution ideas, rank them in two dimensions, elaborate solutions and finally present them to a larger group by walking them through the map. Asked to contrast their working experience to working together on a paper canvas, some reported to prefer the electronic way of collaboration because it made it easier for them to control their flow of work – especially when other more dominant participants were occupying the auditive space.

Learning

Looking at the side of self directed learning, we have seen cases from pupils, who build maps on their own to organize their learning content and personal notes from vocabulary, to project material, as shown in Fig. 8 and 9.

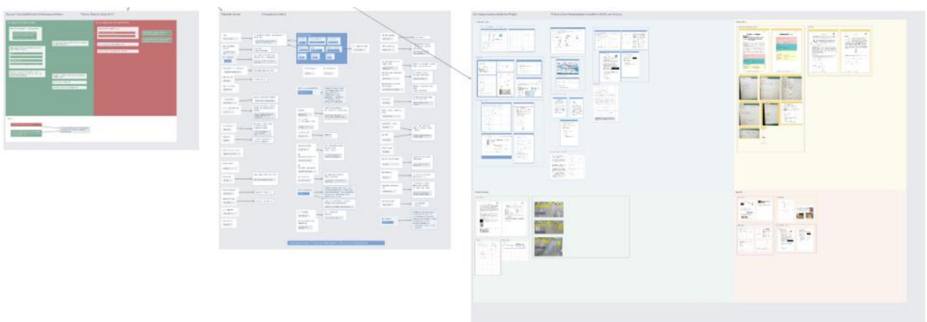


Fig. 8. Part of a 14 year old pupil's map showing a pro and contra analysis (left), a vocabulary map (middle) and an overview of project material (right).

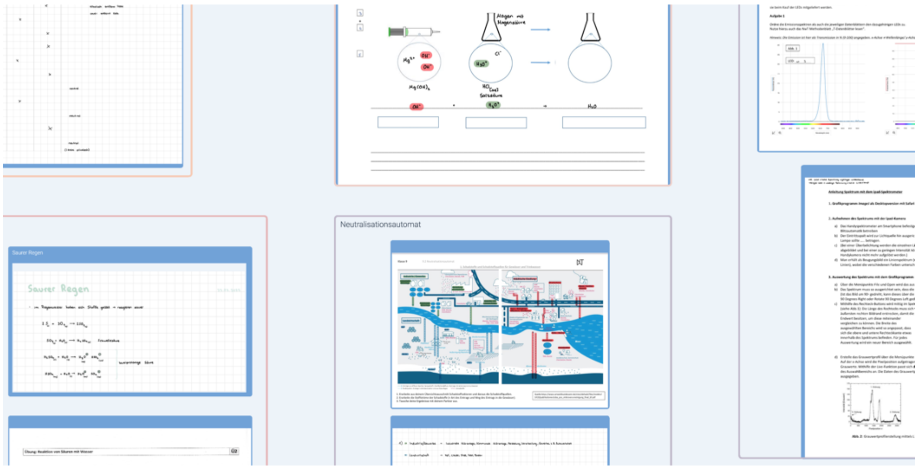


Fig. 9. Detail of the same map as Fig. 8, showing graphical learning material spatially arranged.

A learning practice seen in university students was to take a lecture's slides and spread them out on a map, where they could be arranged in a spatial layout that represents the content structure (similar to Fig. 9 and the right part of Fig. 8). This allows to see models, figures and charts in context rather than one by one only, which they reported to help their learning process.

Research

Finally, a strong use case that our user research has revealed, is *research* – both in the sense of scientific research, where, typically, many results are collected, evaluated and interrelated over a longer period of time, as well as web research in general, where many people report to get bogged down with up to 100 open browser tabs overwhelmed with the loss of overview and the inability to remember which of the intermediate results need further scrutiny.

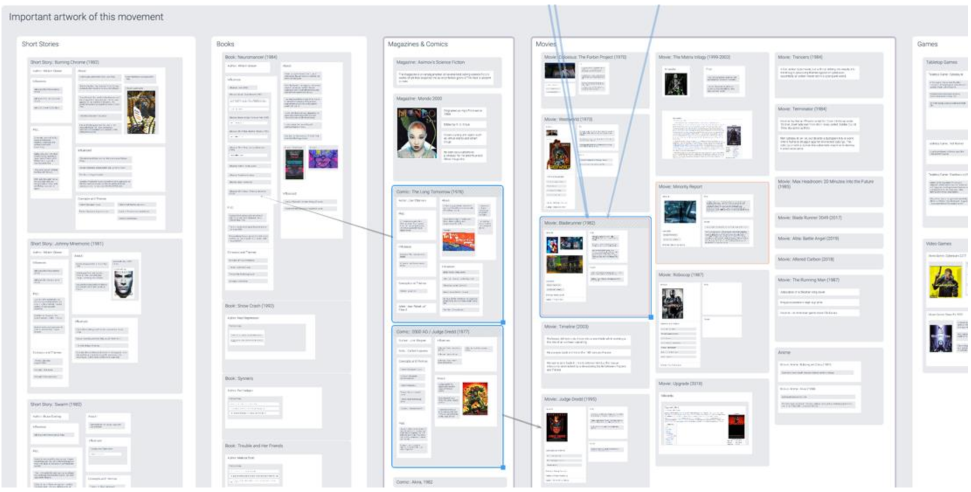


Fig. 10. Part of a research map about the topic of cyber punk.

5. Vision and Outlook

Unfortunately, since the initial evaluation of the iMapping technique (Haller, 2011), there has been no formal research on the effects and benefits iMapping and Infinity Maps, so all evidence reported is anecdotal only. Any scientific research is highly welcome to complement our current more pragmatic user research.

On our long road map to the future of visual knowledge work, some steps will be

- the introduction of mirror-cards to allow content to appear in more than one location
- the integration of infinity maps with more third-party systems to enable the use of more and more existing knowledge artifacts inside infinity maps
- efforts to ease the maintenance of certain structural layouts without the need of manual tweaking in order to further release cognitive load from the user.

It is our aspiration, to not only create a cognitive tool for individual knowledge workers and problem solvers but to provide a visual operating system for collaborative knowledge work of many kinds.

References

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24(1), 87–114. http://doi.org/10.1017/S0140525X01003922

Haller, H. (2011) *User Interfaces for Personal Knowledge Management with Semantic Technologies.* [Doctoral dissertation, Karlsruhe Institute of Technology]. Repository KITopen. https://www.doi.org/10.5445/IR/1000024007

Haller, H. (2003) *Mappingverfahren zur Wissensorganisation.* Unpublished Thesis. https://www.heikohaller.de/literatur/diplomarbeit

Haller, H. and Abecker, A. (2010) iMapping – A Zooming User Interface Approach for Personal and Semantic Knowledge Management. *Proceedings of the 21st ACM conference on Hypertext and hypermedia.* https://doi.org/10.1145/1850770.1836295 https://www.aifb.kit.edu/web/Inproceedings3060

Haller, H., Kugel, F. and Völkel, M. (2006). iMapping Wikis – Towards a Graphical Environment for Semantic Knowledge Management. In M. Völkel & S. Schaffert (eds.), *Proceedings of the First Workshop on Semantic Wikis*, June, 2006: ESWC2006.

Shipman, F.M., Marshall, C.C. (1999a). Spatial hypertext: an alternative to navigational and semantic links. *ACM Comput. Surv.* 31, 4es (Dec. 1999), 14–es. https://doi.org/10.1145/345966.346001

Shipman, F.M., Marshall, C.C. (1999b). Formality Considered Harmful: Experiences, Emerging Themes, and Directions on the Use of Formal Representations in Interactive Systems. *Computer Supported Cooperative Work* (CSCW) 8, 333–352. https://doi.org/10.1023/A:1008716330212

Track – Text Analytics I

An integrated workflow for exploring graduated research landscape using co-occurrence with focus+context visualization

CHIH-CHIEN HOU* JUNE-HAO HOU†

*,†*Graduate Institute of Architecture National Yang Ming Chiao Tung University Hsinchu, 300093, Taiwan*
**E-mail:chihchien@arch.nycu.edu.tw*
†*E-mail:jhou@arch.nycu.edu.tw*

Theses and dissertations, as a means to accumulate concepts, innovations, and implementations, are the main building blocks of the creative learning organization. In this research, we focus on a digital media-oriented graduate program in Architecture, where young researchers often derive their theses from various disciplines and encounter difficulties when exploring unfamiliar research areas. The aim of this paper is to provide a new way to explore the thematic map of knowledge generated in theses. The study proposes a bibliometric analysis method with a visualizing workflow embedded with focus+context visualization to support wayfinding in theses networks. In this study, we extracted terms from 95 design research theses across 2002 to 2022 in an applied knowledge base of a creative institution, and conducted automated data pre-processing to apply the co-occurrence analysis for thematic clusters. We present a framework around 3 abstract processes to design the interactive visualization for exploring multivariate information based on top-down and bottom-up perspectives. The results reveal that digital media-oriented theses papers cover 3 main topics and highlight the shifting pattern of the research interest. The integrated workflow can be applied to different academic works, such as syllabus, and projects, given the possibility to expand the knowledge and explore a more comprehensive graduated research landscape in the future.

1. Introduction

The multidisciplinary fields of research in the creative disciplines are changing dynamically (Gaona-García et al., 2017), especially in the context of theses and dissertations, which is a challenge for students and researchers to understand the contextual overview. The current search system normally handles lookup search tasks with an insufficient visual perspective. Researchers have to switch back and forth in the system repeatedly to gradually form an overview of different topics, rather than developing academic achievements that lead to the evaluation of new knowledge (Börner et al., 2005).

The techniques to visualize knowledge have been widely used in academic networks (Dur, 2014). However, with the evolving disciplines in digital architecture programs, users could traverse the information space without prior knowledge of the domain. Whereas theses and dissertations produce various representations in nodes and edges, thus forming the multivariate network (Shi et al., 2020). It is a challenge to examine the relationships between different entities in multivariate research networks. Furthermore, the decision- making pattern in a real-life wayfinding task could be similar to applying information- seeking strategies in the information space, which require navigators to conceptualize the space as a whole. An understandable structure that matches the state of knowledge of users can help them navigate within an unfamiliar concept of documents and investigate the core research topics.

Science Mapping Analysis (SMA) is an important technique in bibliometrics and has attracted scholars to analyze academic productions (Morris & Van der Veer Martens, 2009; van Eck & Waltman, 2017). However, many focus on collaborations and performances through large-scale documents, rather than analyzing conceptual maps from theses and dissertations in graduate schools (Chaparro & Rojas-Galeano, 2021), especially for multidimensional research papers with unstructured data and diverse concepts, making it difficult to organize them into meaningful clusters. Although most SMA tools allow to conduct many analysis results, those graphs are mostly static, which fails to interact with multivariate networks. How to improve usability to develop user-friendly interfaces that support different exploratory behavior remains an important challenge (Dattolo & Corbatto, 2022).

This study proposes an integrated framework focused on digital media-oriented programs in Architecture to make interdisciplinary themes easy to explore. Our goal is to design a knowledge representation that effectively supports students and researchers in exploring multiple facets of theses networks between overview and detail perspectives. We offer three exploration views shown in Fig. 1. Users can learn different thematic knowledge through an interactive dashboard, which allows users to:

- Explore and lookup different attributes in topics to identify the central entities in the network.
- Compare different topics or advisors' networks to analyze the similarities and differences between clusters.
- Analyze the trends of keywords in papers to examine the evolutions and major themes for future works.

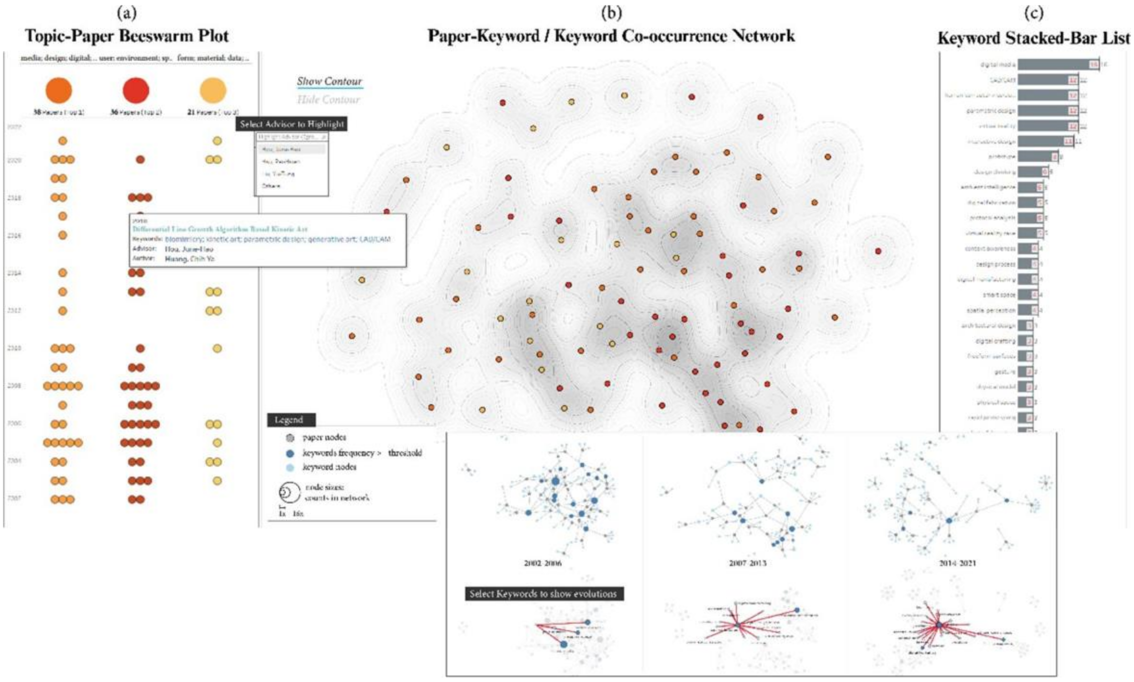


Fig 1. Dashboard overview. (a) A hierarchical overview of theses papers categorized in each topic from the perspective of time. (b) An exploratory network view to investigate the relationships between papers and keywords.
(c) A keyword ordered list present in a stacked bar chart that shows frequency of keyword appearances.

2. Design Framework

2.1. Overview

The framework is summarized based on the classic visualization model and the workflow of science mapping analysis (Card et al., 1999; Cobo et al., 2012). We focus specifically on multivariate based representations which categorize the workflow into different stages (Fig. 2) including data processing, derived data from analysis, visual representation, and user interaction. We further systematically organize workflows into three abstraction levels to design for theses explorations.

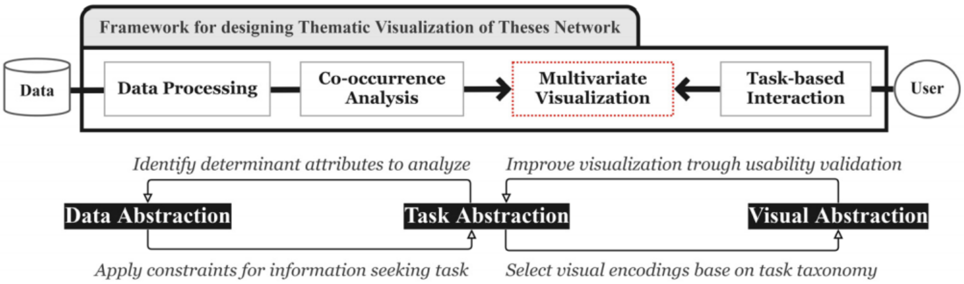


Fig. 2. The overall framework with three abstraction levels.

2.2. Data Abstraction

The data abstraction produces a scientific workflow from data cleaning, wrangling to data analyzing (Fig. 3). The study analyzed the master's theses in architecture published between 2002 to 2022, 328 papers were obtained from the digital library of theses and dissertations. To support exploring multi-dimensional research papers, our analysis then focused on 95 theses papers in digital media-oriented programs, which apply interdisciplinary knowledge and the systematic approaches to design research.

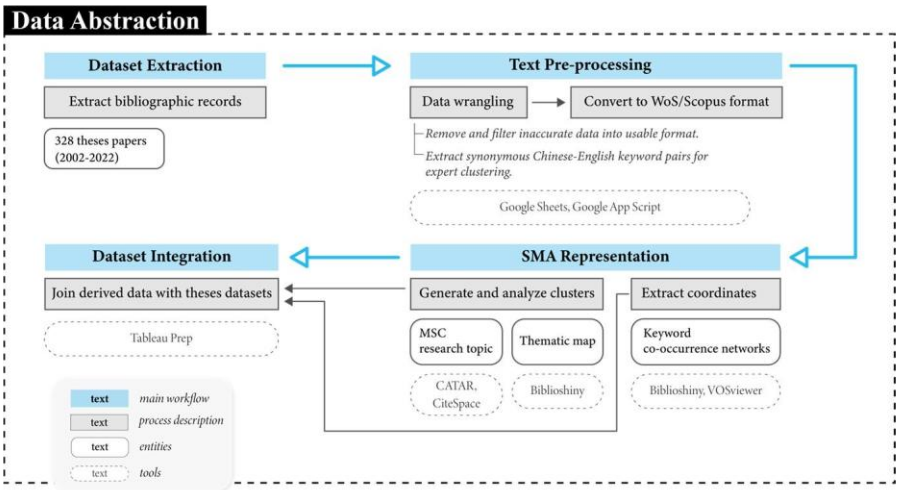


Fig. 3. Data abstraction framework.

2.2.1. Data Preprocessing

The process aims to improve the accuracy of the science mapping analysis and clustering results. The workflow is built within Google Sheets and Google App Script[1] for automatically cleaning data and enables future collaboration and management. After extracting results from SMA tools, we build the data processing pipeline using Tableau Prep Builder[2], which provides flexible connection to different data sources and can be used by various visualization software for implications.

2.2.2. Categorical attribute: Hierarchical clustering

We focus on co-word analysis to reveal the intellectual structure. We selected CATAR[3] for detecting thematic research topics with its multi-stage clustering (MSC) algorithms (Tseng & Tsay, 2013) which generate the effective cluster labeling extracted from titles, abstracts, and keywords. The results also enable us to associate clusters with the original theses datasets, which can be applied for more interactive approaches.

2.2.3. Quantitative attribute: Network coordinates

2D map is generated to illustrate the relative positions of the topics. We extract the x and y coordinates to further create an interactive node-link graph that visualizes the bibliographic networks.

2.3. Task Abstraction

In the framework, we design the information architecture to assist both lookup and exploratory tasks, which involve cognitive mechanisms based on how people fill the knowledge gaps. The task abstraction framework (Fig. 4) is inspired by the task-based information-seeking model (Kim, 2009) and can be divided into three stages: involve prior knowledge, conduct the task, and perform information seeking strategy.

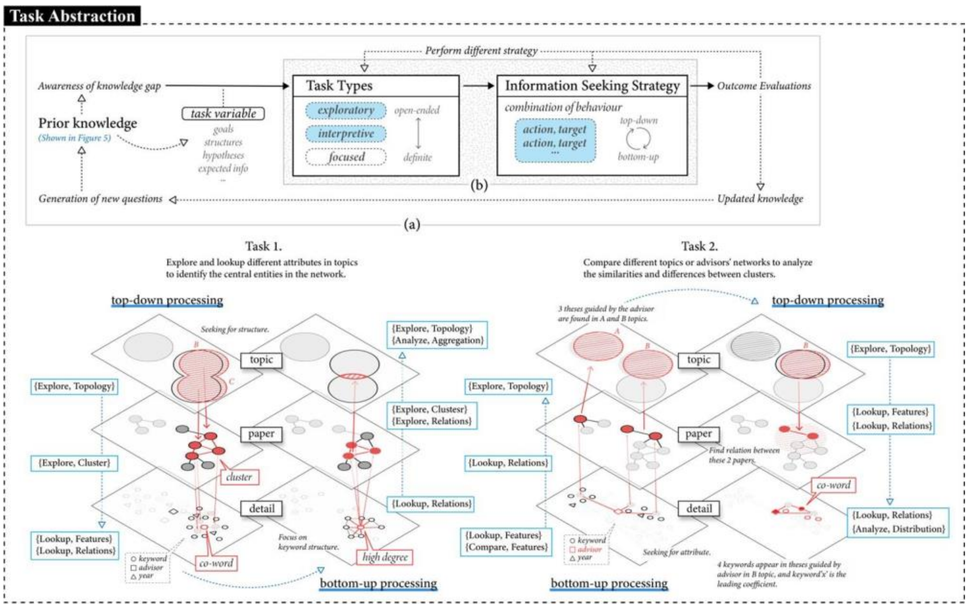


Fig. 4. Task abstraction framework. (a) Task based information seeking model for academic search. (b) Sensemaking processes with action-target pairs.

[1]https://developers.google.com/apps-script
[2]https://www.tableau.com/support/releases/prep
[3]https://web.ntnu.edu.tw/~samtseng/CATAR/

According to wayfinding taxonomy (Schwering et al., 2017; Wiener et al., 2009) and sensemaking mechanisms (Battle & Heer, 2019; Zhang & Soergel, 2014), we assume that people will perform top-down or bottom-up processing sequences to explore theses networks based on their prior knowledge of different design-related fields. Fig. 5 describes four scenarios and indicates that people will take approaches to fill their knowledge gap.

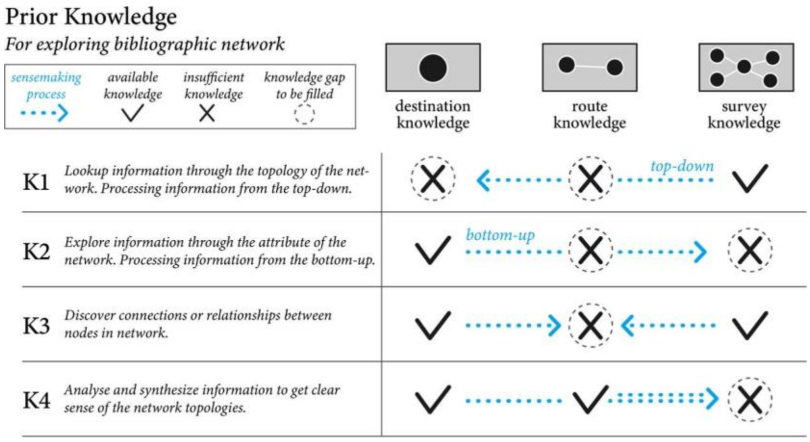


Fig. 5. Prior knowledge for exploring theses networks.

Past researches provided a systematic approach for multivariate network tasks (Nobre et al., 2019; Pretorius et al., 2014). We synthesize the tasks as clusters, paths, and nodes for academic networks and propose hierarchical ordering as the key information when exploring thesis papers (Fig. 4(b) and Fig. 7). Two different tasks are presented in Fig. 4(b), starting from different information processing, high-level hierarchical structures, and the behaviors of action-target pairs (Brehmer & Munzner, 2013).

3. Design Solution

Visual abstraction (Fig. 6) helps choose an appropriate encoding and interaction method that supports users to complete their task.

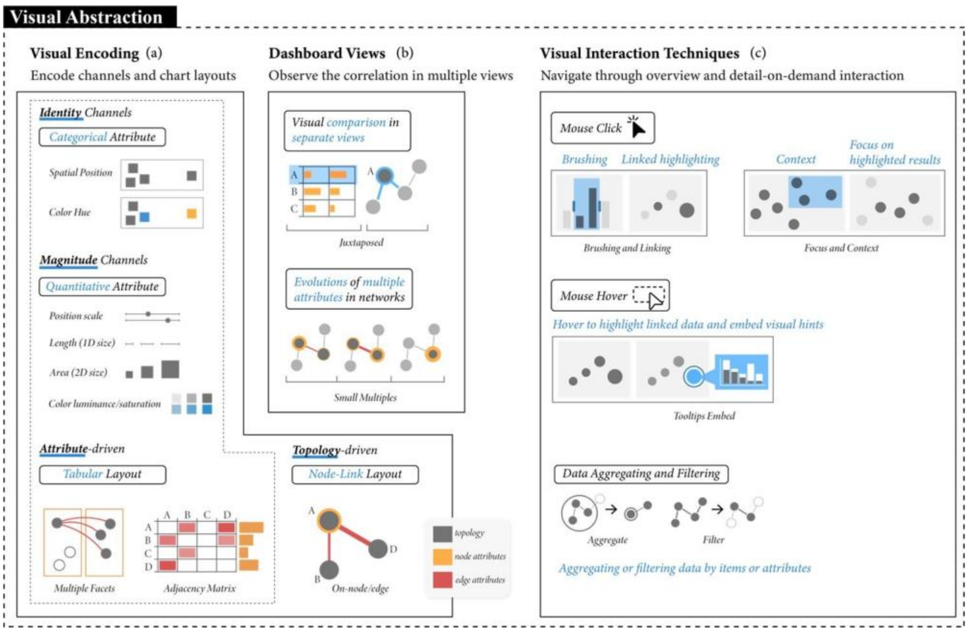


Fig. 6. Visual abstraction framework, adapted from (Munzner, 2014; Nobre et al., 2019). (a) Encoding channels and chart layouts. (b) Multiple views. (c) Navigate in multivariate networks.

In the visual encoding stage (Fig. 6(a)), we select the effective visual channels (Munzner, 2014) to identify and magnitude different attributes. The other guidance on designing multivariate network visualization can refer to Nobre et al (Nobre et al., 2019). One of our charts is topology-driven node-link layout to represent a co-word network. We also provide an attribute-driven facet and adjacency matrix to encode numerical attributes. The design considerations of the dashboard views (Fig. 6(b)) and interaction techniques (Fig. 6(c)) are followed by the information-seeking mantra (Shneiderman, 2003). Our approach is to juxtapose multiple views that visually compare variables within networks to see the information from different perspectives. We also provide small multiples to analyze the evolutions of theses networks. The dashboard offers brushing and linking techniques, which enable users to recognize data points across multiple views. This approach can refer to the focus+context visualization (Hauser, 2006), for example, highlighting the results while preserving context in grayscale to aid exploration (Fig. 7(b)). Tooltips are embedded to provide details-on-demand information when hovering over data.

To this end, we designed an exploratory dashboard shown in Fig. 1. Three linked views are built around the hierarchical structure (Fig. 7), which provides users with seamless navigation between the overview and local views to better understand research topics and focus on details of interest before reading the papers.

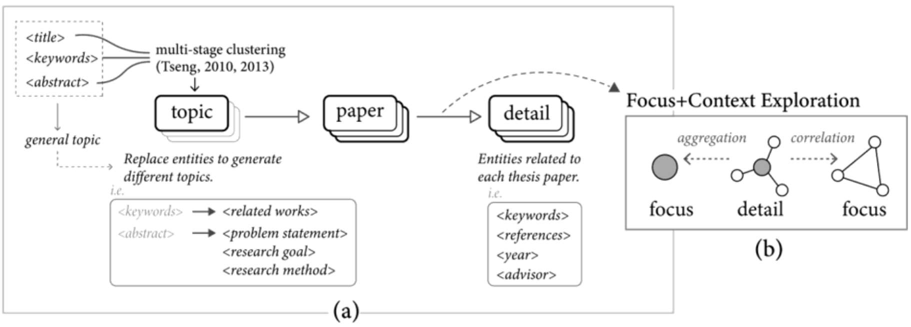


Fig. 7. The main structure to design our dashboard interface. (a) The hierarchical ordering on paper exploration views. (b) The focus+context visualization to explore theses networks.

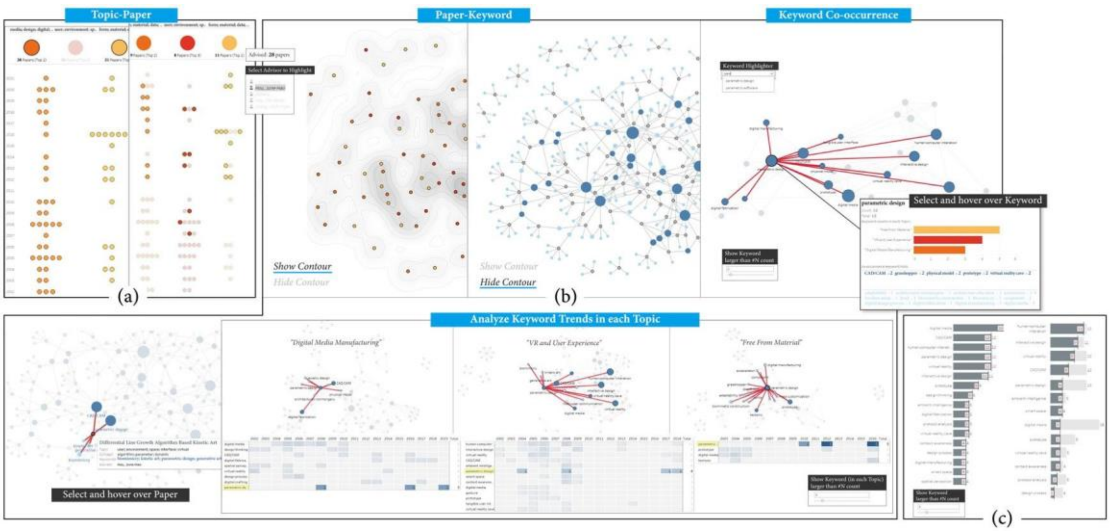


Fig. 8. Interaction across views. (a) Topic-aper comparison and advisor-paper relations. (b) Activate/deactivate the landscape view or the node-link diagram as well as filter the network with co-occurrence values threshold. (c) Stacked bars show topic-keyword for part-to-whole comparison.

3.1. Topic-Paper Beeswarm Plot

The beeswarm plot shows the overall theses papers across years clustering by their main research topics, which serve as the tables-of-contents (Fig. 8(a)). The topic menu is placed above the beeswarm plot as a navigation tool with the text to indicate the overall dominant topic. Brush selecting the topics will filter out irrelevant papers for comparison. We also list the advisors on the floating sidebar menu where users can hover over each advisor to highlight the advised papers. The count of papers will change dynamically to indicate the main research area for the advisor.

3.2. Node-Link Theses Networks

Two network views to capture the context of hierarchical relationships between papers and keywords, or switch layers to focus on keywords co-occurrence (Fig. 8 (b)).

Paper-keyword hierarchical topology network. We display only the paper nodes in the initial view with the contour-map overlayed on entities which enhances the correlation between each paper. The position channel shows the relations based on their author-specified keywords. Frequencies are size- and color-coded to indicate highly used keywords as well. Once the user clicks on a node/links, the corresponding nodes/links will be highlighted to enable users follow the paths and lookup the features.

Keywords co-occurrence network. The view enables users to analyze aggregation over many usages and investigate more application of each keyword. An additional selection menu is provided for users to lookup the keyword of interest. Hovering over keyword node opens a tooltip with the co-occurrence keyword lists juxtaposed with a bar chart showing the number of the target keyword appearing on different topics. Apart from the single view, we allow users to compare and analyze keyword trends using small multiple views with the selected matrices.

3.3. Keyword Bar Chart

When the user selects papers in other views, the corresponding keywords are counted, sorted, and encoded as the darker bar stacked on top of the bar chart (Fig. 8(c)). By interacting with the stacked bar graph, which clearly provides part-to-whole information, users can analyze the keyword distributions and compare the composition of the selected clusters.

4. Discussion and Future Work

In this paper, we proposed a conceptual framework built around three abstraction workflows that help designers transform theses networks into structural visualization. This framework addresses the problem of exploring and analyzing the thematic knowledge of multidisciplinary research.

The research topics are developed with a general approach, using only titles, abstracts, and keywords in bibliometric analysis. An in-depth analysis of the semantic scheme for making sense of scholarly knowledge will benefit researchers investigating specific aspects of research. Our integrated workflow could support this approach by extracting from other text fields, while automated processing and semantic technologies are needed to effectively derive the high-quality knowledge graph (Jaradeh et al., 2019). Furthermore, incorporating additional thematic overlays on the scientific community enables users to identify cutting- edge research, the evolution of cross-disciplinary knowledge, and to apply a more comprehensive knowledge landscape to explore in the future.

References

Battle, L., & Heer, J. (2019). Characterizing Exploratory Visual Analysis: A Literature Review and Evaluation of Analytic Provenance in Tableau. *Computer Graphics Forum*, *38*(3), 145–159.

Börner, K., Chen, C., & Boyack, K. W. (2005). Visualizing knowledge domains. *Annual Review of Information Science and Technology*, *37*(1), 179–255.

Brehmer, M., & Munzner, T. (2013). A Multi-Level Typology of Abstract Visualization Tasks. *IEEE Transactions on Visualization and Computer Graphics*, *19*(12), 2376– 2385.

Card, S. K., Mackinlay, J. D., & Shneiderman, B. (1999). *Readings in information visualization: Using vision to think*. Morgan Kaufmann Publishers.

Chaparro, N., & Rojas-Galeano, S. (2021). *Revealing the research landscape of Master's degrees via bibliometric analyses*. http://arxiv.org/abs/2103.09431

Cobo, M. j., López-Herrera, A. g., Herrera-Viedma, E., & Herrera, F. (2012). SciMAT: A new science mapping analysis software tool. *Journal of the American Society for Information Science and Technology*, *63*(8), 1609–1630.

Dattolo, A., & Corbatto, M. (2022). Assisting researchers in bibliographic tasks: A new usable, real-time tool for analyzing bibliographies. *Journal of the Association for Information Science and Technology*, *73*(6), 757–776.

Dur, B. I. U. (2014). Data Visualization and Infographics In Visual Communication Design Education at The Age of Information. *Journal of Arts and Humanities*, *3*(5), 39–50.

Gaona-García, P. A., Martin-Moncunill, D., & Montenegro-Marin, C. E. (2017). Trends and challenges of visual search interfaces in digital libraries and repositories. *The Electronic Library*, *35*(1), 69–98.

Hauser, H. (2006). Generalizing Focus+Context Visualization. In G.-P. Bonneau, T. Ertl, & G. M. Nielson (Eds.), *Scientific Visualization: The Visual Extraction of Knowledge from Data* (pp. 305–327). Springer-Verlag.

Jaradeh, M. Y., Oelen, A., Farfar, K. E., Prinz, M., D'Souza, J., Kismihók, G., Stocker, M., & Auer, S. (2019). Open Research Knowledge Graph: Next Generation Infrastructure for Semantic Scholarly Knowledge. *Proceedings of the 10th International Conference on Knowledge Capture*, 243–246.

Kim, J. (2009). Describing and predicting information-seeking behavior on the Web. *Journal of the American Society for Information Science and Technology*, *60*(4), 679– 693.

Morris, S. A., & Van der Veer Martens, B. (2009). Mapping research specialties. *Annual Review of Information Science and Technology*, *42*(1), 213–295.

Munzner, T. (2014). *Visualization analysis and design*. CRC press.

Nobre, C., Meyer, M., Streit, M., & Lex, A. (2019). The State of the Art in Visualizing Multivariate Networks. *Computer Graphics Forum*, *38*(3), 807–832.

Pretorius, J., Purchase, H. C., & Stasko, J. T. (2014). Tasks for Multivariate Network Analysis. In A. Kerren, H. C. Purchase, & M. O. Ward (Eds.), *Multivariate Network Visualization* (Vol. 8380, pp. 77–95). Springer International Publishing.

Schwering, A., Krukar, J., Li, R., Anacta, V. J., & Fuest, S. (2017). Wayfinding Through Orientation. *Spatial Cognition & Computation*, *17*(4), 273–303.

Shi, L., Liao, Q., Tong, H., Hu, Y., Wang, C., Lin, C., & Qian, W. (2020). OnionGraph: Hierarchical topology+attribute multivariate network visualization. *Visual Informatics*, *4*(1), 43–57.

Shneiderman, B. (2003). The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In B. B. Bederson & B. Shneiderman (Eds.), *The Craft of Information Visualization* (pp. 364–371). Morgan Kaufmann.

Tseng, Y.-H., & Tsay, M.-Y. (2013). Journal clustering of library and information science for subfield delineation using the bibliometric analysis toolkit: CATAR. *Scientometrics*, *95*(2), 503–528.

van Eck, N. J., & Waltman, L. (2017). Citation-based clustering of publications using CitNetExplorer and VOSviewer. *Scientometrics*, *111*(2), 1053–1070.

Wiener, J. M., Büchner, S. J., & Hölscher, C. (2009). Taxonomy of Human Wayfinding Tasks: A Knowledge-Based Approach. *Spatial Cognition & Computation*, *9*(2), 152– 165.

Zhang, P., & Soergel, D. (2014). Towards a comprehensive model of the cognitive process and mechanisms of individual sensemaking: Towards a Comprehensive Model of the Cognitive Process and Mechanisms of Individual Sensemaking. *Journal of the Association for Information Science and Technology*, *65*(9), 1733–1756.

Exploring Knowledge Graphs

JOHANNES HERCHER

Freie Universität Berlin, University of Applied Sciences Potsdam, Germany

PETER HEISIG

University of Applied Sciences Potsdam, Germany

GÜNTHER NEHER

University of Applied Sciences Potsdam, Germany

Conference abstracts and project reports can be a valuable source for access to information on possible risks or experiences to build on. However, these collections are often not accessible in a convenient manner. This work explores the usage of knowledge graphs to support document discovery with an automated indexing approach. Therefore, an RDF/SKOS graph is built, consisting of nearly 1400 tagging labels manually extracted from literature on knowledge graphs and organized in a thesaurus structure. The feasibility is demonstrated with a small collection of texts from the domain of library IT. The graph is published online for further use (cf. https://w3id.org/kgcm/). This method may be usable to explore other knowledge domains. On the other hand, the graph-building process is both intellectually and technically demanding and needs support by an interplay of cooperative and mechanical processes in the future.

Extractive Automatic Text Summarization Techniques for Afaan Oromoo – Afroasiatic Language in Ethiopia

RAMATA MOSISSA GICHILA*

Department of Information Science,
Jimma University, Mettu, Mettu, Oromia,
Ethiopia E-mail: mosissaebissa@gmail.com

DANIEL ALEMNEH+

Department of Information Science,
University of Pretoria, Pretoria, South Africa
E-mail: Daniel.Alemneh@up.ac.za

Text summary has become a vital and more popular domain to preserve and highlight the core purpose of textual information as the amount of online information and resource texts has grown. Text summarization is the task of extracting key information from a text document. Text summarizing research in Afaan Oromoo is still rare and hasn't been thoroughly assessed. This study's primary goal was to evaluate the performance and method of extractive models on automatic extractive text summarization for Afaan Oromoo. Automatic Text summarization approaches can be classified as extractive or abstractive. Automatic abstractive text summarization was not included in this study. Existing automatic extractive text summarizing algorithms take key sentences from the source manuscript and provide a summary without changing the data. This paper examined and assessed some studies on the Afaan Oromoo Language Text Summarization system with a focus on methods and performance. In addition, the automatic extractive text summarization domain's challenges in Afaan Oromoo's were also discussed. This paper used a systematic literature review method to examine the most recent literature in automatic extractive text summarization as it relates to the Afaan Oromoo language. We used a search engine, Google Scholar, ResearchGate, CiteseerX, peer-reviewed papers, and Academia to gather papers. After the papers were selected, the performance and automatic text summarization methods were evaluated.

Keywords: *automatic text summarization, extractive summarization, abstractive summarization, afaan oromoo language*

1. Introduction

1.1. Background of Text summarization

In the past, humans were used to summarizing the text in a different language in their own way, but in recent days because of the increasing volume of data, it is difficult for human beings to navigate the huge amount of data. To control this problem, text summarization is needed. Text summarization is extremely active in the field of machine learning and generates clearly expressed forms of large documents keeping most of the original information. It is carried out on single or multiple documents of a similar kind. Text summarization has attracted the attention of researchers from 1950 to today. Luhn was the first researcher of computer science and information science who made an automatic text summarization system. Luhn, (1958) specified the advantage of words depends on their frequency measures. He claimed that words that appear frequently in a text give a good notion of the document's content, even if there are terms that appear frequently but are not content bearing.

Text summarization is the practice of distilling the most significant information from a source (or sources) to produce a condensed version for specific users and tasks (Lin & Chen, 2010). It is a tool that can assist users in sorting through a large amount of data and can be carried out by human experts on a certain topic, but it is time-consuming and costly (Regassa, 2017). In today's fast-growing information era, text summary becomes a crucial and timely instrument for assisting and taking text information (Jezek & Steinberger, 2008).

Condensing the original text into a shorter version while keeping the original message of the content is known as text summarization. Manually summarizing massive amounts of text is extremely challenging for humans (Gupta & Lehal, 2010).

Automatic textual content summarization has been considerably utilized in numerous fields like science, medicine, law, engineering, etc. One example is researchers have targeted producing summaries of doctors' prescriptions, and that has proved very beneficial to patients. Similarly, lengthy information articles have been summarized and readers can consume multiple records on numerous subjects inside a brief span.

Different text summarizations depend on the form of summary we need to be produced from sources. There are two categories of techniques: extractive and abstractive. Extractive summarization is a method that identifies important text and reproduces the summary from original documents. The summary produced from this approach is completely selected from the sentences or sequence of words in the original document (Alguliev & Aliguliyev, 2009) In addition to the complete sentences, extractives can have paragraphs and paraphrases. The problem with the extractive approach is that it produces incoherent summaries from original text or lacks a sense of balance and consistency. In this approach, the extraction of sentences can be out of the context meaning and anaphoric references can also be broken (Rejhan et al., 2009).

When we come to abstractive summarization, it can compose novel sentences which are even unseen in the original sources or original document. This approach is usually built from the existing content but uses advanced methods. However, abstractive approaches require deep natural language processing such as inference; semantic representation, and natural language generation (Alguliev & Aliguliyev, 2009). When the abstraction approach is used for text summarization in deep learning problems, it can solve the grammar inconsistencies of the other approach or extractive method.

Afaan Oromoo is a prominent African language that is widely spoken and utilized in much of Ethiopia, as well as some sections of neighboring Kenya and Somalia (Abera, 1988). It is spoken by the Oromo, Ethiopia's largest ethnic group, accounting for 34.5 percent of the country's population. A few individuals of other ethnicities, who come into contact with the Oromos, like the Omotic-speaking Bambassi and the Nilo-Saharan-speaking Kwama in northwestern Oromia, speak it as a second language. Currently, Afaan Oromoo is the official language of the Oromia Regional State.

1.2. Problem Definition

The history of research work on text summarization started in the 1950s using the extractive system approach (Luhn, 1958). He proposed that words existing many times in a text provide a good idea about the content of the document though there are words that appear very frequently but are not content bearing. As a result, he cut off these words by determining a fixed threshold. The idea of the researcher was recognized and used in many automatic information processing systems. The system uses properties like term filtering and word frequency. Luhn carefully observed that the location of a sentence in a text gives some hint about the importance of the sentence. Thus, he suggested word frequency, cue phrases, title and heading words, and sentence location as extraction features.

Digital information in Afaan Oromoo is growing since Afaan Oromoo became the official language of the Oromia regional state. Currently, there are several newspapers, articles, printed media, and other news releases in Afaan Oromo. There are sources of newspapers like Bariisaa, Kallacha OromiyaA, and radio broadcasts in Afaan Oromoo by Ethiopian Radio, Radio Fana, OBN, and VOA Afaan Oromoo. Also, magazines, judiciary documents, and office reports constitute some portion of the documents produced in the language.

Currently, some study in an automatic artificial text summarization has been conducted for Ethiopian languages, mainly for Amharic text in many areas using various methodologies, and a few contributions have also been made for the Afaan Oromoo language on plagiarism detection, sentiment analysis, disambiguation, and extractive automatic text summarizer using traditional analysis (traditional algorithm). Few studies have compared extractive

and abstractive document summarization (Carenini et al., 2006). Findings indicate that extraction and abstraction worked roughly as well as in other languages. The current effort is a contribution to the development of Ethiopian language natural language processing applications, and it expands the area of text summarizing study by looking into its applicability to the Afaan Oromoo language. Therefore, the objective of this paper is to review the literature on the performance, methods, and challenges of extractive automatic text summarizers for Afaan Oromoo.

1.3. Research questions

In the end, this research project will attempt to answer the following study questions:

What is the performance of automatic extractive text summarization for Afaan Oromo?

What are the challenges in Afaan Oromoo automatic extractive text summarization?

What are the methods used in automatic extractive text summarization?

2. Methodology

A systematic literature review was conducted in automatic text summarization for Afaan Oromoo Language. We used a search engine, Google Scholar, ResearchGate, CiteseerX, peer-reviewed papers, and Academia to gather papers. Only studies published after 2013 were considered in the review of papers on Afaan Oromoo text summarization. After the papers were selected, the performance and automatic text summarization methods were evaluated.

3. Literature Review

3.1. Text Summarization

Text summarization research may be traced back to the 1950s when the first extraction system was created (Luhn,1958). He claimed that words that appear frequently in a text give a good notion of the document's content, even if there are terms that appear frequently but are not content bearing. As a result, they attempted to eliminate these terms by establishing a defined threshold. Many automatic information processing systems recognize and use Luhn's concept. It is a technical article summarizing system with a domain-specific single document summarization system. Term filtering and word frequency are among the features used by the algorithm (low-frequency terms are removed). Sentences are weighted according to the significant terms in them, and sentence segmentation and extraction are conducted. He meticulously defined the human extraction principles, noting that the placement of a sentence in a text provides some insight into its relevance. As an extraction feature, he offered word frequency, cue phrases, title and heading terms, and sentence location. Edmundson's system (Edmundson, 1969), like Luhn's, is a single document that is domain-specific (that deals with technical articles). Different text summarizations depend on the form of summary we need to produce from sources. There are two categories of techniques: extractive and abstractive.

3.1.1. Extractive Text Summarization

The extractive summarization method selects important sentences from the original text document and connects them into a shorter form without changing or altering the main text

(Shivangi & Rachana, 2018). According to Kylmenko et al. (2020), extractive summarization methods create summaries by concatenating numerous sentences (text units) from the text to be summarized in the exact order in which they appear. The fundamental goal of these systems is to figure out which sentences are essential enough to include in the summary. For many years, extractive methods have been the primary focus of text summarization researchers. Many recent approaches treat extractive summarization as a sequence labeling task, with each label indicating whether a sentence should be included or excluded from the summary.

3.1.2. Abstractive Text Summarization

According to Pai (2014) natural language processing, semantic representation and modification, text interpretation and production are all topics covered by abstraction summarization. The goal of the abstractive method is to extract essential sentences as a type of computed document summary. This can be accomplished in several ways. Structured and semantic-based approaches to abstractive text summarization are examples. Structured approaches use a variety of schemas to encode important aspects of documents, such as a tree, ontology, lead and body phrases, and template and rule-based schemas, whereas semantic-based approaches are more concerned with the text's semantics and thus rely on the document's information representation to summarize the text. The multimodal semantic method, information item technique, and semantic graph-based method are all semantic-based methodologies (Suleiman & Awajan, 2020).

3.2. Single Document Summarization

Single document text summarization is a summary from a single source document, using single document text summarization (Shivangi & Rachana, 2018). This sort of text summarization technique takes only one document as input, then employs several techniques to extract relevant sentences from the source document, after which a summary is constructed from the retrieved sentences. Summary generation is more comprehensible, syntactically, or semantically correct, and, most importantly, in a reduced form. To minimize repetitions when summarizing various papers, one must identify and locate topic overlaps. It is also required to decide what to do with the rest, deal with potential document contradictions, and, if necessary, arrange events from diverse sources along a single timeline.

3.3. Multi-Document Summarization

The goal of multi-document text summarizing is to create a summary from multiple source documents. The goal of multi-document text summarizing is to extract useful information from each source document and then create a summary that meets the needs of humans (Modi & Oza, 2019). To create the new summary, extractive text summarization involves selecting terms and sentences from the base text. Procedures entail rating the importance of phrases to select only those that are most relevant to the source's implication (Asawa et al., 2020).

3.4. Extractive Summarization Methods

The goal of extractive summarizers is to pull out the most important information by highlighting the most important sentences in the document while also keeping the summary's repetition to a minimum.

3.4.1. Term Frequency-Inverse Document Frequency (TFIDF) method

According to Zhang et al. (2005), the term frequency-inverse document frequency (TF-

IDF) of scores of words is used to denote documents. In this instance, the term frequency refers to the average number of occurrences (per document) across the cluster. The IDF value is calculated using the complete corpus. The summarizer accepts texts that have previously been clustered as input. Each cluster is referred to as a theme. Words with the highest term frequency and inverse document frequency (TF-IDF) scores in that cluster represent the theme.

The conventional weighted term-frequency and inverse sentence frequency paradigm are used to build the bag-of-words model at the sentence level, where sentence frequency is the number of sentences in the document that contain that term. These sentence vectors are then ranked according to their closeness to the query, with the highest-scoring sentences being selected for inclusion in the summary. This is a straightforward application of the information retrieval paradigm to summary.

Nonstop words that occur most frequently in the document(s) can be used as query terms to build a generic summary. These words provide generic summaries since they describe the document's theme. For sentences, term frequency is usually 0 or 1, because the same

content word does not present frequently in a sentence. If users create query words in the same way they do for information retrieval, query-based summary production will become generic.

3.4.2. Cluster-Based Method

According to Shiva Kumar and Soumya (2015), documents are often constructed in such a way that they address several themes one after the other in a logical order. They are usually divided into sections, either expressly or implicitly. This categorization should be applied to summaries as well, which should address various "themes" that arise in the texts. Some summarizers use clustering to add this feature. Document clustering becomes almost required to build a coherent summary when the document collection for which the summary is being prepared is of completely distinct themes.

The graph-theoretic approach combines related sentences using both sentence ranking and clustering, which are both employed in graph models. Singular nonmatrix factorization is used to cluster sentences in the text. Finally, while clustering and ranking sentences in a document, the weighted graph model technique utilized in this approach examines the discourse relationship between sentences. Hariharan and Srinivasan, (2009) studied a way of summarizing news stories using a Graph-Based approach 2009. An adjacency matrix, which is the cornerstone of Graph-Based methods, is used to represent the measure of similarity between phrases in this method. In this work, two different strategies are studied. The authors' first suggestion was the cumulative sum approach. The second strategy studied was the degree of centrality, which was a previously existing method. A novel way for evaluating the adjacency matrix was proposed in this paper using the aforementioned two techniques, which add two metrics: Effectiveness 1 and Effectiveness 2. These are useful for comparing system summaries to human summaries. Comprehensive studies have shown that this method is superior to traditional methods and that there is still room for advancement in this field of text summarization.

3.4.3. Latent Semantic Analysis Method

According to Ozsoy et al. (2011), Latent Semantic Analysis (LSA) is a statistical-algebraic method for extracting latent semantic structures in words and sentences. It is an unsupervised method that does not necessitate any training or prior knowledge. LSA gathers information from the context of the input document, such as whether words are used together, and which common terms appear in different phrases. The presence of a large number of common terms between sentences implies that they are semantically connected. The meaning of a sentence is determined by the words in it, and the meanings of words are determined by the sentences in which they appear. The interrelationships between sentences and words are discovered using Singular Value Decomposition, an algebraic approach. In addition to being able to model links between words and phrases, single value decomposition can also reduce noise, which helps to increase accuracy.

3.4.4. Machine Learning Approach

According to Sirohi et al. (2021), these methods transform the unsupervised summarizing task into a supervised classifying task that works on the text. The sentence from the inputted document is categorized as "instant" (summary) or "non-instanced (non-summary)" using an algorithm trained from examples and a document with training sets (i.e a set of documents and their numerous summaries generated by humans). The goal of the machine-learning-based summarization method is to score the sentence Query-based

3.4.5. Query-based based extractive text summarization

According to Rajendran and Shravan (2020), the sentences in a given document are graded based on the frequency counts of the query-based text summarization system (words or phrases). The sentences that contain query phrases receive a better score than those that only contain single query terms. The highest-scoring sentences are then included in the output summary, together with their structural context. Text fragments can be taken from various parts or subsections. The resultant summary is the result of combining these

extracts. The number of extracted sentences and the extent to which their context is displayed is determined by the summary frame size, which is set to the maximum screen size that can be viewed without scrolling. When a sentence is chosen for inclusion in the summary by the sentence extraction algorithm, some of the heads in that context are also chosen.

4. Related Work in Afaan Oromoo Language

Dinegde and Tachbelie (2014) tried constructing an Open Oromo Text Summarizer (OOTS). The reference summary was created and used to assess the system output's performance (system summary). An intrinsic technique was used to accomplish the evaluation. They built three approaches for summarizing Afaan Oromoo news material and examined their performance both objectively and subjectively in this study. S1 is a summarizer that uses term frequency and position methods without using the Afaan Oromoo stemmer and other lexicons (synonyms and abbreviations), S2 is a summarizer that uses a combination of term frequency and position methods with the Afaan Oromoo stem specific language language-specific lexicons (synonyms and abbreviations), and S3 is a summarizer that uses improved position method and term frequency as well as (synonyms and abbreviations). Its summarizing mechanism is based on sentence extraction. It was utilized for the phases of processing, sentence ranking, and summary generation. Two evaluation types were used: subjective and objective. For a summary derived using three distinct methodologies, subjective and objective evaluations were undertaken. Both evaluations' findings are consistent; in every example, the S3 summary outperformed the other techniques.

The average in formativeness of the summary is (34.37%, 37.5 %, and 62.5 %) for the three methods (S1, S2, and S3), respectively; the average language quality is (59.37 %, 60 %, and 65 %), and the average coherence and structure are (21.87 %, 28.12 %, and 75 %). The objective assessment result, on the other hand, demonstrates that the average f-measure score for (S1, S2, and S3) is (34 %, 47 %, and 81 %). Dinegde and Tachbelie (2014) concluded that the results of the study's summarizers, like those of previous extraction-based summarizers, lack consistency. Based on the result they recommended abstract summarization more advanced way of avoiding such an issue.

Kannaiya Raja et al. (2019) tried to create Afaan Oromoo Text Summarization by Frequency and Sentence Position Methods. To determine the most significant sentence for extracting a summary, the summarizer combines term frequency and sentence position algorithms with language-specific lexicons. The extraction technique for single news text is the technique proposed for this study. The most significant sentences from the document are extracted and shown to the reader using the extraction process. They aimed to create an algorithm that can summarize a document in the Afaan Oromoo language based on its performance, both objectively and subjectively. The technique proposed for this study was the extraction technique for single news text. Using the extraction technique most important sentences from the document are extracted and displayed to the reader. Term frequency and sentence position algorithm language-specific lexicons were employed in this paper to provide weights to the sentences to be extracted for the summary. An intrinsic method was used to accomplish the evaluation. It included subjective (qualitative) as well as objective (quantitative) evaluation techniques. The four human subjects (expert journalists) are involved in both measurements. The linguistic quality, informativeness, and coherence of the automatically generated summaries were assessed using subjective judgment.

The recall and precision criteria were used to assess its performance. It compares the extracts to the reference summary given an input text, a human's (reference) summary, and a summarizer's extract. Tokenizing, stop-word elimination, stemming, and parsing are all part of the preprocessing steps (breaking the input document into a collection of sentences). Following the formatting and stemming of an input document, the document is broken down into a series of sentences, which are then rated based on two key characteristics. Sentence position and term frequency (TF). This paper missed the finding or its performance and recommendation for the prospects.

Based on the user's inquiry, Jilo et al. (2021), developed a document summary for the Afaan Oromo language. The TF-IDF word weight methodology was employed in the development of the query-based architecture. For morphological analysis, development tools such as HornMorpho are used, whereas, for text processing, Natural Language Processing Toolkit is used. The system has experimented with different extraction rates of 10%, 20%, and 30%. For objective analysis, recall, precision, and f-measure were evaluated; subjective analysis was evaluated by language consultants. At a summary extraction, the proposed system registered f-measures of 90%, 91%, and 93%, according to the results of the evaluation. The performance of the system improved to 91.3% F-measure by applying a morphological analysis tool, even though the additional study is still required to improve the Afaan Oromoo text summary.

Gemechi (2021) attempted to make at Afaan Oromo News Text Summarization Using Sentence Scoring Method. The researcher conducted experiments on ten selected topics from a total of 30 gathered topics using the extractive approach. The natural language tool kit (NLTK) created the system using the Python programming language. The developed approach computes the sentence's score by summing the scores of each word and then computing the sentence's score. The method creates the summary by extracting top-scoring sentences at three different extraction rates: 20%, 30%, and 40%. Automatic summary performed 74%, 78%, and 86% in terms of informativeness at 20%, 30%, and 40% extraction rates. The system performed 86.1 % of the objective evaluation with the three metrics recall, precision, and F-score generated. As a result, a summarizer with better performance in this work. However, evaluating the three metrics of precision, recall, and f-measure during a summary review objectively is difficult. The author recommends new features and future research prospects.

Table 1. Comparison of research performance in Afaan Oromoo

S/N	Author/s	Extractive Method/s	P	R	F
1	(Dinegde & Tachbelie, 2014)	preprocessing, sentence ranking, and summary generation	0.54	0.54	0.54
2	(Kannaiya Raja et al., 2019)	Frequency and Sentence Position Methods	missed	missed	missed
3	Jilo et al. (2021)	TF-IDF term weight Method	missed	missed	0.93
4	(Gemechis,2021)	Sentence scoring Method	0.861	0.861	0.861

5. Challenges and Future Research Directions for Extractive Text Summarization

It is difficult to evaluate summaries (either automatically or manually). The fundamental issue with assessment is the inability to create a standard against which the findings of the systems can be compared. Furthermore, determining a proper summary is difficult due to the possibility that the system will provide a superior summary that differs from any human summary that is used as a rough approximation to correct output. According to Lalithamani et al. (2014), the problem contents elections has not been solved. Because people are so different, subjective authors may choose completely different sentences. Paraphrasing is the process of combining two or more sentences expressed in different languages to represent the same concept. There is a method for evaluating summaries automatically using paraphrases (para Eva).

Extractive summarizing (selecting and replicating lengthy sentences from professional publications) is used by the majority of text summarization s systems. Even though humans may cut and paste important material from a text, they frequently rewrite sentences or combine similar facts into a single statement. The low levels of inter-annotator agreement reported during manual evaluations suggest that the future of this field of research is heavily reliant on the ability to develop efficient methods for automatically evaluating systems.

6. Conclusion

This review has revealed a variety of mechanisms for extractive text summarizing. The

process of extractive summarization is highly coherent, less redundant, and cohesive (summary and information-rich). The goal is to provide a thorough examination and comparison of various methodologies and strategies for extractive text summarization.

Even though summarization research has been going on for a long time, there is still a long way to go. In large-scale applications, simple sentence removal has shown satisfactory results. Some trends in automatic summary system evaluation have been highlighted. However, in the context of time and space difficulty, the work has not fully focused on the various issues of extractive text summarization. Extractive text summarization work carried out in Afaan oromoo language is very limited when we compare it with other languages spoken by more than 40 million similar to Afaan Oromoo.

References

Abera N. (1988). Long vowels in Afan Oromo: A generic approach, Master's thesis, School of graduate studies, Addis Ababa University, Ethiopia.

Alguliev, R., & Aliguliyev, R. (2009). Evolutionary Algorithm For Extractive Text Summarization. *Intelligent Information Management*, *01*(02), 128–138. Https://Doi.Org/10.4236/Iim.2009.12019

Asawa, Y., Balaji, V., & Dey, I. I. (2020). *Modern Multi-Document Text Summarization Techniques*. *1*, 654–670. Https://Doi.Org/10.35940/Ijrte.A1945.059120

Dinegde, G. D., & Tachbelie, M. Y. (2014). Afan Oromo News Text Summarizer. *International Journal Of Computer Applications*, *103*(4), 975–8887.

Edmundson, H. P. (1969). New Methods In Automatic Extracting. *Journal Of The Acm (Jacm)*, *16*(2), 264–285. Https://Doi.Org/10.1145/321510.321519

Gupta, V., & Lehal, G. S. (2010). A Survey Of Text Summarization Extractive Techniques. *Journal Of Emerging Technologies In Web Intelligence*, *2*(3), 258–268. Https://Doi.Org/10.4304/Jetwi.2.3.258-268

Hariharan, S., & Srinivasan, R. (2009). Studies On Graph-Based Approaches For Single and Multi-Document Summarizations. *International Journal Of Computer Theory And Engineering*, *1*(5), 519–526. Https://Doi.Org/10.7763/Ijcte.2009.V1.84

Jezek, K., & Steinberger, J. (2008). Automatic Text Summarization (The State Of The Art 2007 And New Challenges). *Proceedings Of Znalosti*, 1–12.

Jilo, J. A., Alemu, A. T., Abera, F. Z., & Rashid, F. (2021). An Integrated Development Of A Query-Based Document Summarization For Afaan Oromo Using Morphological Analysis. *Indian Journal Of Science And Technology*, *14*(38), 2946–2952. Https://Doi.Org/10.17485/Ijst/V14i38.1182

Kannaiya Raja, N., Bakala, N., & Suresh, S. (2019). Nlp: Text Summarization By Frequency And Sentence Position Methods. *International Journal Of Recent Technology And Engineering*, *8*(3), 3869–3872. Https://Doi.Org/10.35940/Ijrte.C5088.098319

Lalithamani, N., Sukumaran, R., Alagammai, K., Sowmya, K. K., Divyalakshmi, V., & Shanmugapriya, S. (2014). A Mixed-Initiative Approach For Summarizing Discussions Coupled With Sentimental Analysis. *Acm International Conference Proceeding Series*, *10-11-Octo*. Https://Doi.Org/10.1145/2660859.2660910

Lin, S. H., & Chen, B. (2010). A Risk Minimization Framework For Extractive Speech Summarization. *Acl 2010 - 48th Annual Meeting Of The Association For Computational Linguistics, Proceedings Of The Conference*, *April*, 79–87.

Modi, S., & Oza, R. (2019). Review On Abstractive Text Summarization Techniques (Atst) For Single And Multi Documents. *2018 International Conference On Computing, Power And Communication Technologies, Gucon 2018*, 1173–1176. Https://Doi.Org/10.1109/Gucon.2018.8674894

Ndyalivana & Shibeshi, 2019 Automatic Text Summarization Using an Advanced Stemmer Algorithm: A Case Study of the Xhosa Language

Ozsoy, M. G., Alpaslan, F. N., & Cicekli, I. (2011). Text Summarization Using Latent Semantic Analysis. *Journal Of Information Science*, *37*(4), 405–417. Https://Doi.Org/10.1177/0165551511408848

Rajendran, S., & Shravan, V. (2020). Query-Based Text Summarization Using Averaged Query. *International Journal Of Psychosocial Rehabilitation*, *July*. Https://Www.Researchgate.Net/Publication/342638073%0aquery-Based

Regassa, M. G. (2017). *Topic-Based Tigrigna Text Summarization Using Wordnet A Thesis Submitted To The Department Of Computer Addis Ababa University College Of Natural Science*.

Luhn H. P. (1958), "The Automatic Creation of Literature Abstracts", pp. 159-165.

Shiva Kumar, K. M., & Soumya, R. (2015). Text Summarization Using Clustering Technique And Svm Technique. *International Journal Of Applied Engineering Research*, *10*(10), 25511–25519.

Sirohi, N. K., Bansal, D. M., & Rajan, D. S. N. R. (2021). Text Summarization Approaches Using Machine Learning & Lstm. *Revista Gestão Inovação E Tecnologias*, *11*(4), 5010–5026. Https://Doi.Org/10.47059/Revistageintec.V11i4.2526

Suleiman, D., & Awajan, A. (2020). *Deep Learning-Based Abstractive Text Summarization: Approaches, Datasets, Evaluation Measures, And Challenges*. *2020*.

Zhang, Y., Zincir-Heywood, N., & Milios, E. (2005). Narrative Text Classification For Automatic Key Phrase Extraction In Web Document Corpora. *Proceedings Of The International Workshop On Web Information And Data Management*

Wide, 51–58. Https://Doi.Org/10.1145/1097047.1097059

Track – Text Analytics II

Dialogue Recognition in Online Health Community via Integrating Text Scene Information

Zhanhua Tang*

School of Information Management,
Wuhan University, Wuhan,
Hubei,430072, China

Renxian Zuo*, Chaocheng He†, Haoyu Zhou, Xiao Huang, Jiang Wu

School of Information Management,
Wuhan University, Wuhan,
Hubei,430072, China

**Zhenhua and Renxian contribute equally.*

† *E-mail: chaochehe@whu.edu.cn.*

Abstract: WeChat group-based online medical community (WGMC) is increasingly accepted by the public because of its high efficiency, convenience, and shared advantages in seeking medical resources. However, the problem to separate a complete dialogue relationship from the chat records is arising because efficient online community management is based on clear dialogue relationships and clear topics. To solve the problem, we proposed a hybrid three-stages BERT method to recognize the dialogue relationships in "Home of Love" --- a central nervous system tumor online healthy community WeChat group. First, based on the social support theory, a multi-layer BiLSTM model is proposed to classify the conversation scenes into five classes. Then, two domain adaptation methods for transfer learning are designed to optimize the BERT pre-training model for specific tasks using the "Haodaifu" as the training corpus. Finally, a hybrid BERT method based on the text scene information and the pre-training model is proposed to recognize the dialogue relationships, and its feasibility is verified by manual labeling. The results show that adding more prior knowledge to the dialogue recognition model by extracting the social support scene information can effectively improve the classification ability and stability of the model.

1. Introduction

Online medical communities are garnering increasing social and scholarly attention. It meets patients' medical needs for specific knowledge and social needs for interpersonal relationships and empowers patients to search for and produce information on their own. As is known that these traditional online medical communities are usually in the form of a forum with predefined settings, which makes it difficult to spread and fit the new situation. To this end, some instant-messaging-software-based (WeChat, QQ) online medical communities are emerging recently in China. In particular, WeChat has become a channel for users to conduct information and social supporting interactions. WeChat group- based medical communities (WGMC) have not only the core sharing function of traditional forums but also a casual and natural communication style, which is more user-friendly. Specifically, in a WGMC, messages from any user can reach all members immediately through a "broadcasting" way. Such real- time reminders in WGMC yield a harmonious atmosphere of mutual help that can largely meet the needs of patients with major or chronic diseases. Patients not merely need specific medical information such as treatment plans or rehabilitation guidance, but also emotional and psychological comfort and encouragement to against their illness and increase their aspirations for a better life after recovery (Umefjord, Petersson et al. 2003). In short, WGMC is not only conducive to the integration of medical resources, but also makes the patient-doctor relationship and patient-patient relationship closer.

Nonetheless, in a WGMC, there are usually multiple evolving topics being discussed and multiple users sharing information simultaneously. On the one hand, concerning new users, it is difficult to integrate themselves into the community because of insufficient context information. On the other hand, regarding community managers or researchers, the massive but semi-structured data in the WGMC makes it impossible to manually label and efficiently utilize the data. Hence, it’s essential to recognize the latent dialogues behind the massive but semi-structured messages in WGMC.

There are extensive studies involving dialogue recognition covering various topics and methods (Li and Choi 2020, Liu, Sui et al. 2020). However, when concerning the dialogue recognition in WGMC, there are still some limitations. First, the existing literature is merely based on text features, ignoring the scene information (Xu and Ren 2020), thus failing to utilize the prior knowledge. Second, the existing complex deep learning models tend to converge too quickly in small samples of labeled data, yielding problems of overfitting or underfitting (Esteva, Robicquet et al. 2019). Third, due to the lack of sufficient textual information for domain adaptation, it is hard to get better recognition performance even if the text representation model or pre-trained model is trained by a large amount of text in the whole domain (Devlin, Chang et al. 2018).

To against this background, we focus on a typical WGMC, “Home of Love”(HOL), and propose a dialogue recognition method using prior knowledge of medical conversation scenes coupled with a small number of labeled data. Specifically, first, we propose a text classification model based on social support theory, which can complete the training with only a small number of manually labeled samples. This can address the dilemma that simple models are difficult to extract features while complex models tend to overfit. Second, based on the idea of full-word masking models in a Chinese context, we further propose two domain adaptation methods for pre-trained models. This can address the problem that pre-trained models trained with large-scale samples in the whole domain can’t perform well in specific tasks and verifies the feasibility of transfer learning. Third, we add more prior knowledge to the model through fusing text semantic information and dialogue scene information.

2. Related works on the online medical community

The online medical community improves the medical resource sharing capability, alleviates the scarcity and unequal distribution of high-quality medical resources to a certain extent (Lieberman, 2007). Also, it improves the overall level of national medical services and promotes the healthy and benign development of the whole medical industry. It’s verified that there is mutual social support among users in the online medical community, including three types of social support: information support, emotional support and accompany (Keating 2013). The online medical community can meet the multi-level needs of users. From the perspective of patients, it can not only provide sufficient medical information but also alleviate their negative emotions through keeping company with other patients. Information support among participants in the community can affect users' perceived empathy (Nambisan 2011).

From the perspective of doctors, existing studies have figured out why doctors participate in the online medical community and how they use it by using a questionnaire or obtaining more real feedback through interviews (Colineau and Paris, 2010). Some scholars also used econometric methods to explore the influencing factors of doctors' active social support in the online medical community, and verify the impact of different incentive methods on users' social support (Narayanan and Onn, 2016). From the perspective of platform, literature has studied the dynamic evolution of user relationship network and knowledge sharing network and knowledge interaction behavior in the online medical community (Ma and Agarwal, 2007). With the continuous improvement of in-depth learning, especially the rapid progress of natural language processing technology, many scholars can conduct more accurate research through the text content generated by users in the community (Vydiswaran, Zhai et al., 2011; Chen, 2012).

3. Model Construction

To address the problem of confusion of dialogue topic crossing WGMC, we propose a dialogue recognition model based on BERT, which integrates the semantic information and scene information of

each text. In specific, the model can be divided into three parts, as shown in Figure 1. In the first part, we use the Multi-BiLSTM model to conduct scene classification of text based on social support theory. In the second part, we extract semantic features of the two texts leveraging the BERT model. In the third part (colored in green), we concatenate the semantic information vector and the scene information vector of the dialogue and input it into a fully connected neural network to conduct the dialogue relationship prediction.

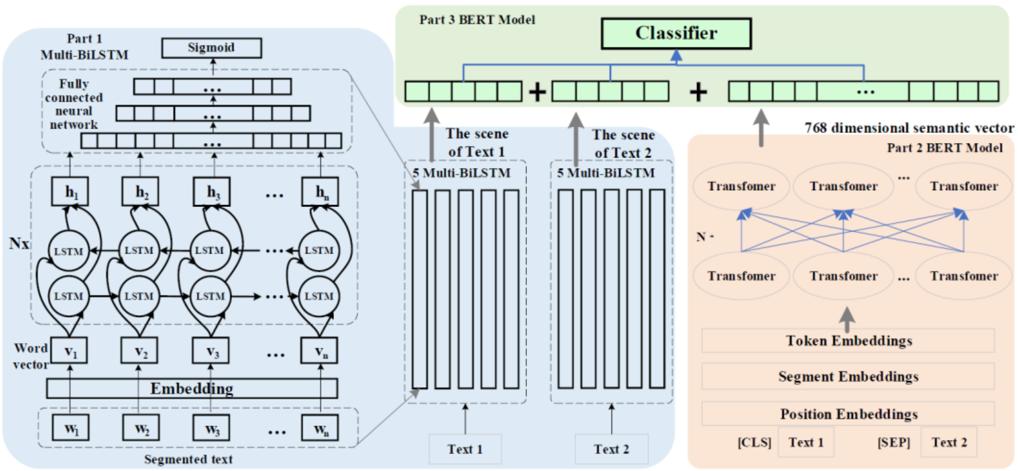


Figure 1: A dialogue recognition model based on BERT

3.1 Scene classification of text via Bi-LSTM

The social support theory posits that there are usually three main types of social support in online medical communities: information support, emotional support, and company (Lieberman, 2007). The information support includes consultation, communication of disease-related information. Emotional support refers to the emotional mutual help between patients or the mental support and encouragement of doctors to patients. Company denotes topics or activities that are unrelated to the disease but can activate the atmosphere. We subdivide information support and emotional support from the perspectives of provision and acquisition. Accordingly, the text information of participants is divided into five categories: information support providing(ISP), information support acquisition(ISA), emotional support providing(ESP), emotional support acquisition(ESA), and company(COM). Table 1 lists some typical instances of the five categories. We can see from table 1 that one text can be categorized into multiple scene categories. The text “怎么会这样呢？孩子几岁了，手术前怎么样？” not only expresses the demand for information acquisition but also expresses a concern for other users, thus can be categorized into both ISA and ESP. Hence, we adopt five binary classifiers for each text. For example, the above text can be expressed as: (0, 1, 1, 0, 0) according to five scene categories: (ISP, ISA, ESP, ESA, COM). Specifically, we propose a multi-layer Bi-LSTM classifier with the same structure but independent partial parameters to complete the task of scene classification of text.

Table 1: Instances of the five scene categories

Scene category	Instances
ISP	髓母多了一个脊髓放疗。 我家也是的，后来发现是脑积水，后来做的分流。
ISA	*怎么会这样呢？孩子几岁了，手术前怎么样？* 既然在这里做了，这时候就相信医生吧。
ESP	*怎么会这样呢？孩子几岁了，手术前怎么样？* 现在恢复的怎么样了？
ESA	我孩子现在活蹦乱跳的，祈祷手术出现奇迹，很害怕。 感谢。我快哭了。[捂脸][捂脸][捂脸]
COM	这个需要报名？还是按日期直接去就行。 才出院没几天很正常。实在疼得难受就吃点镇痛的。

For the binary classifier, to trade-off between a small number of labeled data, the complexity, stability, and generalization of the model, we propose a multi-layer Bi-LSTM classifier, as shown in the blue part of **Figure 1**. In specifically, the model is composed of three parts. The first part is the input and word embedding. The second part applies the multi-layer Bi-LSTM to extract text features. We adopt multi-layer Bi-LSTM for the trade-off between model depth and the small number of training samples. The third part is a fully connected neural network.

In the binary classifications, data sets will suffer a serious imbalance in the proportion of samples in the real scene. As such, for cross-entropy loss, the relatively high sample weight will be virtually enlarged, which will deviate from the optimization direction of the model and cannot be well fitted with the training data. Based on this, Lin et al. (2017) proposed the concept of Focal Loss based on the cross-entropy loss function, in which the weight of easy to classify samples is reduced by adding modulation coefficients, making the model focus more on samples that are difficult to classify, as shown in equation (1),

$$FL(p,y) = \begin{cases} -\alpha(1-p)^{\gamma}\, log(p), & if\;\; y=1 \\ -(1-\alpha)p^{\gamma}\, log(1-p) & if\;\; y=0 \end{cases} \tag{1}$$

We can adjust the parameter α according to the proportion of positive and negative samples, which plays the role of adjusting the parameter $Class\ Weight$ in the training of a deep learning model. γ refers to the modulation coefficient of Focal Loss. The larger γ, the lower the weight for the easily classified samples. When $\gamma = 0$, Focal Loss is the same as cross-entropy. Considering the imbalanced distribution of the five social support classifications, we use Focal Loss as the loss function, by setting different values of $\alpha$and γ to achieve the best classification training effect.

Finally, we apply five such multi-layer Bi-LSTM classifiers with the same structure but independent partial parameters to classify the text in terms of the five scenes.

3.2 BERT-based Dialogue Recognition Model

3.2.1 Domain adaptation of pre-training model

We can integrate all domain information to pre-train a model utilizing a huge heterogeneous corpus. However, when it comes to some specific domains or specific scenarios, there is still considerable room for improvement. To this end, we adopt the domain-adaptive pre-training method, by pre-training again the original pre-trained model, utilizing a more specific domain corpus. Accordingly, the corpus adopted in pre-training a BERT in Chinese comes mainly from Wikipedia in Chinese. However, it's not targeted at the text content of the online medical community.

To better perform dialogue recognition, we propose two similar pre-training models, according to the pre-training of BERT. And we adjust one model to adapt the text content of Chinese WGMC. For the first pre-training model, we follow the line of Whole Word Mask (WWM) to adjust the MLM (Cui, Che et al., 2019). Note that in Chinese texts, the word and words are yet to distinguish. Consequently, when training Chinese texts, MLM only masks the word granularity and ignores the existence of words in Chinese, resulting in semantic bias. To address this, WWM will first judge whether the current word belongs to certain Chinese words. If so, it will mask the related words according to semantic meanings, to realize masking words' granularity. However, the MLM alone can't learn all the features. Here, we set the masking proportion, wherein each word has a probability of 15% to be masked. In the masked words, each word will be masked by "MASK" with a probability of 80%, and be replaced with a probability of 15%, and remain unchanged with a probability of 5%. Table 2 shows an example of the masking language model.

The second pre-training task is the same adjacent sentence prediction task (NSP) as the BERT model. To obtain a better domain adaptation in the online medical community, we adopt the data of doctor-patient dialogue from an online diagnosis and treatment platform, "Haodaifu"[1], an online medical consultation platform that is similar to WGMC. The platform, "Haodaifu", has the same fragmented structure and colloquial characteristics as WGMC. In addition, doctor-patient Q&A data is well structured.

Table 2 An example of MLM

Example	Illustration
Original text	病人状态不好，头晕头痛不敢跑那么远。
Word-segmented text	病人 | 状态 | 不好 | ， | 头晕 | 头痛 | 不敢 | 跑 | 那么远 | 。
Word-level masking	病人状[MASK]不好，头晕头痛不敢[MASK]那么远。
Words-level masking	病人[MASK][MASK]不好，头晕头痛不敢[MASK]那么远。

By incorporating both the data of WeChat and "Haodaifu", we can perform the pre-training tasks of MLM and NSP, to adapt the all-domain-information-based model to the online medical community domain, to better capture the semantic relationships between dialogs.

[1] https://www.haodf.com/

3.2.2 BERT-based Dialogue Recognition Model

To address the problems of confusion of dialogue topic crossing in the online medical community, we propose a dialogue recognition model based on BERT, which integrates the text's semantic information and scene information. We extract semantic information of the two texts pair leveraging the BERT model, as shown in the orange part in **Figure 1**. And the scene information of each text is generated by the scene classification in section 3.1, as shown in the blue part in **Figure 1**. Furthermore, we concatenate the semantic information and scene information of two texts and input them into a fully connected neural network to obtain the final prediction of the dialogue relationship prediction, as shown in the green part in **Figure 1**. Specifically, the input can be divided into scene information and semantic information. For the scene information, we assign a 5-dimensional vector to each text's scene according to the social support theory. Regarding the training process, we concatenate the vectors of the two texts into a 10-dimensional vector, as the dialogue's scene information. For the text information, we first extract the semantic features using the pre-trained BERT model. It's noted that the text needs to be preprocessed according to the input rules. Specifically, the special symbol "[CLS]" is added at the beginning of the text as the mark of the beginning of the text, and the special character "[SEP]" is added between the two dialogue texts as the mark of the distinction between the two texts. After the preprocessing, the text also needs to be word embedded. **Figure 2** illustrates the BERT embedding.

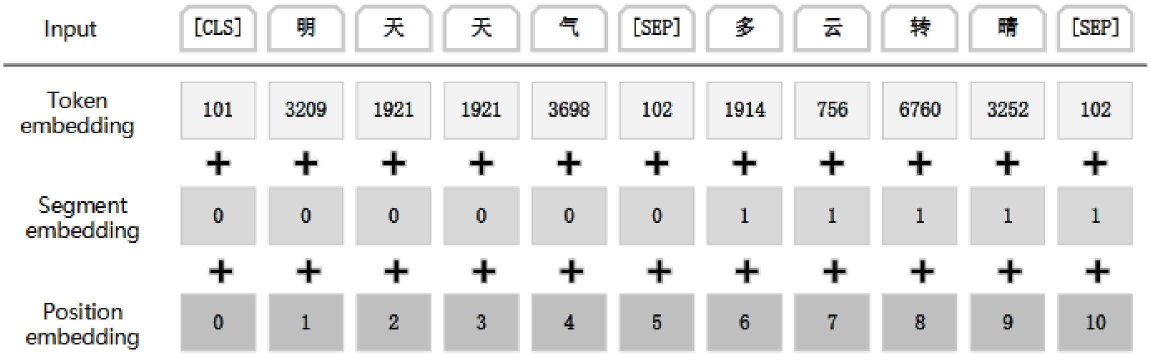


Figure 2: An illustration of BERT embedding

BERT embedding contains three parts: (1) token embedding, (2) segment embedding, and (3) position embedding. The token embedding refers to dividing words into limited common subword units (Wu, Schuster et al., 2016) through WordPiece's segmentation method, to control the size of the dictionary and effectively reduce the number of unregistered words. We encode each word with a unique ID, and each ID can be mapped to a 768-dimensional word vector. As for the position embedding, it adopts a position embedding method that is similar to the token embedding method. Specifically, the model will assign a unique id according to the word's position in the whole text. Furthermore, we can

concatenate the semantic features with the scene information of dialogues as the input of the classifier to obtain the final probability p_{ij} of whether the two texts can form a dialogue as follows in equation 2,

$$p_{ij} = Classifier\left(Concate(cls_{ij}, text_i, text_j)\right) \quad (2)$$

For WGMC, the intervals among dialogues are commonly short. Hence, here we don't consider matching all texts before the focal text. Instead, we set a number threshold, $Number$, and time threshold Gap to obtain the potential dialogue set of the focal text. In specific, when the distribution of the text before the focal text is sparse in terms of time, we adopt the time threshold Gap. In contrast, we adopt the number threshold, $Number$, if the distribution of the text is dense.

After obtaining the potential dialogue set of the focal text j, we can calculate the probability p_{ij} of text i and j to form a dialogue. We filter the probability that is higher than 0.5 as equation 3.

$$p_{ij} = \begin{cases} 0 & p_{ij} < 0.5 \\ p_{ij} & p_{ij} \geq 0.5 \end{cases} \tag{3}$$

Finally, we select the text j with the largest p_{ij} as the dialogue text of the focal text i, using the softmax function in equation 4.

$$s_{ij} = \frac{e^{p_{ij}}}{\sum_i^I e^{p_{ij}}} \tag{4}$$

The overall research line of this paper is shown in **Figure 3**

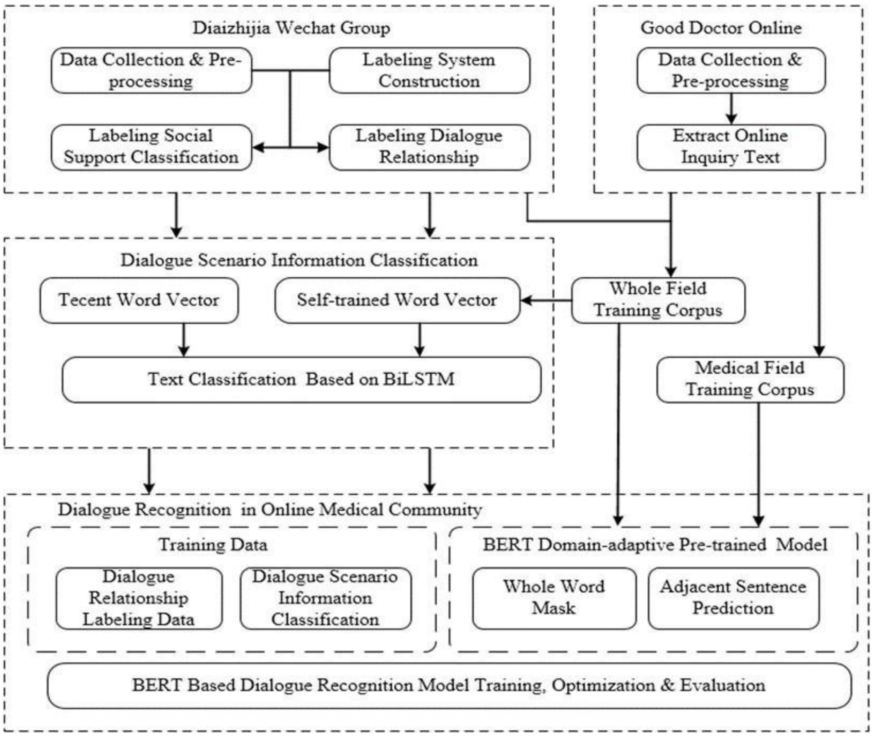


Figure 3: The whole structure of research

4. Experimental Results

4.1 Scene classification of text via Bi-LSTM

We collect the data from a WGMC, HOL. This community is mainly composed of the families of patients with glioma, medulloblastoma, germ cell tumor, and so on. HOL is responsible for the operational support of the whole community. We collected 23 groups before March 14, 2019, with a total of 510843 records.

In WGMC, a large number of text messages are short. Text messages within 10 words account for 39.13% and chat messages within 30 words account for 83.91%, which also proves that the communication of WeChat groups is mainly short texts. After preprocessing the raw data, this paper finally selects some dialogue data of two groups (code: 5007452012 and 5556417190) for the manual label.

In the labeling process, each data will be labeled three times, the results with the same results are selected, and a total of 13313 data are obtained. To avoid mutual influence, separate labeling is adopted. To better represent the information contained in the dialogue text and obtain a better classification effect, we adopt two ways to obtain the word: Tencent word vector (Song, Shi et al., 2018) and self-training word vector of user-defined text. Tencent word vector is a large-scale and high-quality Chinese word vector opened by Tencent AI Lab. However, for specific text processing tasks, when there is sufficient corpus, it is also a common way to train self-training word vectors for specific fields. This paper will use all dialogue text information to train word vectors suitable for WGMC with CBOW model.

After acquisition the text representation model, to compare and verify the classification efficiency of the BiLSTM model, this paper selects Logistic Regression (LR), XGBoost, TextCNN, single-layer LSTM, and BERT model for comparative experiments. The specific parameters of each model are shown in **Table 3**.

Table 3: The parameters of the training model

Models	TextCNN	Base BERT	Multi BiLSTM_t	Multi BiLSTM_c
structure	Embedding+CNN	6*Transformer	Embedding+4*BiLSTM	Embedding+4*BiLSTM
vectors	TextCNN	-	Tecent word vector	Self-trained word vector
loss function	Focal Loss	Focal Loss	Focal Loss	Focal Loss
max length	256	128	256	256
iterations	100	30(5)	100	100
learning rate	1.00E-04	1.00E-05	1.00E-03	1.00E-03
optimizer	Adam	Adam	Adam	Adam
patch number	128	32	128	128

The above parameters are the optimal parameter configuration obtained after multiple tests of each model. After fixing the parameters, we use the same labeled corpus for training. In the training process, k-fold verification is used, in which K is set to 5.

Because there is a serious imbalance between positive and negative samples in ESA in the classification task, the probability threshold for completing the classification is changed according to the model, so the $Accuracy$ can not be completely used as the evaluation standard of the model. It is necessary to add an auxiliary evaluation index ROC curve as the evaluation standard of the model training effect.

After 50% discount verification, the effect of the classification model is shown in **Table 4**. The values in the table are 50% discount verification training and corresponding standard error. From five different classification tasks, we can see that the model with balanced classification training has a higher fitting degree and relatively good stability to the data. On the contrary, the fitting effect of the model is poor, and the deviation of the results obtained from many experiments is large.

Table 4: Model comparison

Models	index	ISP	ISA	ESP	ESA	COM
LR	Accuracy	0.6604	0.6887	0.7385	0.8621	0.6721
		-0.0106	-0.0064	-0.0126	-0.0085	-0.0046
	AUC	0.6483	0.6169	0.6355	0.5532	0.5267
		-0.0092	-0.0103	-0.0123	-0.0253	-0.0062
XGBoost	Accuracy	0.7231	0.8048	0.8529	0.9726	0.7832
		-0.0088	-0.0083	-0.0072	-0.0024	-0.0092
	AUC	0.6928	0.5875	0.612	0.6091	0.5915
		-0.0066	-0.0085	-0.0109	-0.0052	-0.0061
TextCNN	Accuracy	0.7862	0.8476	0.7349	0.9529	0.7951
		-0.0997	-0.0406	-0.0261	-0.0421	-0.0107
	AUC	0.8153	0.7353	0.7319	0.6491	0.5506
		-0.0579	-0.0907	-0.0294	-0.0373	-0.0302
Base Bert	Accuracy	0.7396	0.7963	0.8468	0.9465	0.7926
		-0.0079	-0.012	-0.0096	-0.0053	-0.0048
	AUC	0.7816	0.772	0.758	0.6972	0.5904
		-0.0044	-0.0159	-0.0076	-0.0109	-0.0144
Multi BiLSTM_t	Accuracy	0.8172	0.82	0.8318	0.8975	0.7463
		-0.0186	-0.0224	-0.077	-0.0975	-0.0747
	F1	0.8557	0.6083	0.5952	0.3198	0.4883
		-0.0119	-0.0201	-0.0779	-0.1289	-0.1042
	AUC	0.8738	0.8557	0.85	0.8378	0.7851
		-0.0079	-0.0077	-0.0126	-0.0121	-0.0224
Multi BiLSTM_c	Accuracy	0.8354	0.869	0.9034	0.9565	0.8431
		-0.0044	-0.0156	-0.0057	-0.0094	-0.0165
	F1	0.8637	0.6922	0.7064	0.4161	0.6216
		-0.006	-0.0259	-0.0149	-0.0387	-0.025
	AUC	0.879	0.8723	0.871	0.8043	0.8152
		-0.0069	-0.0111	-0.0116	-0.0312	-0.0096

The comprehensive comparison shows that the multi-layer BiLSTM model has achieved the best accuracy and the highest AUC value in the five classification tasks. Compared with the Tencent word vector with richer content, the self-trained word vector has achieved better results in four models such as ISP, ISA, ESP, and COM. In ESA model, the AUC is slightly lower than the Tencent word vector, but the accuracy and F1 are higher.

Using multi BiLSTM_C predicts all 470 thousand WeChat group chat texts and gets the proportion of various social support types, as shown in **Figure 4**.

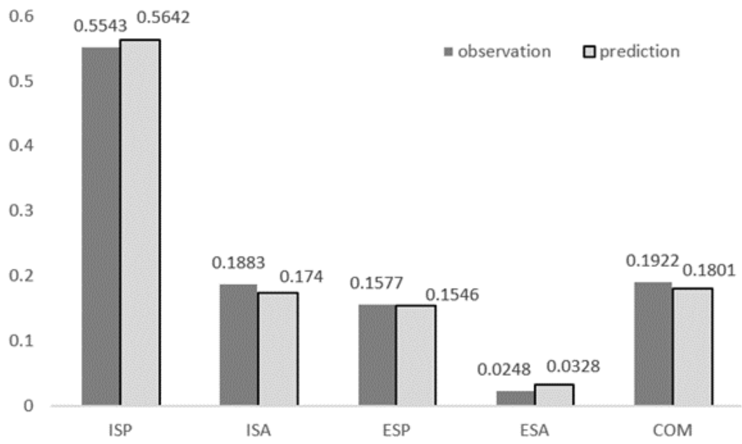


Figure 4: text classification

Except for ESA category, the distribution differs from the actual label is approximately within 5%. However in ESA, due to the serious problem of sample imbalance, the two parameters of *Focal Loss* are increased to correct the deviation of the model during training, which also leads to the problem of too many parameters in the real prediction of the model. The multi-layer BiLSTM model proposed in this paper can maintain a good classification effect and generalization ability even in the case of a small number of labels.

4.2 Results of domain adaption of the pre-training model

To improve the capability of the pretraining model to understand the semantic features in the WGMC, we propose two pre-training tasks before the dialogue recognition. (1) whole word masking (WWM) language model pre-training task, and (2) adjacent sentence prediction task (NSP).

For the whole word masking language model pre-training, we also use the "Haodaifu" doctor-patient Q&A information to supplement the current 370 thousand text data from the WGMC. The “Haodaifu” data was collected in 2018 and has been preprocessed such as removing sensitive information, gift-giving information, voice information, and other non-text information, yielding 11 billion online interrogation records as pre-training corpus. We pre-trained the model for three epochs to obtain a stable loss as shown in **Figure 5**.

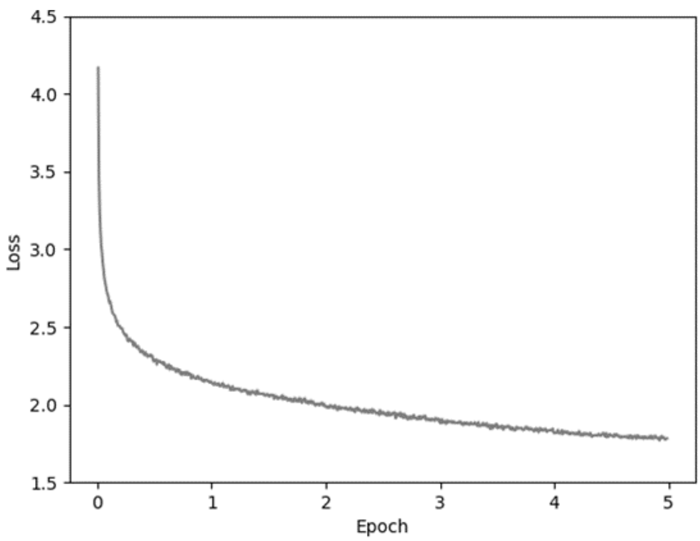


Figure 5: Whole word masking language model pre-training process

In the adjacent sentence prediction task, to avoid overlapping with the subsequent dialogue recognition training, we only trained the "Haodaifu" doctor-patient Q&A record. After data cleaning and short text screening, 5.5 billion doctor-patient Q&A dialogue text records are sorted out as the positive sample of NSP task. We further use all 11 billion original corpora to randomly generate the same amount of negative samples to form a training corpus with balanced positive and negative samples.

The training environment of NSP task is the same as that of the domain adaptation task of the whole word masking language model. The model training process is shown in **Figure 6**. After completing the first generation training, the basic accuracy rate of the model is stable at 93.5% and the loss value is stable at 0.163. Therefore, this paper finally selects the model after training a whole generation.

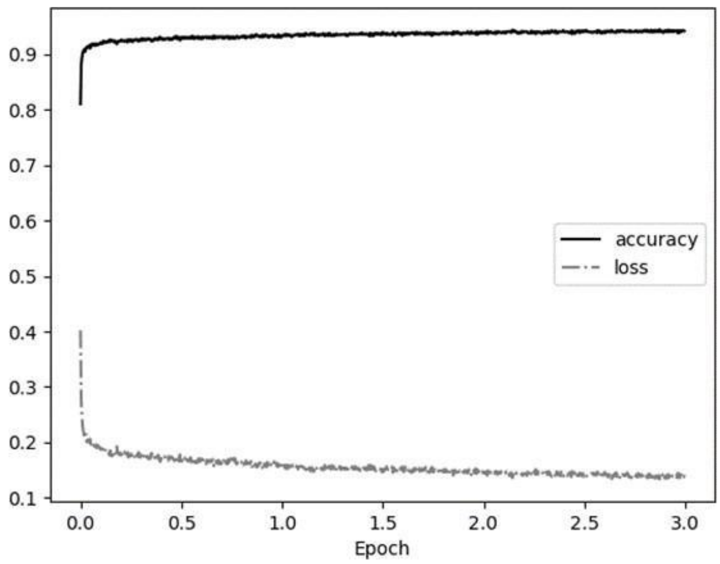


Figure 6: Whole word masking language model pre-training process

In addition, we randomly select 1000 dialogue texts from WeChat group chat texts as the test corpus and performs two operations on each text: single word masking and whole word masking. By comparing the accuracy and loss of multiple models for single word masking language model and whole word masking language model and evaluating the semantic understanding ability of the pre-training model for the online medical community, the results are shown in **Figure 7** below.

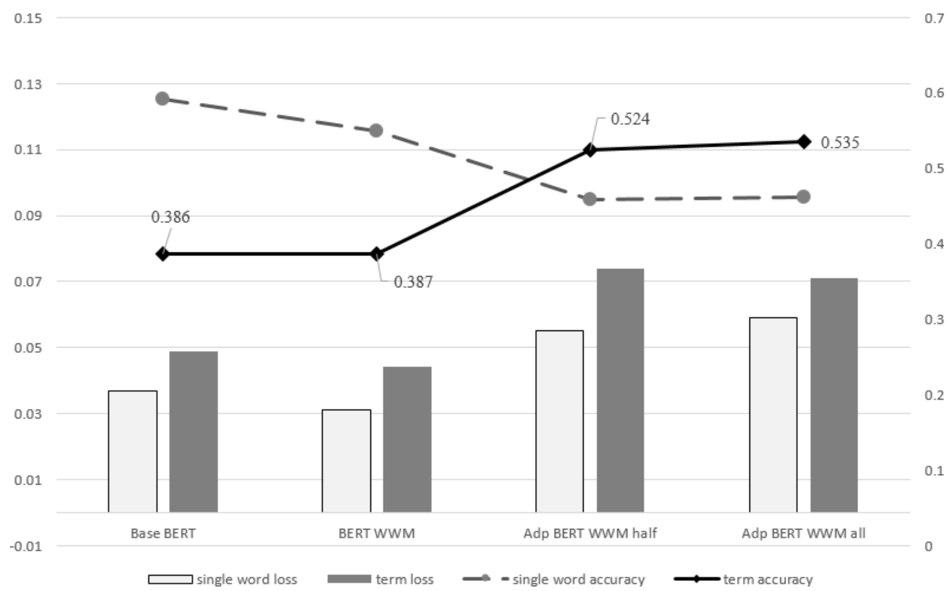


Figure 7 : Comparison of the semantic understanding ability of pre-trained models in the online medical field

In **Figure 7**, the base BERT model represents the BERT pre-training model published by Google. without any processing, the accuracy of single word prediction can reach 59.2% and the accuracy of whole word prediction can reach 38.6%. The BERT WWM model is a BERT pre-training model released by Harbin Institute of technology. The whole word masking language model pre-training is added based on the original BERT, which does not improve the semantic understanding ability of the Chinese online medical community. The Adp BERT WWM half represents the pre-training model obtained when the whole word masking language model is used to complete 1.3 iterations. The output of the model will be marked with "##" for the first word of the phrase and all subsequent words. After five generations of domain adaptation model training, the accuracy is further improved to 53.5%, and the loss value is further reduced.

After the model pre-training of the two domain adaptation tasks, the BERT model has a stronger semantic understanding ability for the WGMC than the original base BERT.

4.3 Results of the dialogue recognition

In this section, we perform dialogue recognition in WGMC. We label dialogue relationships on the dataset from two WeChat groups. After cross-validation, a total of 9196 dialogue relationships were obtained as positive samples. In the meantime, we also randomly select the same number of texts with lengths larger than 3 as negative samples. It's noteworthy that during the process of labeling, we set the number threshold to 30 and the time threshold to 24h. In other words, the potential dialogue set of the focal text is the texts that are posted less than 24 hours before or later than the focal text or are posted less than the 30 texts earlier or later than the focal text.

After constructing the dataset, we select various models as baselines, including Multi-layer BiLSTM, the Base-BERT(Chinese) of Google (Devlin, Chang et al., 2018), the BERT-WWM(Chinese) (Cui, Che et al., 2019). The implementation details are shown in Table 5.

Table 5: Model parameters

Models	Multi-layer BiLSTM	Base BERT	BERT WWM	Adaptive WWM	Adaptive WWM+NSP
Abbreviation	LSTM	base	wwm	adp wwm	adp wwm_nsp
Training corpus	Wechat, Haodaifu	Wikipedia	Wikipedia , ext	Wikipedia, Wechat, Haodaifu	Wikipedia, Wechat, Haodaifu
Loss function	Cross entropy	Cross entropy	Cross entropy	Cross entropy	Cross entropy
Maximum length	256	128	128	128	128
Epoches	200	early-stop	early-stop	early-stop	early-stop
Learning Rate	0.001	0.00001	0.00001	0.00001	0.00001
Optimizer	Adam	Adam	Adam	Adam	Adam
Batchs	128	32	32	32	32

We split all the sample data with 80% as the training set and 20% as the test set. **Table 6** provides the results.

Table 6: Model comparisons

Features	Indexes	LSTM	base	wwm	adp wwm	adp wwm_nsp
Text features	Precision	0.6801	0.8308	0.8404	0.859	0.8603
	Recall	0.7022	0.7546	0.8281	0.8543	0.8478
	F1	0.6775	0.8113	0.8401	0.8537	0.8525
Text features+ Scene features	Precision	0.7235	0.8562	0.8546	0.8625	0.8715
	Recall	0.7329	0.8314	0.8123	0.8727	0.8759
	F1	0.7121	0.8478	0.8433	0.8595	0.8732

Experiments show that the BERT model based on attention mechanism is much better than other models in dialogue recognition, and the recognition accuracy and F1 value can reach more than 80%. However, the multi-layer Bi-LSTM model is not in good performance in dialogue recognition tasks.

The horizontal comparison is shown in **Figure 8**. The two models adapted in the field have better performance, and the accuracy, recall, and stability of the F1 value are significantly higher than those of other models. Through the domain adaptation of the whole word masking language model, the accuracy of using only text features is 85.90%. After integrating text scene features, the accuracy is slightly improved to 86.25%, and the F1 value is stable at 0.8595. Because the domain adaptation of sentence critical prediction task is similar to dialogue recognition task, the improvement effect is more obvious, and the best effect is achieved. The accuracy rate is 87.15%, and the F1 value can also reach 0.8732.

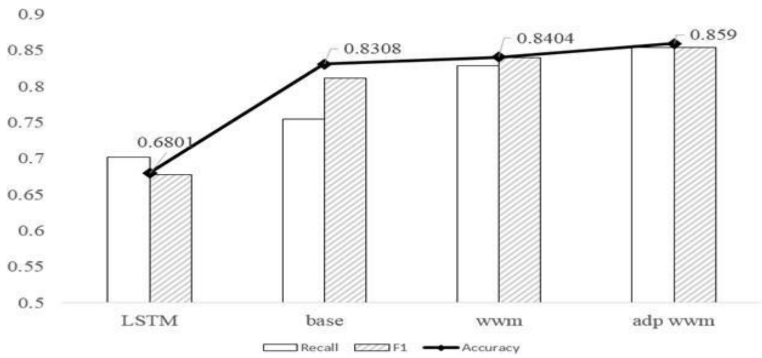


Figure 8: Whole word masking language model pre-training process3

In terms of vertical comparison, for the pre-training model after domain adaptation, the effect of classification and recognition has not been significantly improved, but in terms of stability, recall rate, accuracy, and F1 value have been significantly improved.

5. Conclusions and future works.

WeChat group, an important online medical community in China, has become a medium for many patients and even their families to obtain disease information and exchange ideas. However, the cross topics and complex conversation threads make the managers and users struggle to acquire useful information.

We focus on the Wechat-group HOL as our dataset, and collected communication text data from multiple Wechat groups. First, we classify the dialogue texts into five scenes according to social support theory. Meanwhile, we compare the performances based on the whole field training corpus and the medical field training corpus. In specific, during the process of semantic features extraction, we adopt the domain adaptation approach in transfer learning to implement the pre-training of BERT in learning the semantic feature of the online medical community, to enhance the semantic understanding of the pre- trained BERT model. Therefore, we can further construct the dialogue recognition model to incorporate the text semantic features coupled with scene information.

The experimental results reveal that by properly adjusting the model structure, we can still obtain good fitting results even if given a small labeled sample set. And the model trained can outperform many complex pre-trained models. In addition, the model pre-trained with whole field training corpus also yields competitive performance, manifesting the excellent generalization ability of the large word vectors. In the dialogue recognition model, we pre-trained BERT model, coupled with proper adaption approaches, to advance its' semantic understanding ability in special domains. Furthermore, the scene information based on social support theory can enhance the extraction of semantic features for the pre-training model, and extensively improve the dialogue recognition performance.

References

Chen, A. T. (2012). "Exploring online support spaces: using cluster analysis to examine breast cancer, diabetes and fibromyalgia support groups." *Patient education and counseling* **87**(2): 250-257.

Colineau, N. and C. Paris (2010). "Talking about your health to strangers: understanding the use of online social networks by patients." *New review of hypermedia and multimedia* **16**(1-2): 141-160.

Cui, Y., W. Che, T. Liu, B. Qin, Z. Yang, S. Wang and G. Hu (2019). "Pre-training with whole word masking for chinese bert." *arXiv preprint arXiv:1906.08101*.

Devlin, J., M.-W. Chang, K. Lee and K. Toutanova (2018). "Bert: Pre-training of deep bidirectional transformers for language understanding." *arXiv preprint arXiv:1810.04805*.

Esteva, A., A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, C. Cui, G. Corrado, S. Thrun and J. Dean (2019). "A guide to deep learning in healthcare." *Nature medicine* **25**(1): 24-29.

Keating, D. M. (2013). "Spirituality and support: A descriptive analysis of online social support for depression." *Journal of religion and health* **52**(3): 1014-1028.

Li, C. and J. D. Choi (2020). "Transformers to learn hierarchical contexts in multiparty dialogue for span-based question answering." *arXiv preprint arXiv:2004.03561*.

Lieberman, M. (2007). "The role of insightful disclosure in outcomes for women in peer-directed breast cancer groups: a replication study." *Psycho-Oncology: Journal of the Psychological, Social and Behavioral Dimensions of Cancer* **16**(10): 961-964.

Lin, T.-Y., P. Goyal, R. Girshick, K. He and P. Dollár (2017). Focal loss for dense object detection. *Proceedings of the IEEE international conference on computer vision*.

Liu, J., D. Sui, K. Liu and J. Zhao (2020). Graph-based knowledge integration for question answering over dialogue. *Proceedings of the 28th International Conference on Computational Linguistics*.

Ma, M. and R. Agarwal (2007). "Through a glass darkly: Information technology design, identity verification, and knowledge contribution in online communities." *Information systems research* **18**(1): 42-67.

Nambisan, P. (2011). "Information seeking and social support in online health communities: impact on patients' perceived empathy." *Journal of the American Medical Informatics Association* **18**(3): 298-304.

Narayanan, S. and A. Onn (2016). The influence of perceived social support and self efficacy on resilience among first year Malaysian student. Kajian Malaysia, 34 (2), 1–23.

Song, Y., S. Shi, J. Li and H. Zhang (2018). Directional skip-gram: Explicitly distinguishing left and right context for word embeddings. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*.

Umefjord, G., G. Petersson and K. Hamberg (2003). "Reasons for consulting a doctor on the Internet: Web survey of users of an Ask the Doctor service." *Journal of medical Internet research* **5**(4): e26.

Vydiswaran, V. V., C. Zhai and D. Roth (2011). Gauging the internet doctor: ranking medical claims based on community knowledge. *Proceedings of the 2011 workshop on Data mining for medicine and healthcare*.

Wu, Y., M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao and K. Macherey (2016). "Google's neural machine translation system: Bridging the gap between human and machine translation." *arXiv preprint arXiv:1609.08144*.

Xu, B. and W. Ren (2020). Understand Online Medical Consultation Based on Dialogue State Tracking. *2020 4th International Conference on Computer Science and Artificial Intelligence*.

An interactive web-based dashboard to examine trending topics: application to financial journals

NGOC PHAN[a], NAYANA PAMPAPURA MADALI[b], SAHAR BEHPOUR[b] and TING XIAO[a,b]

[a]Department of Computer Science and Engineering,
University of North Texas, US

[b]Department of Information Science,
University of North Texas, US

ngocphan@my.unt.edu
nayanapampapuramadali@my.unt.edu
Sahar.behpour@unt.edu
ting.xiao@unt.edu

ABSTRACT

Understanding trends is helpful to identify future behaviors in the field, and the roles of people, places, and institutions in setting those trends. Although traditional clustering strategies can group articles into topics, these techniques do not focus on topics over limited timescales; additionally, even when articles are grouped, the generated results are extensive and difficult to navigate. To address these concerns, we create an interactive dashboard that helps an expert in the field to better understand and quantify trends in their area of research. Trend detection is performed using the time-biased document clustering introduced in Behpour et al. (2021) study. The developed and freely available web application enables users to detect well-defined trending topics in financial journals by experimenting with various levels of temporal bias - from detecting short-timescale trends to allowing those trends to spread over longer times. Experts can readily drill down into the identified topics to understand their meaning through keywords, example articles, and time range. Overall, the interactive dashboard will allow experts in the field to sift through the vast literature to identify the concepts, people, places, and institutions most critical to the field.

AUTHOR KEYWORDS

Interactive dashboard; journal trends; time biased clustering; data visualization; trend analysis.

1. INTRODUCTION

One of the most effective methods for visualizing data is creating interactive dashboards. Interactive dashboards are useful since they enable users to edit data across multiple dimensions and choose the analytics focus (Wu et al., 2019). In recent years, many businesses and organizations have turned to data visualization and interactive dashboards to analyze data and spot trends or patterns so that they can better serve their customers.

Trend analysis is a method that examines data over time to spot patterns or trends. It has several benefits, such as the ability to compare, forecast, and identify research gaps. Research related to trend analysis is carried out in various areas. For example, Twells et al. (2014) used historical data on the prevalence of adult obesity in Canada from 1985 to 2011 to detect the prevalence pattern and make future projections. Yu et al. (1993) examined surface water quality data to determine if any patterns of the water composition emerged. Velvizhi et al. (2021) studied trends in poly waste management in India by collecting and analyzing research papers and studies.

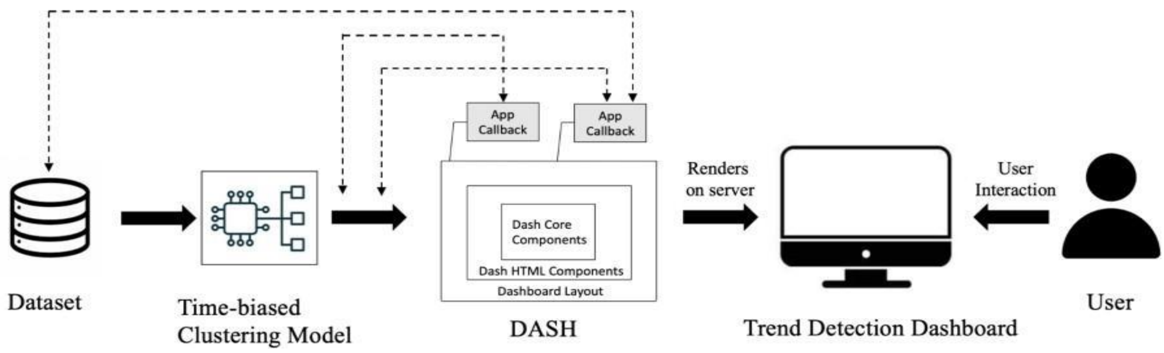
In 2019, the National Institute of Health had a 20% research grant acceptance rate while the Fulbright U.S. Student Program had a 18% research grant acceptance rate (Aswell, 2020). Most research proposals were rejected for funding because the research topics seemed inappropriate for the funding agency (“Getting Grants - Common Reasons Proposals Are Rejected”, 2021). The two potential reasons that contribute to the low research grant acceptance rate are: 1) research topics do not align with the current trends, and 2) the trend for the research topic may become obsolete within a few years. Therefore, it is critical to examine the research topics to comprehend their trends. This will aid in determining how research topics have changed over time and how they may change in the future.

Extensive research related to trend analysis, journal trend analysis, and interactive dashboards have all been carried out, however there has not been much work on combining all of these techniques into one. This study is the outcome of a collaborative effort to bring all of these approaches together and develop an interactive dashboard that automatically detects journal trends. The study extends Behpour et al. (2021) research on automatic trend detection by utilizing their time-biased document clustering model to detect trends in financial journals and offer an interactive web-based dashboard that provides thorough insights into raw data and the detected journal trends. Our study extracts additional features from the raw data and the results provided by Behpour et al. (2021) research for conducting a comprehensive temporal analysis of the raw data and the detected journal trends. The features are derived from the existing features of the raw data and temporal clustering model's output results. These features include abstract length, keyword frequency, keyword timeline, topic timeline, and topic unique terms.

The key contribution of this research paper is the interactive dashboard which offers a variety of information about journal trends. The trend detection application, which was built using Python, Plotly and Dash, provides a lightweight application that can be easily deployed to a Flask-based server. The interactive dashboard has three tabs: the journal trends tab, the dataset tab, and the vocabulary tab. While the journal trends tab provides information about the topics and their trends, the dataset tab provides a quick overview of the data. The vocabulary tab is included to enhance the user's understanding of terms used in the application. There are three sub-tabs under the journal trend tab: simple UI, advanced UI, and diagnostics. For the scope of our study, we report the findings for trending topics in financial journals.

Figure 1 provides a system diagram for our trend detection application. Raw data would be preprocessed and available to be used by the time-biased clustering model and the application. Every time the user initiates an action on the dashboard, the application would retrieve the preprocessed data, run the time-biased clustering algorithm to detect the trending topics, render relevant data and model's results on the server, and finally update the dashboard.

Figure 1
System diagram showing different processes and workflow of the interactive journal trend dashboard



2. LITERATURE REVIEW

2.1. Trend Analysis

Trend analysis is a statistical method typically used for interpreting data. It aids in the comprehension of historical trends and the forecast of future trends based on prior patterns. Trend analysis has been researched for a long time, and it is used in various fields such as medicine, environmental science, biology and so on (Abdolalizadeh et al., 2021; Ceribasi & Ceyhunlu, 2021; Wang et al., 2021). Asfaw et al. (2018) broke down the meteorological factors concerning environmental change using trend analysis. Their Mann-Kendall pattern investigation test result revealed a growing pattern for mean and least normal temperatures over time, while the highest temperature pattern showed a non-critical growing pattern. Pourghasemi et al. (2020) employed "spatial modeling, risk mapping, change detection, and outbreak trend analysis" on COVID-19 pandemic data in Iran to determine the trends that might help with the administration and control of the COVID-19 emergency. Their findings demonstrate that Iran's death trends are similar to those of the rest of the world, but that the world's death trends are more volatile. Noghabaei et al. (2020) performed a trend analysis for the reception of virtual and augmented reality in the architecture, engineering, and construction industry. Their findings show that older generations defined as age 35 and older are essentially more certain about the future of virtual and augmented reality innovations and see more advantages in virtual and augmented reality use. Romano et al. (2021) used time-trend analysis to examine the impact of the COVID-19 outbreak on outpatient drug sales and shortages. Overall, their study showed that COVID-19 led to a surge in drug demand and reported a deficiency in drug availability in the early stages of outbreak (Romano et al., 2021).

2.2. Journal Trends

Much research is carried out to understand journal trends and analyze journal topics. Journal trends related to various fields are studied to identify the trending research topics, authorship patterns, author productivity, patterns of co-authorship and so on (Sadik-Batcha & Ahmad, 2021). Jang et al. (2017) analyzed research publications from the Journal of Korean Society Dental Hygiene (JKSDH) and found trends and directions in dental hygiene research for the variables such as research topics, type of research and so on. Similarly Karnak et al. (2020) investigated Journal of Applied Behavior Analysis (JABA) articles and identified research topics and other publication trends for various characteristics such as gender of the author, country of origin and so on. Sadik-Batcha & Ahmad (2021) analyzed the publication patterns in two journals: the Indian Journal of Information Sources and Services (IJISS) and the Pakistan Journal of Library and Information Science (PJLIS). According to their findings, many of the articles in both journals were multi-authored. They also discovered that IJISS had an author productivity of 0.53 and PJLIS had an author productivity of 0.50. Many techniques and tools such as bibliometric analysis, topic modeling, google trends, neckinger 4 (Ali et al., 2021; Behpour et al., 2021; Russo et al., 2020; Lee & Kim, 2019) have been used to perform trend detection in journals. Despite the fact that extensive research is done to discover journal trends, they are ineffective due to the lack of a user interface.

2.3. Interactive Dashboard using Python, Plotly and Dash

Interactive dashboards have various benefits, such as offering a deeper understanding of raw data, making data processing easier, and assisting in improved learning and better data-driven decisions. Toasa et al. (2018) conducted a literature review to gain a better understanding of data visualization, its approaches, and existing dashboard platforms. They underlined the necessity of effective dashboard design such as customizing existing platforms and building customized boards, in their article because a dashboard is a powerful tool for presenting data. Brath and Peters (2004) address the relevance of dashboard design by claiming that a well-designed dashboard aids in efficiently communicating essential information to the users. In their study of COVID-19 pandemic, Clement et al. (2020) used Python modules, Dash and Plotly, to build an interactive dashboard that analyzed COVID-19 trends. The dashboard displayed the current trends related to the COVID-19 pandemic globally as well as in Canada.
Smieszek et al. (2021) created an interactive dashboard to assist clinicians in analyzing and interpreting health data acquired through Apple Watches. The authors used Plotly to depict the user's health data, including heart rate, exercise, and other activities. Their interface made it easier to spot data trends and see how they influenced the user's health. Enyew et al. (2020) examined the impact of Hurricane Florence on local food bank operations and created an interactive dashboard using Dash and Plotly. The dashboard was intended to enable foodbank managers to interactively visualize and understand food donation and distribution statistics, allowing them to better prepare, respond, and recover from disasters (Enyew et al., 2020).

From the above-mentioned literature, it is clear that trend analysis, journal trend detection, and interactive dashboards techniques are studied extensively. But there hasn't been much work on merging all these methods into one. This research is the result of a collective effort to combine all these methodologies and build an interactive dashboard that automatically detects the journal trends in financial journals.

3. METHODOLOGY

The research team obtained the raw data and temporal clustering algorithm from the Behpour et al. (2021) study to build a web-based interactive dashboard for detecting journal trends. The implementation model proposed by Behpour et al. (2021) for developing a temporal clustering was inefficient as that study incorporated both the data preparation process and model development process into one extensive process. This would increase the application's run time because we would need to prepare the data for training a time-biased clustering model every time we used the algorithm to generate trending topics. Since we only needed to run a data preparation process once whenever new data is collected, we split Behpour et al. (2021) implementation into two sub-processes including data preparation and model development to reduce the application's run time. As a result, the proposed application had four main processes: 1) data preparation, 2) temporal clustering model preparation, 3) model's evaluation results generation, and 4) dashboard construction.

3.1. Data Preparation

The first step in data preparation was data collection. Since the data provided by the Behpour et al. (2021) study only included a collection of financial economics articles from January 1974 to March 2020, we collected additional financial economics articles published between April 2020 and December 2020 from ScienceDirect for a complete analysis of 2020 data. The resulting dataset had an additional 115 articles, which resulted in a total of 2,973 articles.

The second step of data preparation was data preprocessing, which included the following operations: feature extraction, data cleaning and data normalization. For feature extraction, the following derived features have been identified based on the content of existing features in the dataset:

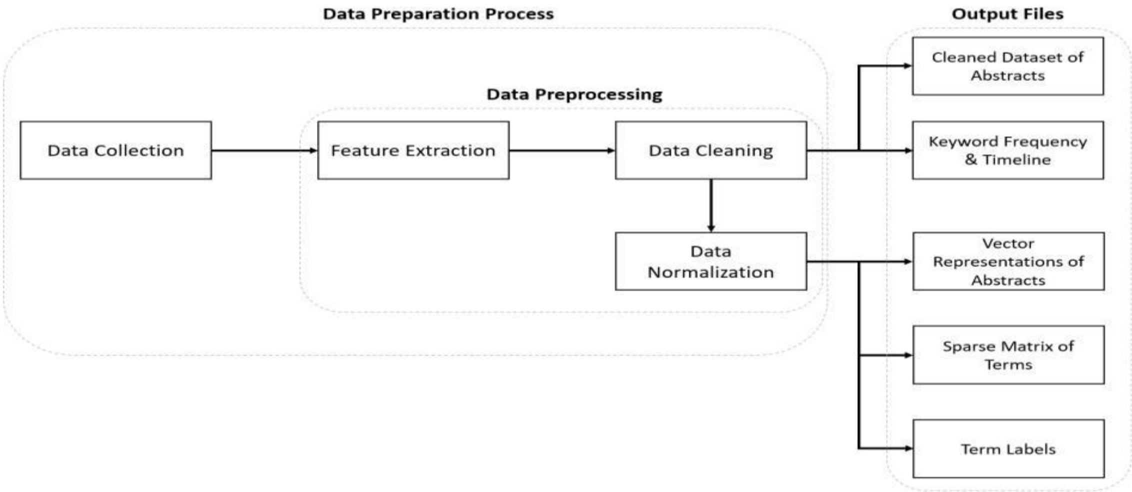
- Abstract length includes the number of words in the abstract.
- Number of keywords includes a count of the keywords for each article.
- Keyword frequency includes the number of articles that contain a given keyword.
- Keyword timeline includes the time period of the articles that contain the keyword.
- Normalized year of publication is computed by applying standard scaling on the year of publication.

For data cleaning, we first converted the month of publication from month name to a corresponding numeric value and assigned a value of -1 for any invalid months. Then we applied the following text processing strategies such as lowercasing, stemming, stop words and non-words removal to both article abstracts and keywords. For data normalization, we performed the following steps:

1. Perform hyperparameter tuning to select the best number of components for Latent Dirichlet Allocation (LDA) and Singular Value Decomposition (SVD). This is done by fitting several KMeans models on a matrix representation of abstracts created using different combinations of the number of components for LDA and SVD. For a clustering run that has a highest Silhouette score, the recognized best number of components for LDA and SVD are 50 and 10, respectively.
2. Create a term frequency vector of the abstracts using Python's sklearn module, CountVectorizer.
3. Extract 50 features from the term frequency vector using LDA. This produces a feature matrix of abstracts.
4. Transform the feature matrix to a normalized Term Frequency-Inverse Document Frequency (TF-IDF) representation.
5. Reduce the number of features for each observation to 10 components using SVD.
6. Convert a sparse matrix representation of abstracts produced from the preceding operation to a dense Numpy array. This produces a Numpy array with a size of 2,973 x 10.
7. Normalize a Numpy array of abstracts using Python's sklearn module, Normalizer. Then divide each feature by the standard deviation of its own column.
8. Build a sparse matrix of terms for the entire corpus using Python's sklearn module, TfidfVectorizer.
9. Create a list that includes a term label for each item in the term matrix.

Figure 2 provides a work process flow diagram for the data preparation process discussed above. The process starts with dataset collection followed by data preprocessing operations. At the end of the data preparation process, five output files are produced for model building and dashboard construction.

Figure 2
Work process flow diagram for data preparation process



3.2. Temporal Clustering Model Preparation

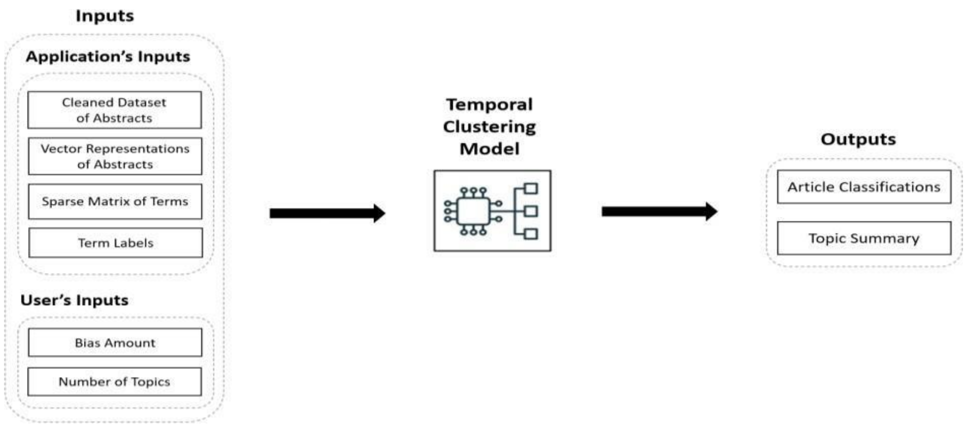
After preprocessing the dataset for model training, we prepared a temporal clustering algorithm that performed the following:

1. Multiply normalized year of publication by a bias amount provided as user's input.
2. Append the computational result from the preceding step to the vector representation of abstracts as a feature. The resulting Numpy array has a size of 2,973 x 11.
3. Fit a Spectral Clustering model on the Numpy array created from the previous step.
4. Convert a sparse matrix of terms to a dense Numpy array and get the labels of top 20 terms for each topic.
5. Compute the following metric scores for each topic:
   (a) Silhouette score is computed by taking the average of the Silhouette coefficients returned from sklearn's function, metrics.silhouette_samples.
   (b) Average standard deviation of years is computed by taking the standard deviation of the normalized year of publication.
   (c) Trend score is computed following two steps: 1) divide Silhouette score by average standard deviation of years, and 2) take the average of the computational result in step 1.

As a result, the temporal clustering algorithm produced an article classification file and a topic summary file. The article classification file included all fields in the raw dataset and a "Topic Id" field that specified the trending topic's id for a given document. The topic summary file includes the following fields: topic id, topic terms, topic timelines, trend score, silhouette score, average standard deviation of years, number of articles and percentage of articles.

Figure 3 displays the inputs required for building the time-biased clustering model along with the outputs returned from the model. According to Figure 3, there are two types of input: application's inputs and user's inputs. The application inputs include the four files prepared from the data preparation process. The user's inputs include bias amount and number of topics. After receiving the inputs, the temporal clustering model generates the trending topics and returns two Pandas data frames: article classifications and topic summary.

Figure 3
Required inputs for temporal clustering model along with outputs returned from the model



3.3. Model's Evaluation Results Generation

As illustrated in Figure 3, the temporal clustering model takes two user's inputs: bias amount and number of topics, denoted as k value. Without knowledge of the model's performance on various bias amounts and k values, it is difficult to specify a bias amount and a k value that would enable the model to identify well-formed trending topics. To enable the user to evaluate the temporal clustering model's performance, we fit several K-Means models, one for each k value ranging from 5 to 30 and compute the average silhouette score for each clustering run. As a result, the best k value would be the one that has the highest average silhouette score. In addition, to enable analysis of the model's performance using bias amounts, we perform 200 clustering runs on every aforementioned k value for various bias amounts ranging from 0.01 to 20 for a total of 5,200 clustering runs. The evaluation result returned from the process for each k value is saved as a csv file for later retrieval.

3.4. Dashboard Construction

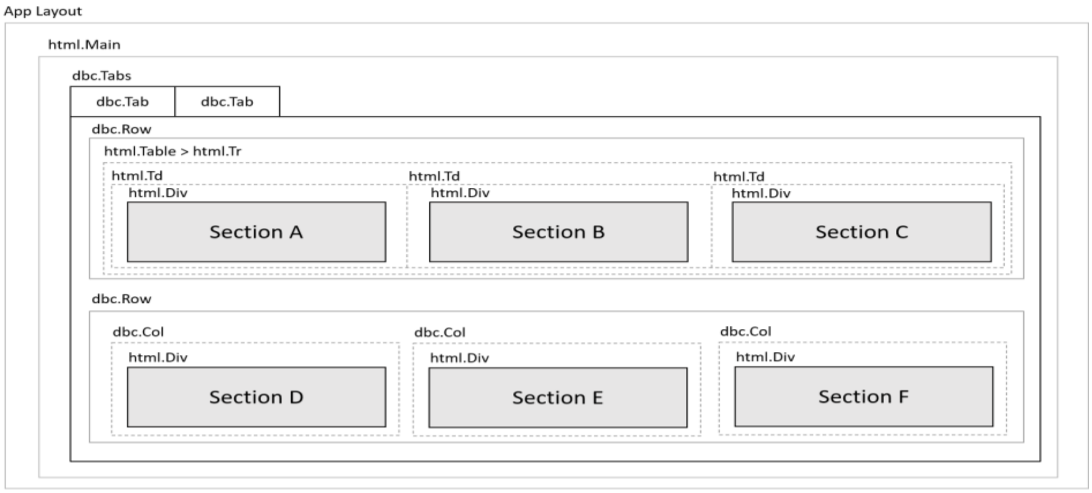
For dashboard construction, we followed Sadiku et al.'s (2016) approach to building an interactive dashboard: 1) create visualizations using Plotly, 2) build a web-based application using Dash, and 3) organize the contents of the dashboard into tabs. In contrast to Sadiku et al.'s (2016) study, which emphasized analysis of COVID-19, our work specifies the approaches for constructing the dashboard and then shows the users how to use the dashboard to detect journal trends.

3.4.1. Application Layout

To build the user interface for our dashboard, we utilized Dash's HTML components and bootstrap components to lay out the contents on the web page. The HTML components (Dash HTML Components, n.d.) are used to generate HTML tables and divisions for arranging various sections on the web page. In contrast to Dash's HTML components, Dash's bootstrap components (Dash Bootstrap Components, n.d.) arrange the contents of the web page using tabs, rows and columns, instead of tables and divisions.

Figure 4 shows a general layout of a web-based application that we used as a template for building our dashboard. First, we created an application layout containing the HTML's main component. Inside the HTML's main component, we used bootstrap's tab components, dbc.Tabs and dbc.Tab, to place the tabs on the web page. Next, we used bootstrap's row component (dbc.Row) to separate the contents of the tab horizontally. To arrange the sections vertically and then align the sections horizontally, we outlined two different approaches, as depicted on Figure 4. The first approach, illustrated as the gray box under the first dbc.Row, used HTML's table component (html.Table) to lay out the sections. The illustrated table (html.Table) had one row (html.Tr) and three cells (html.Td). Inside each table's cell, we placed an HTML's division component (html.Div) that included the contents of a given section. The second approach, illustrated as the gray box under the second dbc.Row, used bootstrap's column component (dbc.Col) to arrange the sections vertically and then placed an html.Div inside the column to show the contents of a specific section.

Figure 4
A general layout of a web-based application that serves as a template for building the dashboard. The gray lines are invisible on the web page.



3.4.2. Application Functionality

To enhance users' interactivity within the application, we incorporated Dash's data table (Dash DataTable, n.d.) which is used to build a table with search enabled, Dash HTML's button component which is used to create a button, and several Dash's core components (Dash Core Components, n.d.), including:

- Input component is used to build a text box that enables the user to enter a value as an input.
- Graph component is used to display a visualization created by Plotly.
- RangeSlider component is used to build a horizontal or vertical slider that enables the user to specify a range of values as inputs.
- Slider component is used to build a horizontal or vertical slider that enables the user to specify a single value as an input.
- Checklist component is used to build a group of checkboxes that enables the user to select multiple values as inputs.
- Loading component is used to display a progress bar while waiting for the data to be loaded on the screen.
- Store component is used to store the data inside the web page as JavaScript Object Notation (JSON) format.

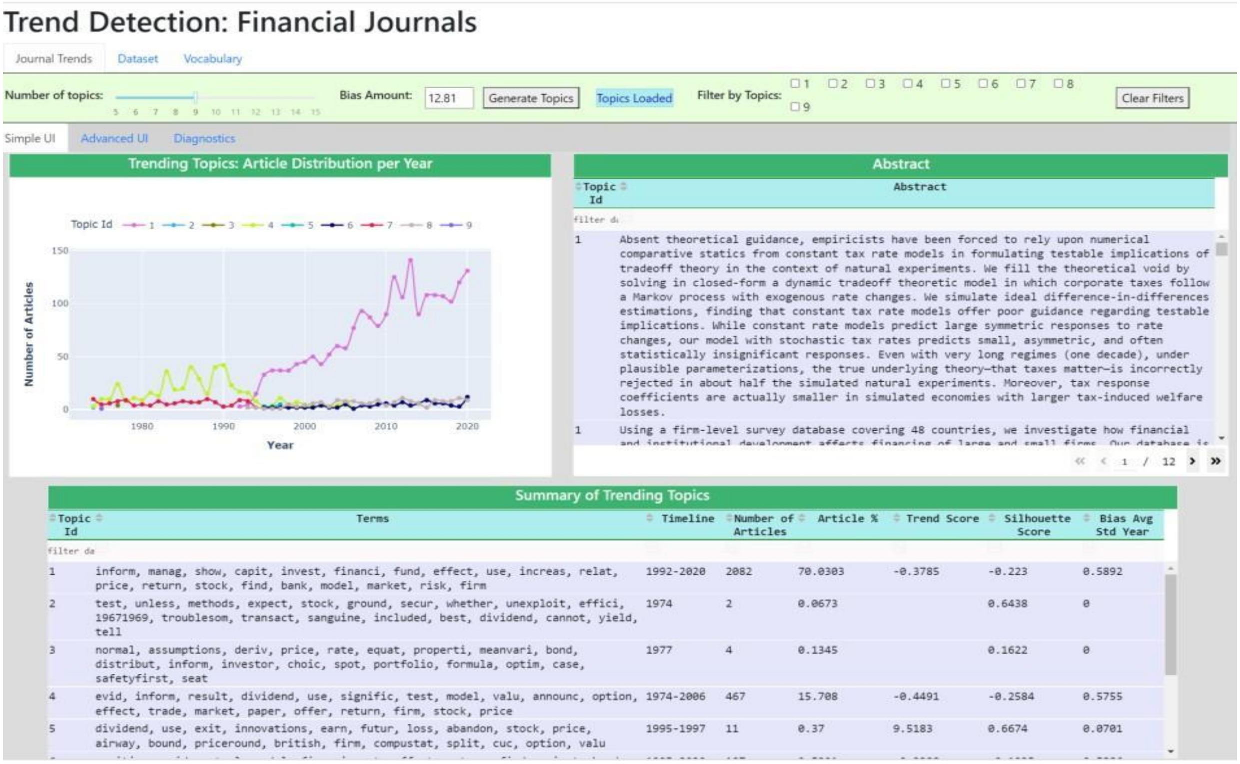
After building the components that enable interaction between the user and the application, we defined several callback functions (Basic Dash Callbacks, n.d.) to update the web page's contents in response to the user's input. One drawback of the callback function is that every time the user changes the value of a single component such as a text box, the application automatically applies the user's input and loads the new data on the screen. This means that if we have three different text boxes, the application would reload the screen three times when the user changes the values for all three text boxes. For a simple application, this may not be an issue. However, for a complex application, the user may have to wait a long time for the application to load the data that reflects all of the user's changes to the text boxes. To prevent the application from reloading data on the screen until a button is clicked, instead of passing the value of a Dash core component or a Dash HTML component to the callback function as an Input, we passed the corresponding value as a State (Basic Dash Callbacks, n.d.). This tells the application not to refresh data on the screen until a desired action, such as a clicked button, is triggered.

4. DASHBOARD DESIGN

The journal trend dashboard has three main tabs: 1) journal trends, 2) dataset, and 3) vocabulary. The contents of the journal trends tab are displayed on the screen when the user first accesses the application. Figure 5 displays a screenshot of the journal trends tab which is visible when the user first accesses the dashboard.

Figure 5

A screenshot of the Trend Detection dashboard. The contents of the Journal Trends tab is visible on the screen when the user first accesses the application.



4.1. Journal Trends Tab

The Journal Trends tab has two main sections: the selection section and the content section (Figure 5). The first section is the selection section which is located in the first row of the journal trends tab. The selection section has a light green background and serves as a query section for driving the information shown on the content section based on the user selections. This section has three main components arranged left to right horizontally: 1) a slider for selecting the number of trending topics, 2) an input text box for specifying a bias amount, and 3) a group of topic checkboxes that enables the user to select the trending topics to be displayed on the screen. The second section is the content section which has a light gray background. The content section contains three sub-tabs: Simple UI, Advanced UI, and Diagnostics.

4.1.1. Simple UI Tab

The objective of the Simple UI sub-tab is to provide high-level information about the topics and their trends over time. As shown in Figure 5, the Simple UI tab has three sections: 1) a trending topics section which includes a figure of the yearly article distribution per trending topic, 2) an abstract section which contains a table of abstracts belonging to particular trending topics, and 3) a summary of trending topics section which includes a summary table of trending topics.

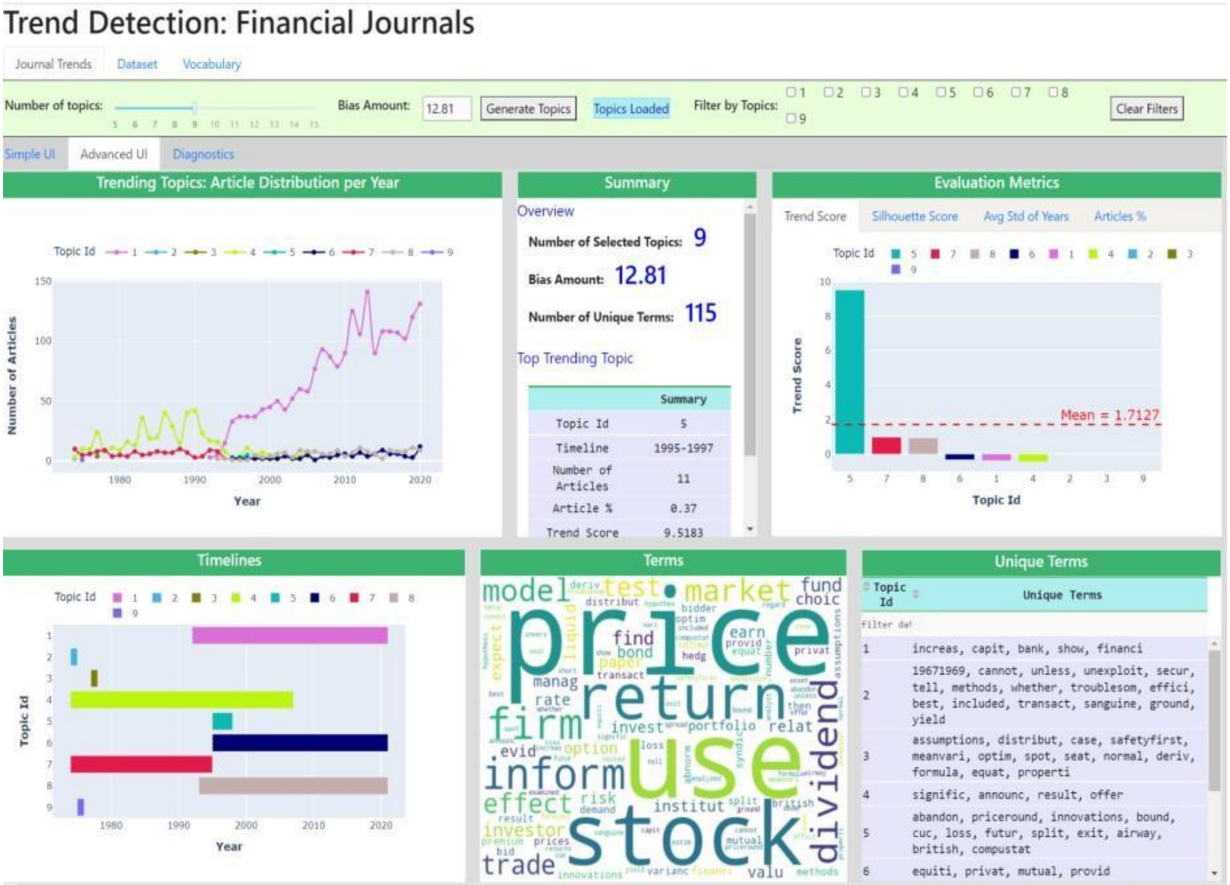
4.1.2. Advanced UI Tab

The advanced UI tab enables the user to conduct a comprehensive analysis of the generated trending topics.

Figure 6 provides a screenshot of the advanced UI tab.

Figure 6

A screenshot of the Advanced UI Tab of Journal Trend Detection dashboard



As shown in Figure 6, the advanced UI tab contains the following six sections:

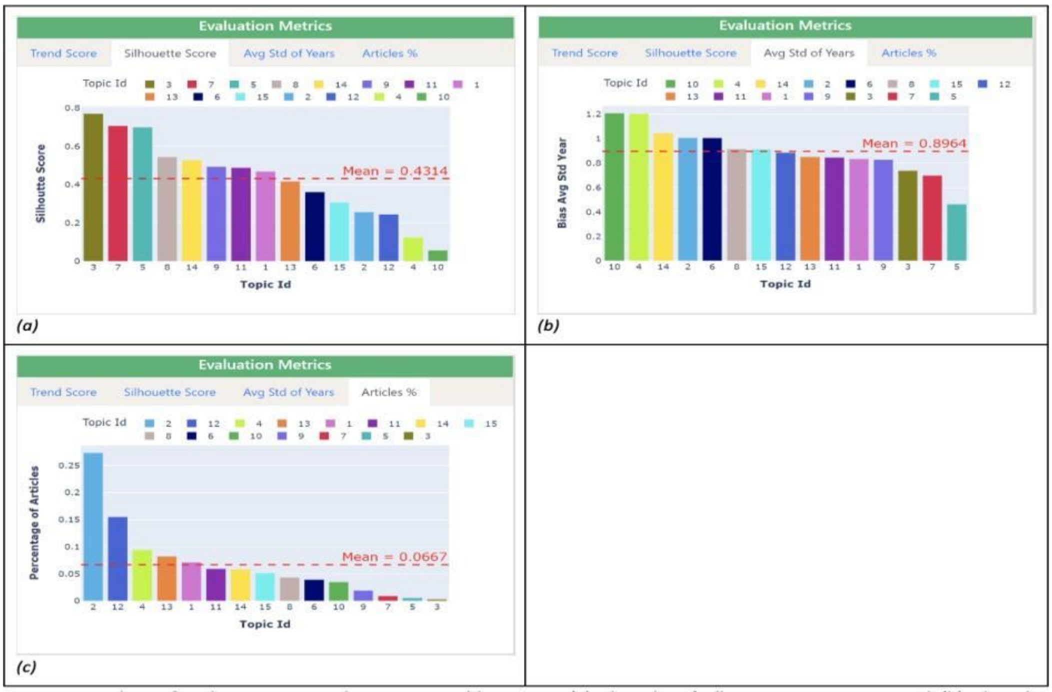
- The trending topics section includes a figure of the yearly article distribution per trending topic.
- The summary section provides a general overview of the trending topics. These include the number of topics shown on the screen and the evaluation metrics for the top trending topic.
- The evaluation metrics section contains four tabs. Each tab displays a vertical bar plot of an evaluation metric including trend score, silhouette score, average standard deviation of years, and percentage of articles per topic.
- The timelines section includes a plot of timelines for the trending topics.
- The terms section includes a figure of a word cloud of the identified terms for all trending topics. The bigger the size of a term in the word cloud, the greater the number of trending topics containing the corresponding term.
- The unique terms section contains a table of a table of unique terms per

trending topic. Figure 7 displays the bar plots of evaluation metrics that are not

visible in Figure 6.

Figure 7

Screenshots of evaluation metrics that are not visible in Fig. 6: (a) a bar plot of Silhouette score per topic id, (b) a bar plot of bias average standard deviation of year per topic id, and (c) a bar plot of



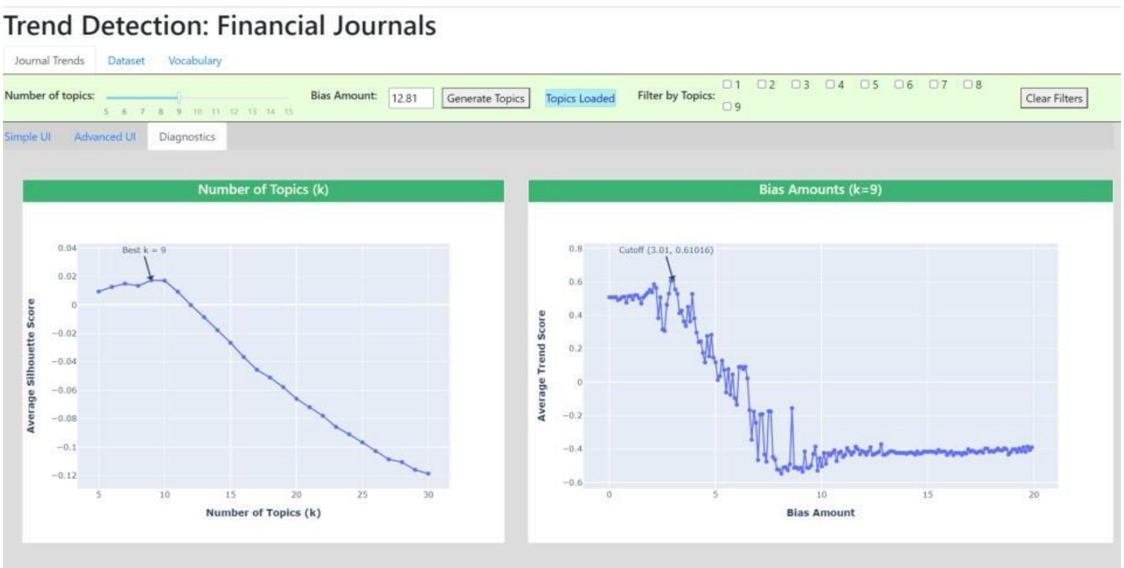
article percentage per topic id.

4.1.3. Diagnostics Tab

The diagnostics tab contains two plots for evaluating the temporal clustering model's performance on various k values and bias amounts. As illustrated in Figure 8, the first plot, located under the Number of Topics section, displays the average silhouette score for various k values ranging from 5 to 30. The second plot, located under the Bias Amounts section, displays the average trend score for various bias amounts ranging from 0.1 to 20 for a number of topics (k value) specified by the user.

Figure 8

A screenshot of the Diagnostics sub-tab included in the Journal Trends tab

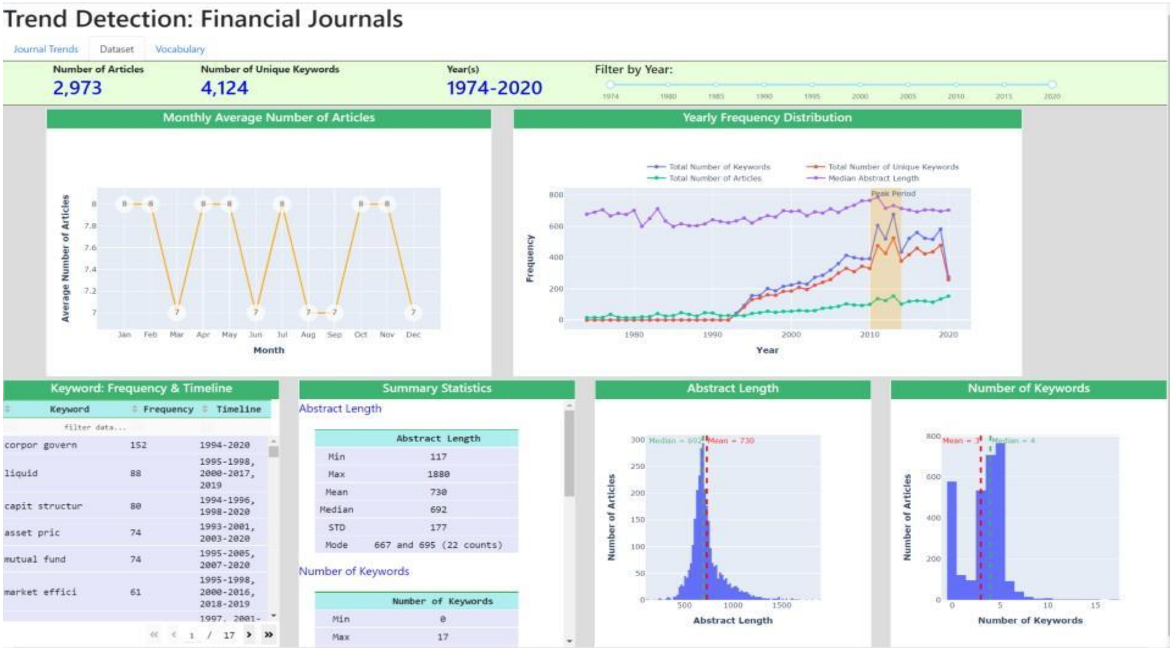


4.2. Dataset Tab

The dataset tab has two main purposes: 1) enables the user to gain an insight of the raw data through exploratory data analysis, and 2) enables the user to detect and correct any discrepancies in the dataset such as data entry errors, missing values and so on. Figure 9 shows a screenshot of the dataset tab.

Figure 9

A screenshot of the Dataset tab of Trend Detection dashboard



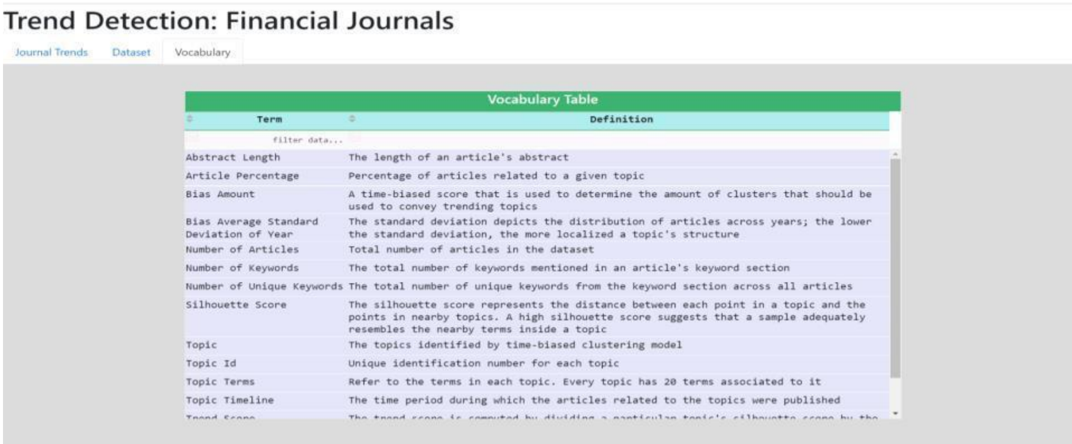
According to Figure 9, the dataset tab contains two main sections. First, the header section, which has a light green background, contains general information regarding the raw data such as number of articles, number of unique keywords, and the time periods of the data shown on the screen. Included in the header section is a light blue slider which enables the user to select the time periods for the data. Second, the content section, which has a light gray background, contains the following six sub-sections:

- The monthly average number of articles section includes a line plot of the average number of articles published per month.
- The yearly frequency distribution section includes line plots of the following:
    - median abstract length (purple line)
    - total number of keywords (blue line)
    - total number of unique keywords (red line)
    - total number of articles (green line)
- The keyword's frequency & timeline section includes a table that displays the frequency and timeline for each keyword in the raw data.
- The summary statistics section displays the summary statistics tables for abstract length, number of keywords, and year and month of publication.
- The abstract length section includes a histogram of the abstract length.
- The number of keywords section includes a histogram of the number of keywords.

4.3. Vocabulary Tab

The vocabulary tab, shown in Figure 10, provides a definition for each term included in the dashboard.

Figure 10
A screenshot of the Vocabulary tab of the Trend Detection dashboard

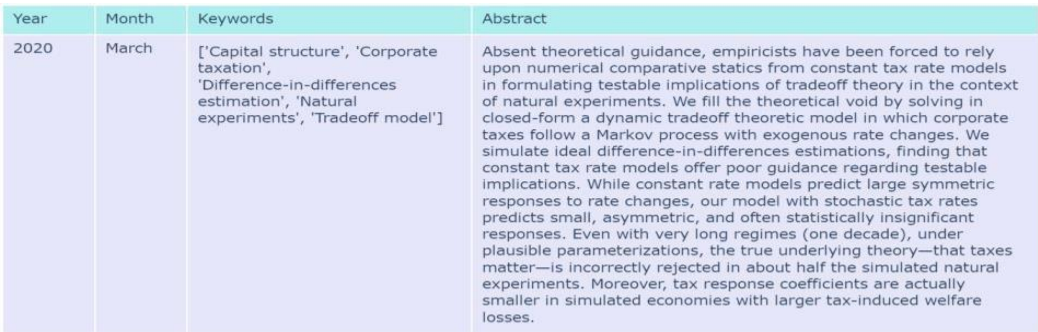


Term	Definition
Abstract Length	The length of an article's abstract
Article Percentage	Percentage of articles related to a given topic
Bias Amount	A time-biased score that is used to determine the amount of clusters that should be used to convey trending topics
Bias Average Standard Deviation of Year	The standard deviation depicts the distribution of articles across years; the lower the standard deviation, the more localized a topic's structure
Number of Articles	Total number of articles in the dataset
Number of Keywords	The total number of keywords mentioned in an article's keyword section
Number of Unique Keywords	The total number of unique keywords from the keyword section across all articles
Silhouette Score	The silhouette score represents the distance between each point in a topic and the points in nearby topics. A high silhouette score suggests that a sample adequately resembles the nearby terms inside a topic
Topic	The topics identified by time-biased clustering model
Topic Id	Unique identification number for each topic
Topic Terms	Refer to the terms in each topic. Every topic has 20 terms associated to it
Topic Timeline	The time period during which the articles related to the topics were published

5. DATASET

The dataset contains a collection of 2,973 article abstracts published between 1974 and 2020 in the “Journal of Financial Economics” retrieved from ScienceDirect. Each record in the dataset has four features: year, month, keywords, and article’s abstract. A sample record from the dataset is shown in Figure 11.

Figure 11
Sample record from the dataset



Year	Month	Keywords	Abstract
2020	March	['Capital structure', 'Corporate taxation', 'Difference-in-differences estimation', 'Natural experiments', 'Tradeoff model']	Absent theoretical guidance, empiricists have been forced to rely upon numerical comparative statics from constant tax rate models in formulating testable implications of tradeoff theory in the context of natural experiments. We fill the theoretical void by solving in closed-form a dynamic tradeoff theoretic model in which corporate taxes follow a Markov process with exogenous rate changes. We simulate ideal difference-in-differences estimations, finding that constant tax rate models offer poor guidance regarding testable implications. While constant rate models predict large symmetric responses to rate changes, our model with stochastic tax rates predicts small, asymmetric, and often statistically insignificant responses. Even with very long regimes (one decade), under plausible parameterizations, the true underlying theory—that taxes matter—is incorrectly rejected in about half the simulated natural experiments. Moreover, tax response coefficients are actually smaller in simulated economies with larger tax-induced welfare losses.

Summary statistics of journal articles are provided in Table 1. According to Table 1, the abstract length ranged from 117 to 1,880 words with the mean and median of 730 and 692 words, respectively. The most typical abstract lengths were 667 and 695 words, each having a corresponding 22 articles. The number of keywords range from 0 to 17, with a mean and median of 3 and 4 keywords, respectively. The mode value indicates that most articles have five keywords.

Table 1
Summary statistics of journal articles

	Min	Max	Mean	Median	Standard Deviation	Mode
Abstract Length	117	1,880	730	692	177	667 and 695 (22 counts each)
Number of Keywords	0	17	3	4	2	5 (766 counts)

Figure 12 provides the data distribution of abstract length (a) and the number of keywords (b), respectively.

The data distribution for abstract length is right-skewed, with most articles having abstract lengths close to the median. Few articles have an abstract length of less than 400 words or greater than 1,200 words. Similarly, the data distribution for the number of keywords is also skewed, with most articles having four or five keywords while there are only a few articles with more than five keywords.

Figure 12
(a) Distribution of abstract length (b) Distribution of number of keywords

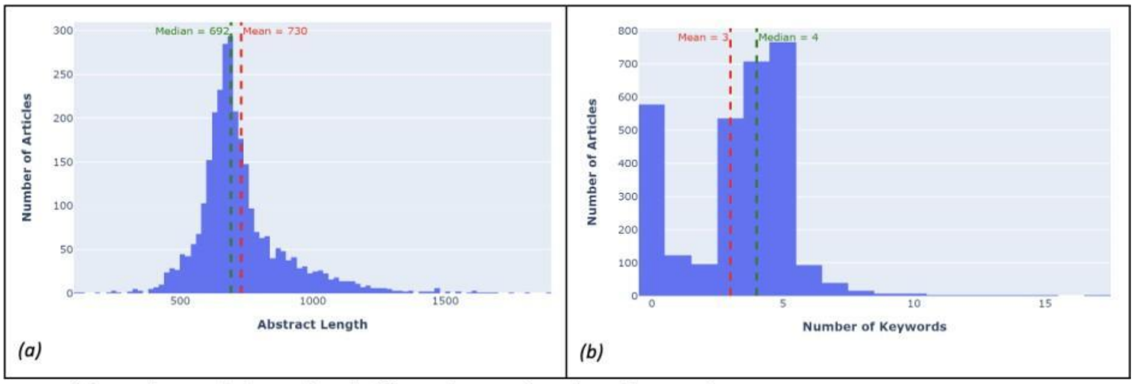


Figure 13 displays the monthly average number of articles published over 46 years. According to the figure, on average, 7 or 8 articles are published each month.

Figure 13
Monthly average number of article distribution over 46 periods

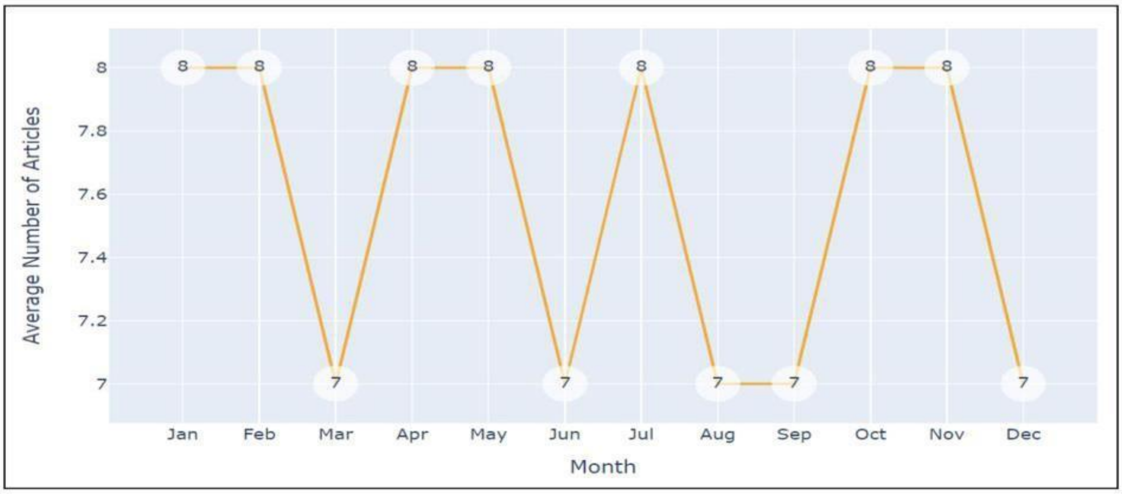


Figure 14 shows the frequency distribution of the total number of articles, the total number of keywords, the total number of unique keywords, and the median abstract length in the form of a line graph. From Figure 14, it is evident that the highest total number of articles was published between 2011 and 2013. Corresponding to the highest total number of articles, the median abstract length, the total number of keywords, and the total number of unique keywords were also highest during the same period. Overall, the general trend for the total number of articles, the total number of keywords, and the total number of unique keywords seems to be increasing. One more interesting observation from Figure 13 is that articles published prior to 1993 did not contain keywords in the print volumes.

Figure 14
Frequency Distribution per Year

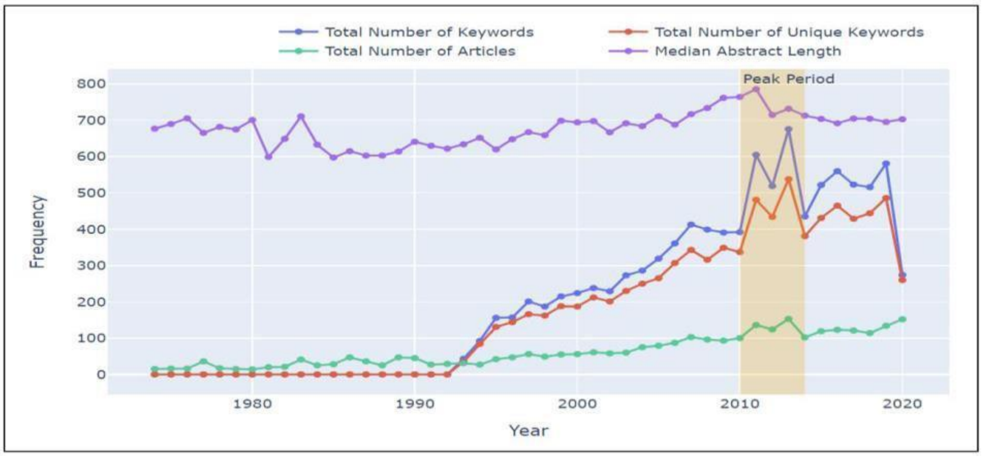


Figure 15 shows the top 15 keywords and their corresponding timeline, sorted in descending order of their frequency. According to the figure, the keyword "corporate governance" has the highest frequency, followed by the two keywords, "liquidity" and "capital structure." In addition, from the timeline column, it can be inferred that the trend for the keyword "corporate governance" is continuous during the periods between 1994 and 2020. Unlike the trend of the keyword "corporate governance," all other keywords have discontinuous trends. Despite having discontinuous trends, the keywords "liquidity" and "market efficiency" have the longest continuous timeline of 17 and 16 years, respectively.

Figure 15
Top 15 keywords with corresponding timeline

Keyword	Frequency	Timeline
corpor govern	152	1994-2020
liquid	88	1995-1998, 2000-2017, 2019
capit structur	80	1994-1996, 1998-2020
asset pric	74	1993-2001, 2003-2020
mutual fund	74	1995-2005, 2007-2020
market effici	61	1995-1998, 2000-2016, 2018-2019
merger and acquisit	55	1997, 2001-2005, 2007-2009, 2011-2019
execut compen	55	1995, 1997, 1999-2003, 2005-2014, 2016-2020
bank	52	1994-1998, 2001, 2003-2005, 2007-2009, 2011-2019
initi public off	46	1993-1997, 1999-2003, 2005-2010, 2012-2014, 2016, 2019
invest	46	1996, 1998-2001, 2003-2009, 2011-2013, 2015-2020
return predict	45	1999, 2001-2009, 2011-2015, 2017-2020
hedg fund	44	2002, 2004-2005, 2009-2019
institut investor	44	1995-1997, 1999, 2002-2005, 2007-2015, 2017-2019
merger	41	1994-1996, 1998-2014, 2017, 2019

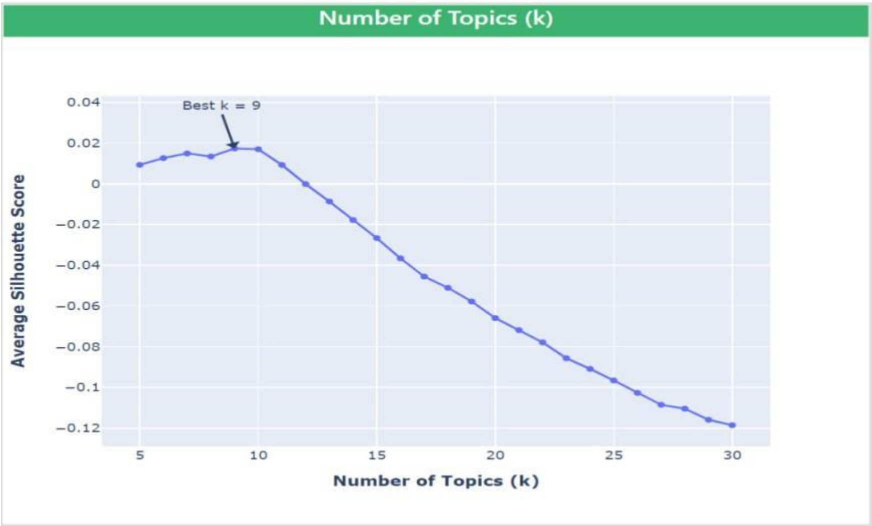
6. FINDINGS

Before using the dashboard to detect journal trends, we access the Diagnostic sub-tab to identify the best number of topics along with a low and a high bias amount value. Then, we use the Simple UI and Advanced UI sub- tabs to analyze the trending topics for the optimal k value on each bias level.

6.1. Best Number of Topics (k value) Identification

To select the optimal number of topics, we analyze the figure under the Number of Topics section of the Diagnostics sub-tab. As shown in Figure 16, the best number of topics is 9 along with the corresponding average silhouette score of 0.0173. Furthermore, the average silhouette score appears to rise from k=5 to k=9, with a tiny dip at k=8, before beginning to fall after k=9.

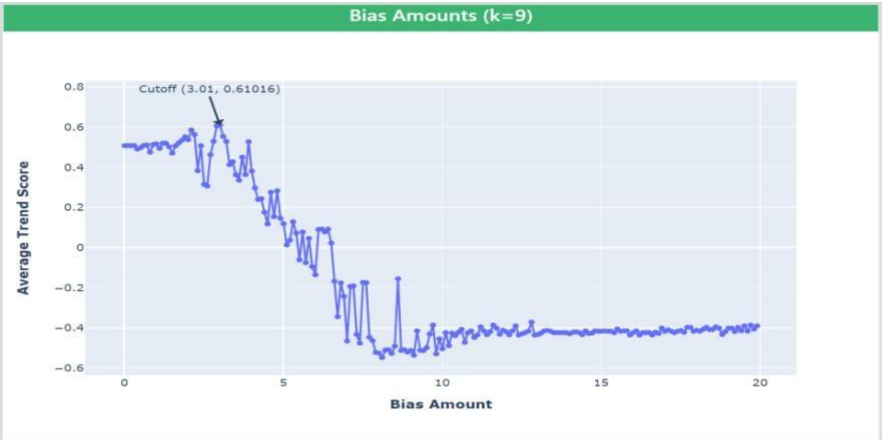
Figure 16
Plot of average silhouette score per number of topics (k)



6.2. Bias Levels Identification

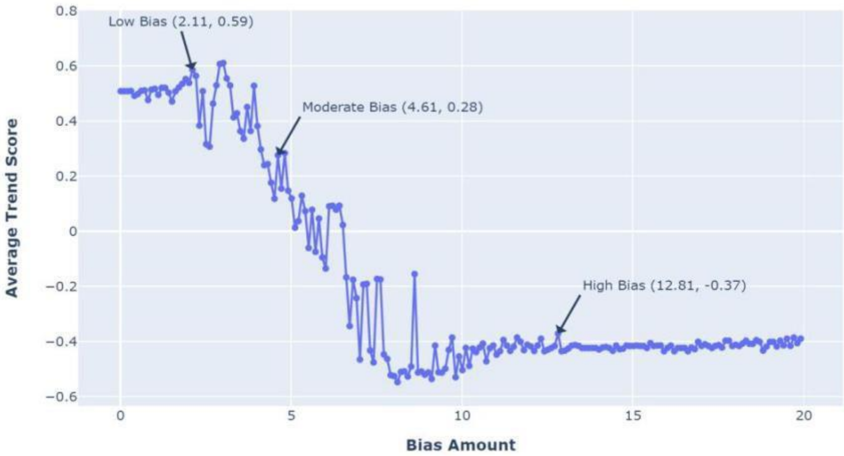
Following the same strategy for bias levels identification proposed by Behpour et al. (2021), we identify a bias amount with a highest average trend score as a suitable cutoff point. At the cutoff point, we recognize the corresponding average trend score as the baseline trend score. Figure 17 displays the average trend score for each bias amount ranging from 0.1 to 20 at selected k value of 9, the optimal number of topics. As shown in Figure 17, the baseline trend score is 0.61.

Figure 17
Line plot of average trend score per bias amount for k = 9. A cutoff point is identified at a bias amount of 3.01 and an average trend score of 0.61.



To identify a low bias level, we look for a bias amount with a local maximum average trend score, or the local's highest average trend score, before the cutoff point. Using the same plot under the Bias Amounts section, we recognize a bias amount of 2.11 as a low bias amount. Since the average trend scores seem to stay constant beyond the bias amount of 12.81, we recognize 12.81 as a high bias amount. For a moderate bias amount, we pick a value between low and high bias amounts, which is 4.61. Figure 18 displays the average trend score at each three bias levels: low, moderate, and high.

Figure 18
Line plot of average trend score per bias amount for k = 9. The average trend score for low, moderate, and high bias amounts are 0.59, 0.28, and -0.37, respectively.



6.3. Analysis of Trending Topics

Having identified the optimal number of topics and the baseline trend score, along with the low and high bias amounts, we further use the journal trend dashboard to analyze the trending topics for each bias level on the best k value, which is 9. First, we specify the number of topics and the bias amount on the header section of the Journal Trends tab. Second, we access the Simple UI sub-tab and filter the table under the Summary of Trending Topics section to display only the topics that have a trend score above the baseline trend score, which is 0.61. Third, we filter the topics by checking the checkbox for topics that have the trend score above the baseline. Fourth, we access the Advanced UI tab and analyze the timeline, evaluation metrics, and topic terms. Next, we repeat the same process for a different bias level on the same k value.

6.3.1. Low Temporal Bias

For k=9 and a low temporal bias amount of 2.11, we identified four top trending topics which had a trend score above the baseline. Topics are listed in Figure 19 in the order of highest to lowest trend score. As shown in Figure 19, the recognized trending topics at the low bias level seem to have a high bias average standard deviation of year ranging from 0.74 to 0.90.

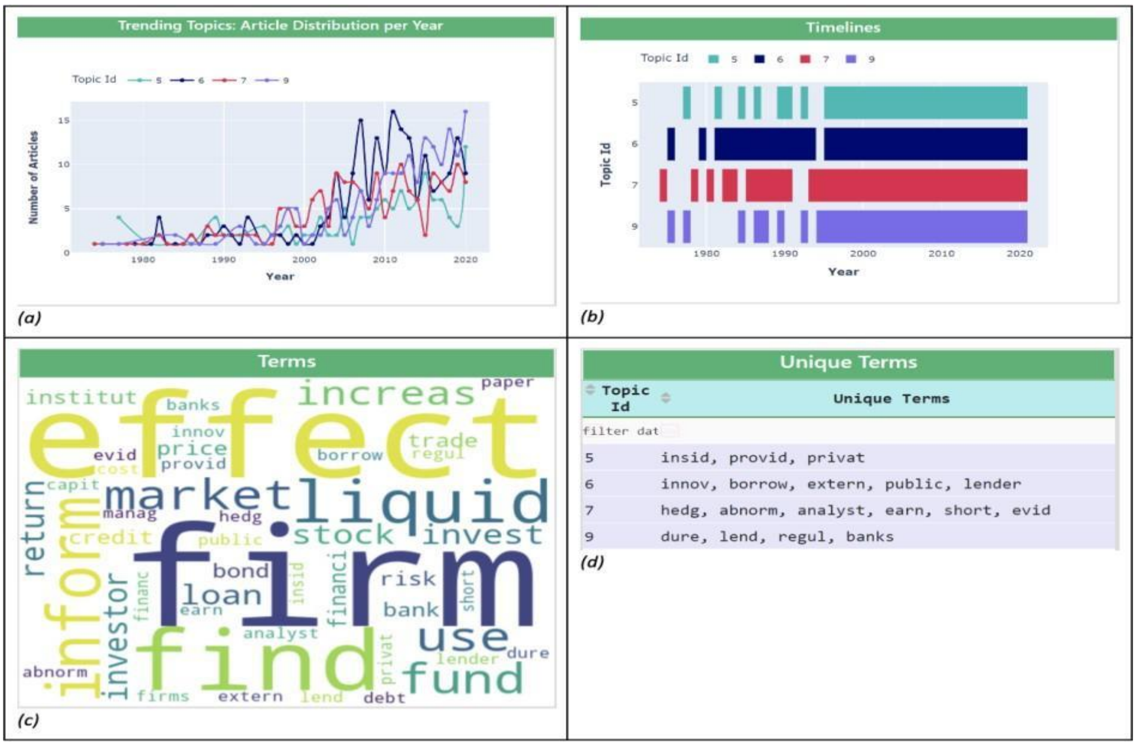
Figure 19
A screenshot of summary of trending topics table for k=9 and bias amount = 2.11, filtered by trend score (>=0.61)

Summary of Trending Topics

Topic Id	Terms	Timeline	Number of Articles	Article %	Trend Score	Silhouette Score	Bias Avg Std Year
filter da					>=0.61		
9	cost, use, invest, find, effect, risk, dure, regul, increas, fund, capit, market, liquid, firm, loan, credit, financi, lend, banks, bank	1975, 1977, 1984, 1986-1987, 1989, 1992, 1994-2020	187	6.2899	0.7848	0.5882	0.7495
5	insid, paper, provid, stock, invest, privat, firm, effect, manag, find, return, bond, use, price, investor, institut, liquid, trade, inform, fund	1977, 1981, 1984, 1986, 1989-1990, 1992, 1995-2020	126	4.2381	0.6229	0.5616	0.9017
6	bond, innov, risk, extern, increas, lender, public, financi, firms, market, find, effect, borrow, inform, financ, credit, bank, debt, loan, firm	1975, 1979, 1981-1993, 1995-2020	215	7.2318	0.68	0.5602	0.8239
7	effect, hedg, institut, evid, use, abnorm, short, firm, investor, earn, find, price, fund, return, inform, analyst, liquid, market, stock, trade	1974, 1978, 1980, 1982-1983, 1985-1990, 1993-2020	179	6.0209	0.6285	0.5249	0.8351

Based on the plot of yearly distribution of number of articles per trending topic (Figure 20a), the identified trending topics are highly overlapped. Furthermore, each trending topic does not have a well-defined trends timeline as there are several gaps in the timeline for each trending topic (Figure 20b). Additionally, the most common terms across all four topics are ‘effect’ and ‘firm’ (Figure 20c). There also seems to be a few unique terms identified for each trending topic (Figure 20d).

Figure 20
Timeline analysis of top four low temporal bias clusters for k=9 and bias amount = 2.11, filtered by trend score (>=0.61): (a) yearly distribution of number of articles per trending topic, (b) timeline of the top four trending topics, (c) a word cloud of all recognized terms for the top four trending topics and (d) recognized unique terms per trending topic.



6.3.2. Moderate Temporal Bias

For k=9 and a moderate bias amount of 4.61, we identified three topics having a trend score of 0.61 or above (Figure 21). Among the three trending topics, topic 2 is more localized because topic 2 has the lowest bias average standard deviation of year, Silhouette score, percentage of articles, and timeline range (7 years).

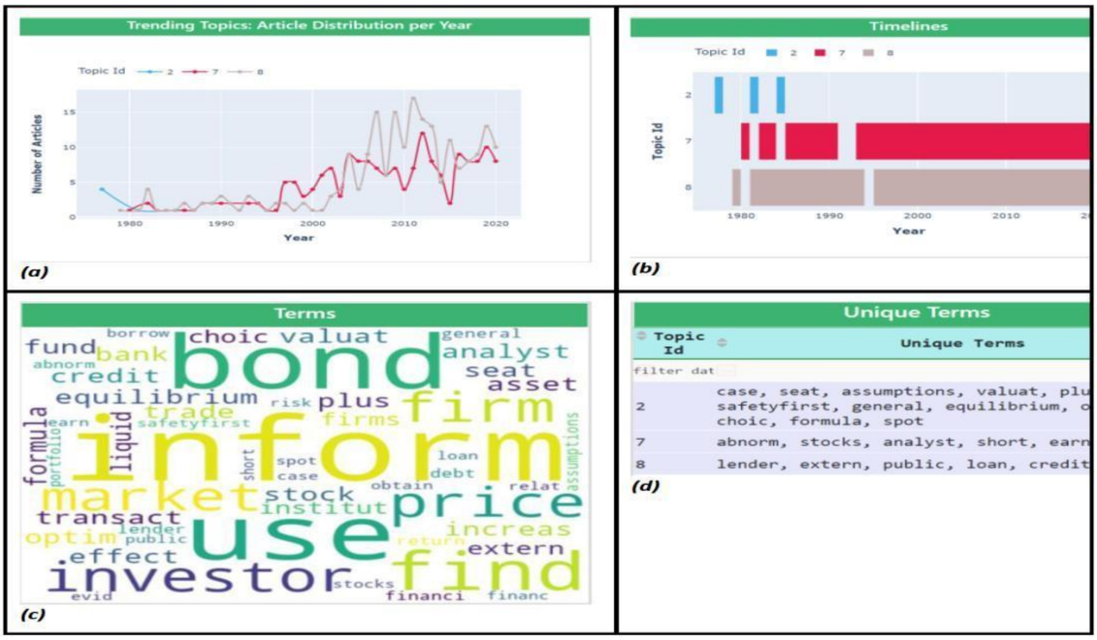
Figure 21
A screenshot of summary of trending topics table for k=9 and bias amount = 4.61, filtered by trend score (>=0.61)

Summary of Trending Topics

Topic Id	Terms	Timeline	Number of Articles	Article %	Trend Score	Silhouette Score	Bias Avg Std Year
filter da					>=0.61		
2	inform, obtain, asset, general, valuat, plus, choic, transact, equilibrium, spot, portfolio, formula, optim, investor, assumptions, price, bond, safetyfirst, seat, case	1977, 1981, 1984	6	0.2018	1.4942	0.3437	0.23
7	stocks, relat, institut, use, evid, abnorm, firm, short, earn, investor, find, price, fund, return, inform, liquid, analyst, market, stock, trade	1980, 1982-1983, 1985-1990, 1993-2020	179	6.0209	0.6894	0.5334	0.7737
8	bond, use, lender, risk, public, increas, extern, financi, firms, market, find, effect, borrow, inform, financ, credit, bank, debt, loan, firm	1979, 1981-1993, 1995-2020	217	7.299	0.7164	0.5701	0.7958

As shown in figure 22b, the trend for topic 2 starts 2-3 years before the trends for topics 7 and 8 start and ends shortly 5 or 6 years after. All three topics share common terms such as ‘investor’, ‘use’, ‘bond’, ‘inform’, and so on (Figure 22c). These terms when combined together indicate a common trend among the three topics such as firm investors use market information to find bond price. Before 2006, the trends for topic 7 and 8 seem highly overlapped (Figure 22a). After 2006, the trend for topic 8 seems stronger than that of topic 7. Figure 22d also indicates that topic 2 has the most number of unique terms although topic 2 has a shorter timeline than that of topic 7 and 8.

Figure 22
Timeline analysis of top three moderate temporal bias clusters for k=9 and bias amount = 4.61, filtered by trend score (>=0.61): (a) yearly distribution of number of articles per trending topic, (b) timeline of the top three trending topics, (c) a word cloud of all recognized terms for the top three trending topics and (d) recognized unique terms per trending topic.



6.3.3. High Temporal Bias

For k=9 and a high bias amount of 12.81, we identified three top trending topics which had a trend score above the baseline (Figure 23). Each identified trending topic also has a high trend score which is above 0.9. The top trending topic (topic 5) has a very low percentage of articles (0.37) and bias average standard deviation of year (0.07) compared to those of topics 7 and 8.

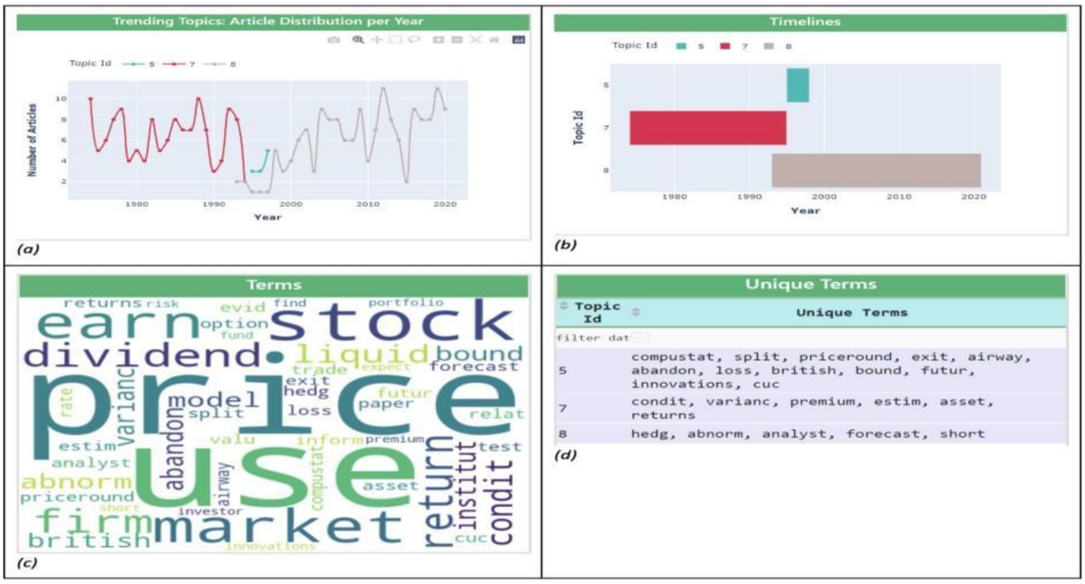
Figure 23
A screenshot of summary of trending topics table for k=9 and bias amount = 12.81, filtered by trend score (>=0.61)

Summary of Trending Topics

Topic Id	Terms	Timeline	Number of Articles	Article %	Trend Score	Silhouette Score	Bias Avg Std Year
filter da					>=0.61		
5	dividend, use, exit, innovations, earn, futur, loss, abandon, stock, price, airway, bound, priceround, british, firm, compustat, split, cuc, option, valu	1995-1997	11	0.37	9.5183	0.6674	0.0701
8	abnorm, hedg, forecast, evid, use, earn, institut, short, firm, price, investor, find, return, fund, inform, liquid, analyst, market, stock, trade	1993-2020	164	5.5163	0.9369	0.5584	0.596
7	use, relat, varianc, dividend, returns, condit, premium, rate, paper, test, stock, market, portfolio, risk, estim, price, asset, expect, model, return	1974-1994	135	4.5409	0.9774	0.4957	0.5072

As indicated in Figure 24a, the trends for the high bias topics seem to be well defined with a little overlap in the timeline. However, the timeline for each trending topic is continuous with no gaps. There seems to be a connecting trend between topic 7 and topic 8 where the trend for topic 7 ends in 1994 and the trend for topic 8 starts in 1993 and then continues to 2020 (Figure 24b). Though the trends for topic 5 overlap with the trends for topic 8, the trends for topic 5 only last for 2 years from 1995 to 1997. Additionally, the most common terms across all three topics are ‘price’ and ‘use’ (Figure 24c). There also seem to be more unique terms identified for the top trending topic, topic 5 (Figure 24d).

Figure 24
Timeline analysis of top three high temporal bias clusters for k=9 and bias amount = 12.81, filtered by trend score (>=0.61): (a) yearly distribution of number of articles per trending topic, (b) timeline of the top three trending topics, (c) a word cloud of all recognized terms for the top three trending topics and (d) recognized unique terms per trending topic.



6.3.4. Discussion of Results

From sections 6.3.1 to 6.3.3, it is evident that as the bias amount increases, the number of trending topics with a trend score of 0.61 or above decreases. Low temporal bias trending topics are highly overlapped, and there is no well-defined timeline for all the top trending topics. With a moderate bias amount, we start seeing one or more trending topics that have a better-defined timeline. The line plot of moderate temporal bias trending topics is also less cluttered compared to that of low temporal bias trending topics. In contrast, high temporal bias trending topics seem to be less overlapped and have a well-defined and continuous timeline. In addition, the number of unique keywords increases as the bias amount increases. Overall, increasing the bias amount results in fewer trending topics with a better-defined timeline of trends. The resulting trending topics are also less cluttered as the bias increases.
However, as the value of bias goes beyond the high level, the trending topics are highly dependent on time instead of the information in the existing term corpus.

7. CONCLUSION AND FUTURE WORK

Trend analysis is crucial since it gives in-depth information and helps in making better decisions. We created an interactive dashboard that displays various information regarding trends in a financial journal. Following Behpour et al. (2021) paper, a temporal clustering method was used to extract the topics and to generate journal trends. Flask web development framework, Plotly's Dash, and Python were used to create the dashboard. This paper makes a significant contribution in the form of an interactive trend detection dashboard that visualizes the most popular topics in a journal over time and drives insight into actions that lead to optimal outcomes. The dashboard can be used to understand and experiment with the data and journal trends. The dashboard also enables the user to understand the effect of bias amount on the topics and their timelines.

The trend detection dashboard has four potential applications. First, journal editors can use this dashboard to see what topics are in trend and understand the development of knowledge created in the field. Thus, journal editors can use this dashboard to deliver information to their users about their journal's trending topics. Second, the trend dashboard enables researchers to explore the trending topics related to their field of study by offering the capability of performing ad-hoc analysis. This further enables the researchers to select a high-trending topic that would attract the readers' interest while increasing their chance of obtaining funds. Third, research funding institutions can use the trend dashboard to learn the trending topics and efficiently and effectively allocate more funds to the most prominent and high trending research projects. Last, because this dashboard includes granular information about data, topics, and terms, it paves the way for an in-depth investigation into why particular topics were trending at a given time. Because there aren't many studies done to produce an interactive dashboard for journal trends in the financial field, the interactive dashboard is an attempt to fill in the existing research gap.

The dashboard built in this study uses data from only the Journal of Financial Economics. More journal data from various areas can be collected and added to this dashboard in the future to examine journal trends in various fields. As the amount of data collected as part of a more extensive study grows, the number of topics can also be accommodated. When a trending topic contains all articles that are published within the same year, we are unable to compute the trend score because the bias average standard deviation of years would have a value of zero. For future research, the time-biased clustering algorithm can be modified to view only those topics with a trend score. Finally, the current research hosts the interactive dashboard locally. The interactive dashboard can be enhanced in the future so that it can be hosted in the cloud or on servers.

External Repositories

The dashboard is available on Github repository. The link for the repository is https://github.com/Deep6Lab/Trending-Topics-Dashboard

REFERENCES

Abdolalizadeh, P., Chaibakhsh, S., & Falavarjani, K. G. (2021). Global burden of paediatric vision impairment: a trend analysis from 1990 to 2017. *Eye , 35*(8), 2136–2145.

Ali, N., Naveed, M., Aslam, S., & Bhatti, M. W. (2021). Research Trend of Journal of Information Science: A Bibliometric Analysis through Web of Science Database.

Aparicio, M., & Costa, C. J. (2015). Data visualization. *Commun. Des. Q. Rev, 3*(1), 7–11.

Asfaw, A., Simane, B., Hassen, A., & Bantider, A. (2018). Variability and time series trend analysis of rainfall and temperature in northcentral Ethiopia: A case study in Woleka sub-basin. *Weather and climate extremes, 19,* 29-41.

Aswell, S. (2020, October 20). *34 grant statistics for 2020*. https://blog.submittable.com/grant-statistics/

Behpour, S., Mohammadi, M., Albert, M. V., Alam, Z. S., Wang, L., & Xiao, T. (2021). Automatic trend detection: Time-biased document clustering. *Knowledge-Based Systems, 220,* 106907.

Brath, R., & Peters, M. (2004). Dashboard design: Why design is important. *DM Direct, 85,* 1011285–1011281.

Ceribasi, G., & Ceyhunlu, A. I. (2021). Analysis of total monthly precipitation of Susurluk Basin in Turkey using innovative polygon trend analysis method. *Journal of Water and Climate Change, 12*(5), 1532–1543.

Clement, F., Kaur, A., Sedghi, M., Krishnaswamy, D., & Punithakumar, K. (2020). Interactive Data Driven Visualization for COVID-19 with Trends, Analytics and Forecasting. *2020 24th International Conference Information Visualisation (IV)*, 593–598.

Dash core components. (n.d.). Retrieved November 24, 2021, from https://dash.plotly.com/dash-core-components

Dash bootstrap components. (n.d.). Retrieved November 24, 2021, from https://dash-bootstrap-components.opensource.faculty.ai/

Dash DataTable. (n.d.). Retrieved November 24, 2021, from https://dash.plotly.com/datatable

Dash HTML Components. (n.d.). Retrieved November 24, 2021, from https://dash.plotly.com/dash-html- components

Enyew, B., Odubela, K., Jiang, S., & Davis, L. (2020). An Interactive Dashboard to Study the Impact of Hurricane Florence on Food Bank Operations. *2020 International Conference on Computational Science and Computational Intelligence (CSCI)*, 1173–1178.

Getting grants - common reasons proposals are rejected. (n.d.). Retrieved October 7, 2021, from https://www.geneseo.edu/sponsored_research/common-reasons-proposals-are-rejected

Jang, J.-H., Won, B.-Y., Jang, G.-W., Kim, S.-A., Oh, S.-H., Kim, Y.-J., Ha, M.-O., Lee, Y.-E., Lee, G.-R., Park, S.-

S., & Um, S. (2017). Trend analysis of research in the journal of Korean society dental hygiene from 2001 to 2015. *Journal of Korean society of Dental Hygiene, 17*(4), 693–704.

Kranak, M. P., Falligant, J. M., Bradtke, P., Hausman, N. L., & Rooker, G. W. (2020). Authorship trends in the Journal of Applied Behavior Analysis : An update. In *Journal of Applied Behavior Analysis* (Vol. 53, Issue 4, pp. 2376–2384). https://doi.org/10.1002/jaba.726

Lee, Y.-J., Kim, E. J., & Kim, J. S. (2019). Research Trend Analysis of Publications in the Journal of Home Economics Education Association Using Network Text Analysis. *Journal of Korean Home Economics Education Association, 31*(4), 1–18.

Noghabaei, M., Heydarian, A., Balali, V., & Han, K. (2020). Trend Analysis on Adoption of Virtual and Augmented Reality in the Architecture, Engineering, and Construction Industry. *Data, 5*(1), 26.

Part 3. Basic Callbacks. (n.d.). Retrieved November 24, 2021, from https://dash.plotly.com/basic-callbacks Pourghasemi, H. R., Pouyan, S., Heidari, B., Farajzadeh, Z., Fallah Shamsi, S. R., Babaei, S., Khosravi, R., Etemadi,

M., Ghanbarian, G., Farhadi, A., Safaeian, R., Heidari, Z., Tarazkar, M. H., Tiefenbacher, J. P., Azmi, A., & Sadeghian, F. (2020). Spatial modeling, risk mapping, change detection, and outbreak trend analysis of coronavirus (COVID-19) in Iran (days between February 19 and June 14, 2020). *International Journal of Infectious Diseases: IJID: Official Publication of the International Society for Infectious Diseases, 98,* 90– 108.

Romano, S., Galante, H., Figueira, D., Mendes, Z., & Rodrigues, A. T. (2021). Time-trend analysis of medicine sales and shortages during COVID-19 outbreak: Data from community pharmacies. *Research in Social & Administrative Pharmacy: RSAP, 17*(1), 1876–1881.https://doi.org/10.1016/j.sapharm.2020.05.024

Russo, G. I., di Mauro, M., Cocci, A., Cacciamani, G., Cimino, S., Serefoglu, E. C., Albersen, M., Capogrosso, P., Fode, M., Verze, P., & EAU-YAU Men's Health Working Group. (2020). Consulting "Dr Google" for sexual dysfunction: a contemporary worldwide trend analysis. *International Journal of Impotence Research, 32*(4), 455–461.

Sadik Batcha, M., & Ahmad, M. (2021). Publication Trend in an Indian Journal and a Pakistan Journal: A Comparative Analysis using Scientometric Approach. In *arXiv [cs.DL].* arXiv. http://arxiv.org/abs/2102.12914

Smieszek, M., Kindermann, A., Amr, A., Meder, B., & Dieterich, C. (2021). An Apple Watch Dashboard for HiGHmed Heart Insuffcency Patients. *Studies in Health Technology and Informatics, 283*, 146–155.

Toasa, R., Maximiano, M., Reis, C., & Guevara, D. (2018). Data visualization techniques for real-time information—A custom and dynamic dashboard for analyzing surveys' results. *2018 13th Iberian Conference on Information Systems and Technologies (CISTI),* 1–7.

Twells, L. K., Gregory, D. M., Reddigan, J., & Midodzi, W. K. (2014). Current and predicted prevalence of obesity in Canada: a trend analysis. *CMAJ Open, 2*(1), E18–E26.

Velvizhi, V., Billewar, S. R., Londhe, G., Kshirsagar, P., & Kumar, N. (2021). Big data for time series and trend analysis of poly waste management in India. *Materials Today: Proceedings, 37*, 2607–2611.

Wang, C., Hu, J., Blaser, M. J., & Li, H. (2021). Microbial trend analysis for common dynamic trend, group comparison, and classification in longitudinal microbiome study. *BMC Genomics, 22*(1), 1-16.

Wu, D. T. Y., Vennemeyer, S., Brown, K., Revalee, J., Murdock, P., Salomone, S., France, A., Clarke-Myers, K., & Hanke, S. P. (2019). Usability Testing of an Interactive Dashboard for Surgical Quality Improvement in a Large Congenital Heart Center. *Applied Clinical Informatics, 10*(5), 859–869.

Yu, Y.-S., Zou, S., & Whittemore, D. (1993). Non-parametric trend analysis of water quality data of rivers in Kansas. *Journal of Hydrology, 150*(1), 61–80.

Social media and people perception of global warming during critical environmental events: the impact of misinformation through the lens of social noise

NAYANA PAMPAPURA MADALI[a], MANAR ALSAID[b] and SULIMAN HAWAMDEH

University of North Texas, US

nayanapampapuramadali@my.unt.edu[a]
manaralsaid@my.unt.edu[b]

ABSTRACT

Global warming is the term used to describe critical environmental issues and concerns. Social media such as Twitter provides a platform for people to share information, exchange ideas, and express their opinions about current and timely issues. This study utilized contextual analysis to analyze data collected from Twitter for the hashtag "global warming" during the period 2010 & 2011. Using sentiment analysis and topic modeling, the study aimed first at assessing people's perception towards global warming issues, and second study the impact of misinformation from the standpoint of social noise on people's perception of global warming during critical environmental events. The outcome of this study helps create a better understanding of the environmental issues discussed on social media. The sentiment analysis from the data analyzed so far shows that most of the tweets were based on Twitter users' personal opinions and not science. The topic modeling results suggest that Twitter users typically tweeted when a major environmental event occurred due to global warming. Topic modeling also aids in the identification of terms that is associated with social noise. The presence of social noise suggests that misinformation does exist and spreads faster.

AUTHOR KEYWORDS

Global warming; misinformation; social noise; topic modeling; environmentalists.

1. INTRODUCTION

Global warming is the term used to describe the changes in the environment such as increased pollution and deterioration of air quality, melting of the glaciers and rising of seawater, deforestation, and so on. Global warming mainly focuses on increasing temperature and is sometimes used to question the legitimacy of climate change debate when weather events such as the record level of snowfalls in the Eastern U.S. in 2010 happened (Schuldt et al., 2011). In 2004, former vice president of U.S. Al Gore warned from cold weather or extreme weather events as an indication of changes in the climate and the inability to look at the big and connect the dots could be costly in the long term (Drudge, 2004; Lineman et al., 2015).

People tend to search on social media for global warming information (White, 2011). Recent studies have also shown that people's reliance on social media as sources for information on various topics is creating a fertile environment for spreading misinformation (Fischer-Preßler et al., 2019; Gruber et al., 2015). Social media and associated social networks are an essential forum for the people's debate where people express their opinions freely. Also, they have a direct impact on individual attitudes and behavior. On social media, people receive a high number of environmental disasters breaking news. Schäfer (2012) stated that social media's are already an important and credible source of global warming and climate change information. According to Williams et al. (2015), people actively participating in global warming online debates tend to have strong attitudes of either activists or skeptics. In contrast, the people who have a neutral opinion are less likely to participate in this debate (Williams et al., 2015).

Online users sometimes tend to share misinformation without much thinking due to the existence of social noise (Zimmerman, 2020). To understand the misinformation problem, there is a need to understand the factors that lead people to engage in social noise activities. Zimmerman first introduced the concept of social noise in 2020 and defined social noise as the influence of personal and relational elements on information received through social media, which can mislead, distort, or even modify the original message (Zimmerman, 2020). Under the effect of social noise, a social media user may alter their information behavior in an attempt to display oneself in a more appealing manner in order to build their social capital. The concept of social noise helps understand the factors that affect social media users' behavior (Alsaid et al., 2021; Pampapura et al., 2022).

This research examines the impact of misinformation from the point of social noise on people's perception of global warming on social media. Notably, the factors that affect the level of awareness and knowledge of global warming issues. In this research, Twitter data related to hashtag "global warming" is collected over two years period (2010 & 2011) and analyzed.

2. LITERATURE REVIEW

Global warming is a serious environmental issue. Misinformation about environmental issues creates uncertainty and undermines efforts to address the global warming problems. Bjornberg et al. (2017) and Dunlap & McCright (2011), through their studies, identify the various actors and organizations that systematically attempt to discredit global warming and climate change, including scientists, governments as well as political and religious organizations such as think tanks, foundations, and institutes. Another factor in dealing with the environment is the conflict of interest where certain actors in the industry, such as oil, coal, steel, mining, and car companies, work hard to play down the danger of global warming. Media can also be influenced by money and lobbyists, particularly those with political affiliations. Global warming and climate change misinformation is often associated with three concepts which are skepticism, contrarianism, and denial (Boussalis et al., 2018; Pearce et al., 2019).

In recent decades, many studies related to social media and global warming are carried out. Boulianne (2015) examines the relationship between social media use, global warming, climate change activism and engagement. There is convincing evidence that social media platforms affect people's opinions and behaviors related to global warming and climate change. The behaviors toward climate change perception strongly influence peers' attitudes (Kahan et al., 2012). Twitter users have shown different attitudes towards global warming. People with the same like-mindedness typically carry positive sentiments in their tweets. While the users with other skeptics and activists, have a negative sentiment in their messages. One of the studies reported by Arlt et al. (2011) showed that some people used online media to influence global warming and climate change policies. However, studies have indicated that social media did not help change people's lifestyles, such as driving or obtaining energy to mitigate global warming or climate change (Boulianne, 2015; Arlt et al., 2011; Diehl et al., 2021). Most of the time, social media users are polarized about the environment and tend to interact with like-minded users. They are exposed to hoaxes and misinformation (Colleoni et al., 2014; Feldman et al., 2014, Jang & Hart, 2015). US Twitter users are more likely to describe global warming as a hoax and create doubt about the science (Corbett & Durfee, 2004).

The diffusion of misinformation in social media via social noise is a severe problem. Studies suggested that organizations with corporate funding and philanthropic actors have a vested interest in the widespread diffusion of misinformation about climate change (Treen et al., 2020). People in power, such as politicians, and influencers, can amplify misinformation via the "influencers echo chamber." When misinformation is placed in social media, it becomes available to everyone, reinforced through social noise, and echoed through sharing and dissemination (Treen et al., 2020).

The social noise concept aims to understand the social media user's behaviors when engaging in misinformation diffusion unintentionally. Zimmerman's defines social noise as the influence of either personal or relational elements that can either mislead or modify or change the original message or information that is received from social media. Zimmerman's initial notion of social noise comprises four constructs. The four constructs are image curation, relationship management, cultural agency, and conflict involvement. A social media user's intentional or unconscious effort to construct their online identity and create a personal presentation that satisfies them is referred to as image curation. Relationship management is the desire of a user to build ties with people or organizations that have high social value to them. Cultural agency refers to a user's understanding of their roles and responsibilities within social institutions and their public involvement in personal concerns. Conflict involvement refers to a user's level of comfort with conflict.

Pampapura et al. (2022) studied the impact of social noise on Black Lives Matter data on Twitter. The study supported the constructs identified by Zimmerman and added two additional social noise constructs, which are "affiliation and politics'" and "norms and beliefs" (Pampapura et al., 2022).

3. METHODOLOGY

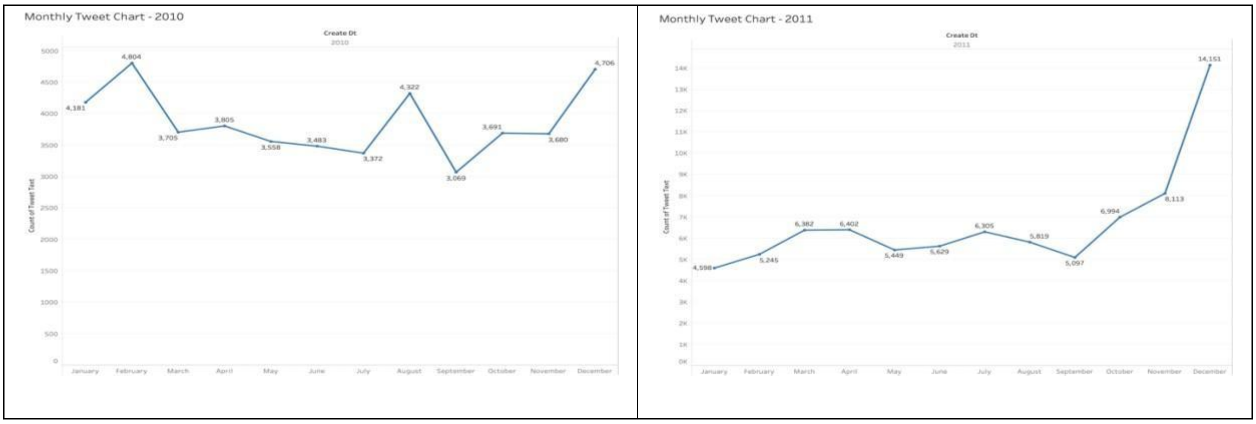
The mixed-method approach used in this study, aimed at understanding issues that impact social media on users' perception of global warming and climate change. A combination of data analytics tools, sentiment analysis and topic modeling, are used to analyze Twitter data. Twitter data for the hashtag “global warming” was collected for this research. This paper reports the results from the data captured during major environmental events in 2010 and 2011.

3.1. Data Collection

Twitter was selected as the data source for two primary reasons. One is that Twitter is one of the most popular social media sites, and Twitter provides a real-time global newswire for individuals and organizations (Fischer-Preßler et al., 2019). Second, Twitter provides Application Programming Interface (API) that makes it easier to acquire data. The Twitter API allowed us to capture the complete archive of Twitter data. In our analysis, we included tweets that were retweeted & quoted. Previous studies emphasize retweets, an adequate measure of how messages are understood as important in the network (Larsson and Moe, 2012).

For 2010, 46,376 tweet records related to global warming were retrieved, whereas, for 2011, 80,184 records were retrieved. The data collected included the tweet creation date, user id, tweet text, tweet language, source of tweet, tweet id, and so on. For easier analysis, both 2010 and 2011 data were stored in two different data files. In this research, only selected data such as the tweet create date and the tweet text data were used for analysis. Figure 1 shows the monthly tweet chart for 2010 & 2011. From figure 1, it can be observed that there is no consistent pattern in the way Twitter users tweeted about global warming for each month.

Figure 1
Monthly global warming tweet chart for the year 2010 and 2011 respectively



3.2. Data Cleaning and Analysis

Exploratory data analysis was performed on the collected data to understand the data better. The tweet text data was then cleaned to remove extra text such as "RT," #," "_," hypertext, and so on. On the cleaned text, two data analysis methods were carried out. First, sentiment analysis was performed to understand the perspective of the Twitter users about global warming as well as the sentiments of global warming tweets. For this research, sentiment analysis was carried out using the Textblob library. TextBlob is a Python package with a simple API to interact with its functions and do basic NLP tasks (Jain, 2018). Textblob sentiment analysis yields two types of results, one is the subjectivity score, and the other is the polarity score. Subjectivity score will help understand whether the text is based on the user's personal opinion (subjective) or if the text is based on the facts (objective). Polarity scores will help understand the sentiment of the text, which can be either positive, negative, or neutral.

Further, to gain in-depth insights about the people's perception and to identify the misinformation thorugh social noise constructs, BERT (Bidirectional Encoder Representations from Transformers) topic modeling method was carried out on the tweet text data. BERTopic is a topic modeling technique that uses transformers and the c-TF- IDF to create dense clusters that allow for easily understandable topics while keeping critical words in the topic descriptions (Bertopic, n.d.). Bert has several advantages, including the fact that it is bidirectional, meaning it can read a text in either direction, which aids the model in learning the context of a word based on its surroundings. In contrast to typical topic modeling techniques, which focus on individual words, Bert uses a minimal vocabulary, and a BERT model considers understanding the "context" of a sentence (LinkedIn, n.d.). On Google Colab, a python code that leverages BERTopic was written and run to extract the topics and their associated keywords/terms for this study.

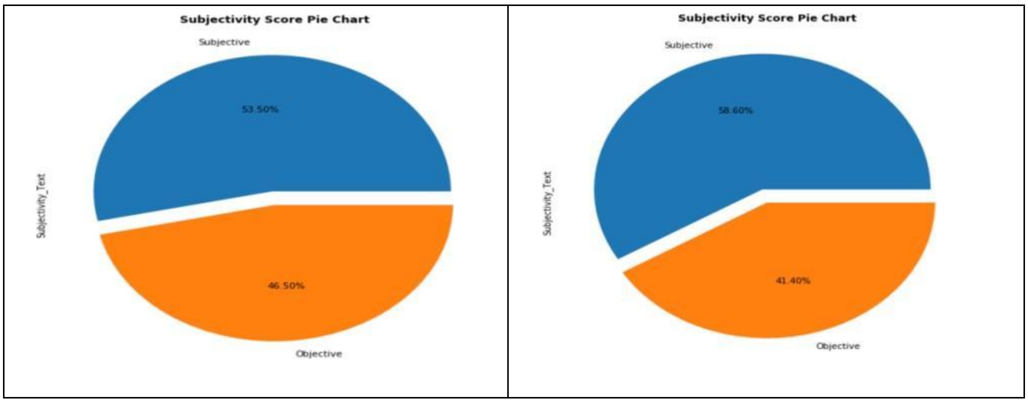
4. FINDINGS

4.1. Sentiment Analysis Results

The text blob package performed sentiment analysis on 2010 and 2011 global warming data. As mentioned earlier, the text blob provides two scores: the subjectivity score and the polarity score. Sentiment analysis subjectivity score of the 2010 global warming data revealed that 53.50 percent of the tweets (24,811 tweets) were subjective or based on the Twitter user's personal view, while only 46.50 percent of the tweets (21,565 tweets) were objective or fact-based. Similarly, a sentiment analysis subjectivity score of 2011 global warming data revealed that 58.60 percent of the tweets (46,986 tweets) were subjective, whereas 41.40 percent (33,198 tweets) were objective. This shows that most tweets in 2010 and 2011 were based on users' personal opinions rather than facts. From the results, it is clear that individuals are concerned about global warming, and the majority of tweets are based on personal opinions and are influenced by personal beliefs. Figure 2 shows the pie chart for the subjectivity scores for 2010 and 2011.

Figure 2

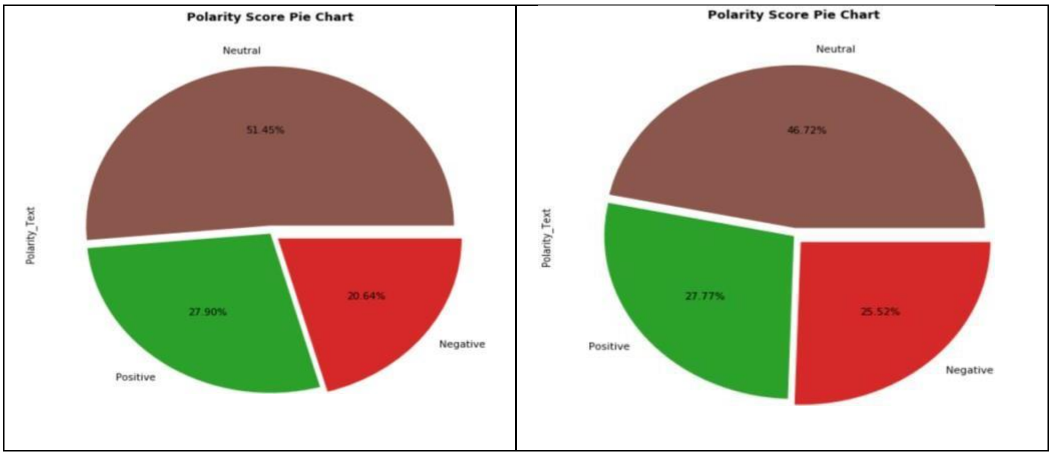
Subjectivity score of tweets for the year 2010 and 2011 respectively



Sentiment analysis also generates polarity scores, which aid in deciphering the sentiment expressed in tweets. According to the examination of polarity scores on 2010 global warming data, 51.45 percent of tweets (23,861 tweets) were neutral, 27.90 percent were positive (12,941 tweets), and 20.64 percent were negative (9,574 tweets). The analysis of polarity scores on 2011 global warming data revealed that 46.72 percent (37,458 tweets) were neutral, 27.77 percent (22,267 tweets) were positive, and 25.52 percent were negative (20,459 tweets). According to the polarity scores, majority of tweets in

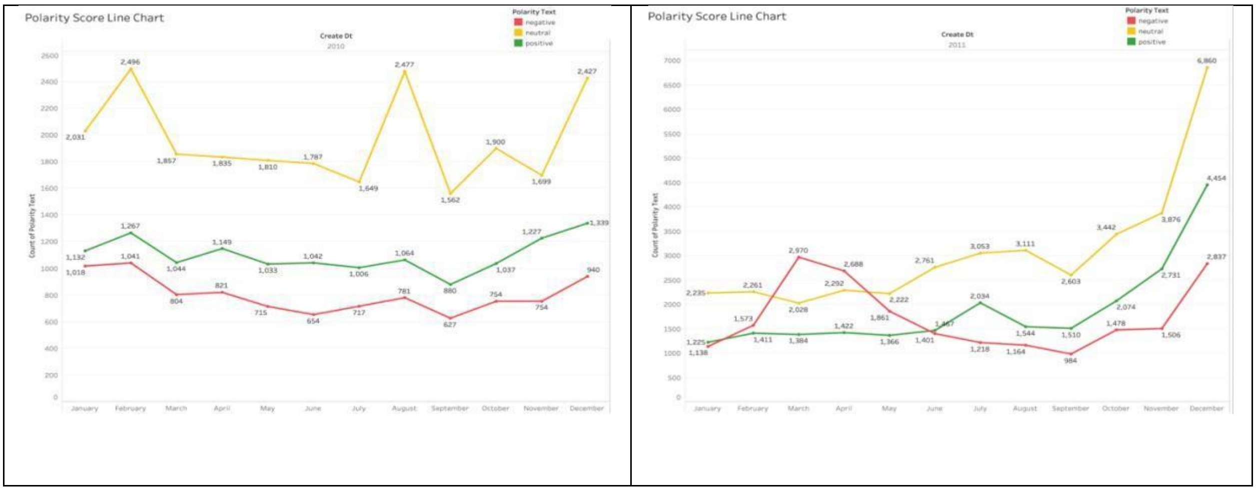
2010 and 2011 were neutral, followed by positive tweets. This demonstrates that most Twitter users have neutral emotions towards global warming. Figure 3 shows the polarity scores as a pie chart.

Figure 3
Polarity score of tweets for the year 2010 and 2011 respectively



The line chart for the polarity score count in 2010 and 2011 is shown in Figure 4. A closer look at the polarity scores reveals that neutral tweets prevailed during the whole year of 2010. Positive tweets followed, with negative tweets remaining minimal over the months. However, in 2011, the trajectory of tweet sentiment was slightly different, with neutral tweets dominating for the first two months of the year before being overtaken by negative tweets in March and April. In May, however, the number of neutral tweets outnumbered the number of positive tweets, and the number of negative tweets remained low for the rest of the year.

Figure 4
Tweet polarity count for the year 2010 and 2011 respectively



4.2. Topic Modeling Results

BERTopic technique was used to understand Twitter users' perception of global warming. Many cities in the mid-Atlantic broke all-time snowfall records in February 2010. When the BERTopic was used to analyze tweets about snow, the topics were compiled, and the corresponding tweet frequency was plotted; from the plot (figure 5a), it is evident that the majority of the tweets about snow were tweeted in February 2010. Similarly, a deepwater horizon oil leak occurred between April 2010 and September 2010. When we plotted the tweet frequency for this issue, it was evident that most of these tweets were tweeted between April and September, when the oil leak was occurring (figure 5b). Figure 5 depicts a line chart showing the frequency of tweets during these events.

Figure 5

Line chart for the two natural disasters that occurred in 2010.

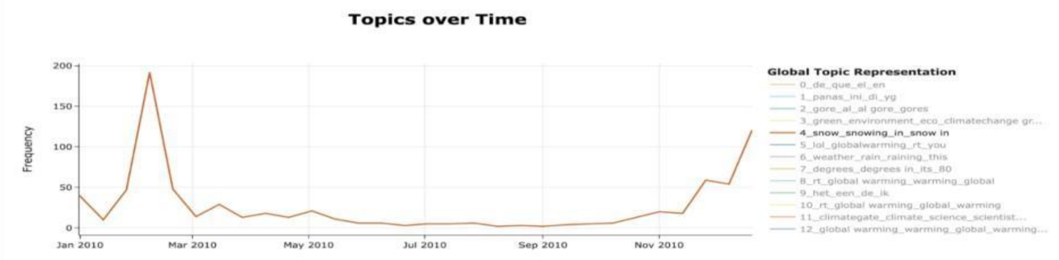


Fig. 5a. Tweet frequency line chart related to Mid-Atlantic Cities Snowfall Crisis in 2010.

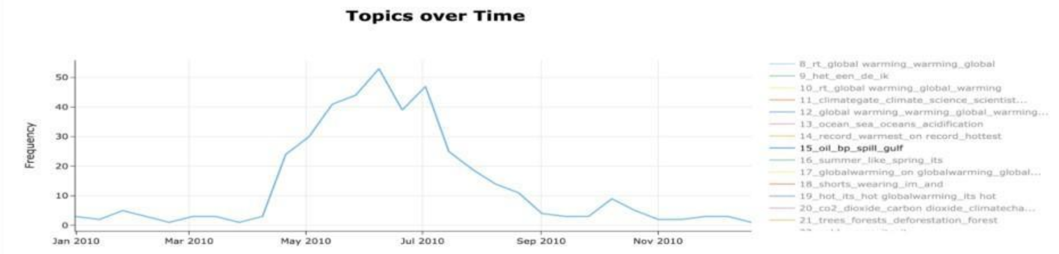


Fig. 5b. Tweet frequency line chart related to Deepwater Horizon oil spill in 2010.

In March of 2011, Japan was hit by the Tohoku earthquake and tsunami. BERTopic developed a topic based on the tweets about the Tohoku earthquake and tsunami, and its tweet frequency was plotted. The line graph revealed that most of these tweets were sent in March and during the time of the natural disaster (figure 6a). In July, many cities experienced record-breaking heat. When the BERTopic was used to analyze tweets about heat, the topics were collated, and the related tweet frequency was plotted, it was clear that in July 2011, the tweet frequency increased fast (figure 6b). When comparing both the line charts (figure 6) connected to natural catastrophes in 2011, it is clear that when a natural disaster occurred due to global warming, Twitter users expressed their dissatisfaction with the situation and started to call for action.

Figure 6

Line chart for the two natural disasters that occurred in 2011

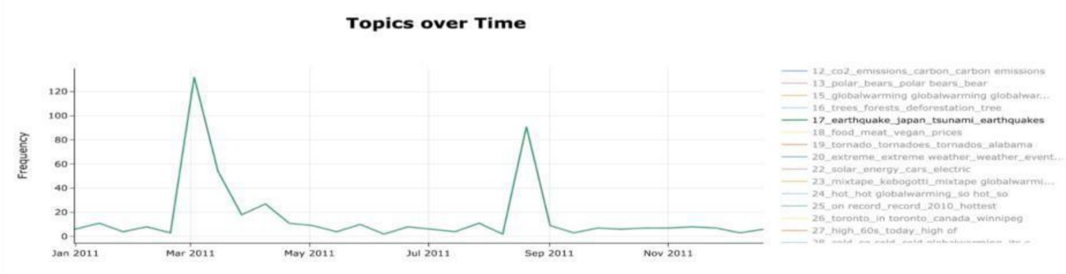


Fig. 6a. Line chart showing the frequency of tweets for the Tohoku earthquake and tsunami event.

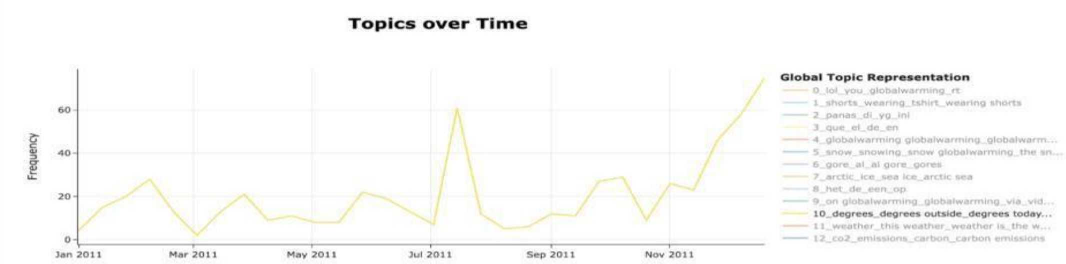


Fig. 6b. Line chart showing the frequency of tweets for the July high temperatures event.

To further understand the general perspective of Twitter users about global warming, line charts were plotted for the topic of global warming for the years 2010 and 2011. From figure 7a and 7b, it is clear that there is no identical tweeting pattern during the months about global warming discussion for both the years 2010 and 2011.

Figure 7

Line chart for the topic global warming

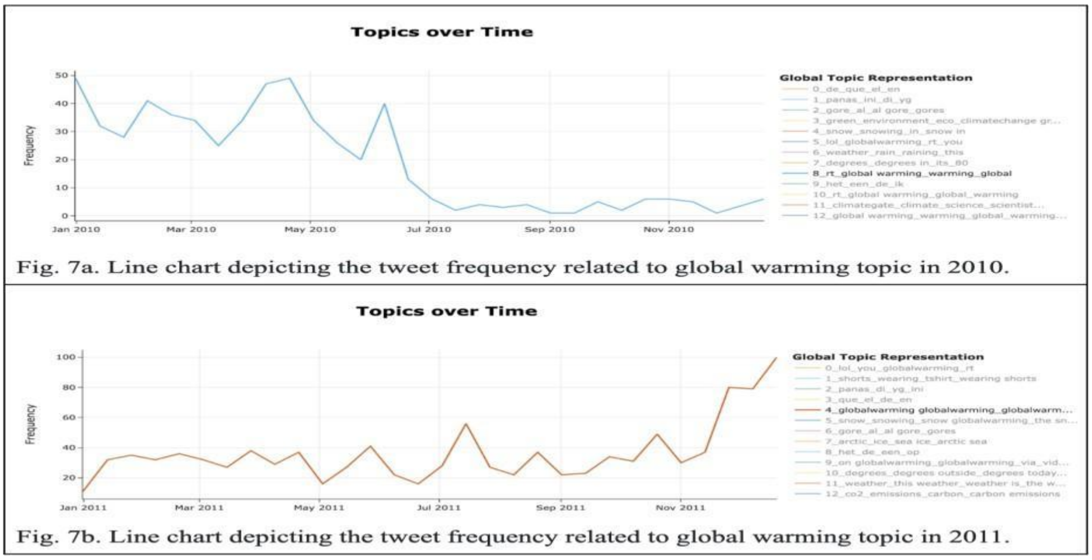


Fig. 7a. Line chart depicting the tweet frequency related to global warming topic in 2010.

Fig. 7b. Line chart depicting the tweet frequency related to global warming topic in 2011.

BERTopic automatically generated 136 topics based on 2010 data, each topic with ten terms connected with it. BERTopic automatically created a total of 262 topics for the 2011 data, each topic with ten terms associated with it. The topics' terms was evaluated for both years to uncover keywords relevant to social noise constructs. The study found terms linked to the six social noise constructs that Pampapura et al. (2022) identified in their publication. There were a lot of terms related to image curation, cultural agency, and conflict participation. This demonstrates that Twitter users used global warming to construct an online image for themselves, voice their opinion and rights, or generate global warming-related conflicts. The topic modeling results further demonstrates that people do participate in social noise which indirectly results in the spread of false information or misinformation. The social noise constructs and their associated phrases are shown in Table 1.

Table 1

Revised social noise constructs and their associated terms from global warming

Construct	Definition	Sample Keywords - BLM	Sample Keywords – Global Warming
Image Curation	Is the effort by a social media user, consciously or unconsciously, to craft their online identities	Please, willing, identity, recommend, report, thank, determination, express	agree, climate experts agree, finds no grounds, global warming drops, global warming scam exposed, help reverse global, hottest year, rising temperature, so hot, too hot, ur own remarkable
Relationship Management	Refers to a user's understanding of their roles and responsibilities within social institutions as well as the level of confidence in their personal beliefs	Love, want, help, please, equivalent, humanity, stoptheviolence, awareness, defense, influence, mutual aid, absurdity, ordinated, advising	Help the environment, can help, can help stop, gonna be, help, help stop, kindle, love, make your own, me again thankfully, need, plz, protect health, protect public health, safety, save, taking steps, to protect health, to reduce
Conflict Engagement	Is the level of social conflict with which a user is comfortable	Report, stop, humiliation, mistreat, misconduct, controversial, discord	blame global warming, blamed, bullshit, change could kill, climate change fraud, climate denial, climate disruption, climategate, disgraced, disruption, fraud, global warming fraud, global warming hoax, global warming scam, hackoff, hoax climate global warming, hot head, may kill, no grounds for, scam exposed, smackoff, undeniable

Cultural Agency	Is characterized by civic participation in social issues and is exhibited by individuals who believe in their own power to be heard and to shape culture and beliefs	radical, recommend, we demand justice, our lives matter, revolution, leadership, diversity, community, social, systemic	brainwashed, can reduce global warming, change near, experts agree humans, fair ambitious, fight global warming, fighting chance, humans are causing, impact, me in urging, participate global warming, power help, reduce global warming emissions, reduce your carbon, save the world, saveworld, stop global warming, to stop global
Affiliation and Politics	Is characterized by loyalty to a political party, religion, or an organization. This could include people paid to carry out certain activities or advertise certain products.	Fox news is racist, system is racism, supremacist, supremacy, racistwhiteleft, fascist, abiding	all white affair, supporting, our community, climate change denial, blame it on, scam
Norms and Beliefs	Is characterized by deep beliefs, culture, ideology, religion, a cause etc.	changeculture, we demand justice, equity, inclusion, fight racism, socialists, communists, anarchists, conservative, extremist	denialism, don't believe, dreaming, myth, regulation, sure about global warming

5. CONCLUSION

It is clear from the data collected that discussion of climate change and global warming peaks around major environmental events such as the snowfall records in February 2010, the deepwater horizon oil leak occurred between April 2010 and Japan Tohoku earthquake and tsunami in March 2011. The sentiment analysis results from the data associated with these environmental events have shown that most people are neutrally concerned about global warming, and the majority of tweets were based on twitter users' personal opinion and were mostly neutral in nature. Topic modeling analysis resulted in the identification of terms related to social noise constructs. Most of the terms identified are associated with image curation, cultural agency, and conflict engagement. This shows that Twitter users tweeted about global warming to either create an online image about themselves, voice their opinion and their rights, or create conflicts related to global warming for cultural agencies or political affiliations. The findings of the study reveal the prevalence of social noise in global warming data, implying the possibility of misinformation. As part of an ongoing research project, the analysis in this paper is limited to data about major events in 2010 and 2011. To further expand this study, more data related to global warming is being collected and analyzed. The results will be reported in an expanded version of this paper.

REFERENCES

Alsaid, M., Madali, N. P., Zimmerman, T., & Hawamdeh, S. M. (2020). COVID19: Mask misinformation and social noise. UNT Digital Library. URl: https://digital.library.unt.edu/ark:/67531/metadc1813454/

Arlt, D., Hoppe, I., & Wolling, J. (2011). Climate change and media usage: Effects on problem awareness and behavioural intentions. International Communication Gazette, 73(1–2), 45–63. https://doi.org/10.1177/1748048510386741

Bertopic. (n.d.). PyPI. Retrieved January 16, 2022, from https://pypi.org/project/bertopic/

Bjornberg, K. E., Karlsson, M., Gilek, M., & Hansson, S. O. (2017). Climate and environmental science denial: A review of the scientific literature published in 1990–2015. Journal of Cleaner Production, 167, 229-241. DOI: https://doi.org/10.1016/j.jclepro.2017.08.066

Boulianne, S. (2015). Social media use and participation: A meta-analysis of current research. Information, communication & society, 18(5), 524-538.

Boussalis, C., Coan, T. G., & Holman, M. R. (2018). Climate change communication from cities in the USA. Climatic Change, 149(2), 173-187. URL: https://link.springer.com/article/10.1007/s10584-018-2223-1

Colleoni, E., Rozza, A., & Arvidsson, A. (2014). Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data. Journal of communication, 64(2), 317-332. DOL: https://doi.org/10.1111/jcom.12084

Cook, B. R., & Overpeck, J. T. (2019). Relationship-building between climate scientists and publics as an alternative to information transfer. Wiley Interdisciplinary Reviews: Climate Change, 10(2), e570.

Corbett, J. B., & Durfee, J. L. (2004). Testing public (un) certainty of science: Media representations of global warming. Science Communication, 26(2), 129-151

Diehl, T., Huber, B., Gil de Zúñiga, H., & Liu, J. (2021). Social media and beliefs about climate change: A cross- national analysis of news use, political ideology, and trust in science. International Journal of Public Opinion Research, 33(2), 197-213.DOI: https://doi.org/10.1093/ijpor/edz040

Drudge, M. (2004). Gore to warn of global warming on New York City's coldest day in decades!.Drudge Report, January 14. Database. accessed: October 5, 2021. URL: http://www.drudgereportarchives.com/ data/2004/01/15/20040115_193004_agwarm.htm

Dunlap, R. E., & McCright, A. M. (2011). Organized climate change denial. The Oxford handbook of climate change and society, 1, 144-160. URL: https://books.google.com/books?hl=en&lr=&id=RsYr_iQUs6QC&oi=fnd&pg=PA144&dq=Organized+cli mate+change+Denial&ots=r7NH9aod7I&sig=b-jJ_DLLAi2xEnAPcuM8telks6k#v=onepage&q=Organized%20climate%20change%20Denial&f=false

Feldman, L., Myers, T. A., Hmielowski, J. D., & Leiserowitz, A. (2014). The mutual reinforcement of media selectivity and effects: Testing the reinforcing spirals framework in the context of global warming. Journal of Communication, 64(4), 590-611. DOI: https://doi.org/10.1111/jcom.12108

Fischer-Preßler, D., Schwemmer, C., & Fischbach, K. (2019). Collective sense-making in times of crisis: Connecting terror management theory with Twitter user reactions to the Berlin terrorist attack. Computers in Human Behavior, 100, 138-151. 12

Gruber, D. A., Smerek, R. E., Thomas-Hunt, M. C., & James, E. H. (2015). The real-time power of Twitter: Crisis management and leadership in an age of social media. Business Horizons, 58(2), 163-172.

Jain, S. (2018, February 11). Natural Language Processing for Beginners: Using TextBlob. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2018/02/natural-language-processing-for-beginners-usingtextblob/#:~:text=A%20big%20advantage%20of%20this,library%20for%20performing%20NLP%20t asks

Jang, S. M., & Hart, P. S. (2015). Polarized frames on "climate change" and "global warming" across countries and states: Evidence from Twitter big data. Global Environmental Change, 32, 11-17.

Kahan, D.M., Peters, E., Wittlin, M., Slovic, P., Ouellette, L.L., Braman, D., Mandel, G. (2012). The polarizing impact of science literacy and numeracy on perceived climate change risks. Nat. Clim. Change 2, 732–735.

Lineman, Maurice, Yuno Do, Ji Yoon Kim, and Gea-Jae Joo.(2015). "Talking about climate change and global warming." PloS one 10, no. e0138996. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0138996

LinkedIn. (n.d.). Retrieved January 20, 2022, from https://www.linkedin.com/pulse/bert-next-generation-topic- detection-sentiment-analysis

Pampapura, N. M., Alsiad, M. and Hawamdeh, S. (2022) The Impact of Social Noise on Social Media and the Original Intended Message: BLM as a Case Study. Journal of Information Science (accepted).

Pearce, W., Niederer, S., Özkula, S. M., & Sánchez Querubín, N. (2019). The social media life of climate change: Platforms, publics, and future imaginaries. Wiley Interdisciplinary Reviews: Climate Change, 10(2), e569.

Schäfer, M. S. (2012). Online communication on climate change and climate politics: a literature review. Wiley Interdisciplinary Reviews: Climate Change, 3(6), 527-543.

Schuldt, J. P., Konrath, S. H., & Schwarz, N. (2011). “Global warming” or “climate change”? Whether the planet is warming depends on question wording. Public opinion quarterly, 75(1), 115-124. DOI: https://doi.org/10.1093/poq/nfq073

Treen, K. M. D. I., Williams, H. T., & O'Neill, S. J. (2020). Online misinformation about climate change. Wiley Interdisciplinary Reviews: Climate Change, 11(5), e665. URL: https://wires.onlinelibrary.wiley.com/doi/full/10.1002/wcc.665

White, C. M. (2011). Social media, crisis communication, and emergency management: Leveraging Web 2.0 technologies. CRC press. URL: https://books.google.com/books?hl=en&lr=&id=bz08DhXc4u0C&oi=fnd&pg=PP1&ots=qUX_bRvCM5& sig=0aEaPiBe3-F7unhMsEMNR6FBUng#v=onepage&q&f=false

Williams, H. T., McMurray, J. R., Kurz, T., & Lambert, F. H. (2015). Network analysis reveals open forums and echo chambers in social media discussions of climate change. Global environmental change, 32, 126-138.

Zimmerman, T (2020) Examining Human Information Behavior on Social Media: Introducing the Concept of Social Noise. (Doctoral dissertation), University of North Texas. Denton, Texas

Panels

Sense-making: Panel of Discovery

Composed by

JOHN TURNER
University of North Texas, USA

SULIMAN HAWAMDEH
University of North Texas, USA

JEFF ALLEN
University of North Texas, USA

DAVE SNOWDEN
Director & Founder – The Cynefin Centre, Cognitive Edge, UK

This panel is made up of seminal researchers who have been identified as critical to the sense-making field of study. This panel is designed to provide an overview of the different methodologies and theories of sense-making from several of the seminal originators of sense-making. This panel will begin by having each panelist identify their own definition of sense-making, what sense-making involves, and why sense-making is relevant in today's globally interconnected environment. This panel will then move to a discussion on how sense-making can be applied in today's environment along with asking recommendations for the next generation of methods and theories for the field of sense-making. Lastly, the audience will have an opportunity to ask questions of the panelists.

Changing Landscape of Scholarly Communications: Open Access

Facilitated by

ANJA ZELTNER
Open Access Brandenburg, University of Applied Sciences, Potsdam, Germany

Composed by

DANIEL GELAW ALEMNEH
University of North Texas, USA

SULIMAN HAWAMDEH
University of North Texas, USA

INA FOURIE
University of Pretoria, South Africa

ABEBE RORISSA
University of Tennessee, USA

ANGELA FORD
Judson University, USA

SHIMELIS ASSEFA
University of Denver, USA

While Open Access (OA) to scholarly and scientific information is not new, various factors, including the shift to emergency remote teaching and national mandates for sharing the products of funded research outputs drive scholars to rethink traditional scholarship models. Today, due to the global COVID-19 pandemic, we are witnessing unprecedented levels of research output generated in a short period of time and shared globally with those who are embarking on finding vaccines and cures in an accelerated fashion. The pandemic indeed, highlighted the importance of transparency, open, and timely access to data and information. Government and private institutions worldwide are reacting to the new situation where researchers, educators, students, and staff are trying to adjust to remote teaching and learning as well as telecommuting.

With the changing landscape of academic publishing and the growing interdisciplinary field of studies, open access facilitates amalgamations of a diverse set of methods, theoretical frameworks, tools and processes that originate from different disciplines and continue to be developed, adopted, and extended through further research, teaching, and practices. In this panel, panelists argue that open access to scholarly knowledge production should be the modus operandi in the time and age we live in. Open access to knowledge is critical not just to accelerate advances in finding solutions to societal issues, but also to meet the growing expectations around higher education institutions' social responsibilities in times of uncertainties. Beyond improving access to scholarly and scientific research for all, OA supports implementation of guiding principles, including the FAIR (findable, accessible, interoperable, and reusable) and FATE (fairness, accountability, transparency, and ethics) principles

Workshop II

Understanding and Co-Creating Sustainable Climate Resilience in Society

Facilitated by

PAULA KUUSIPALO-MÄÄTTÄ
CEO of Paulos Consulting, Finland

OSSI MÄÄTTÄ
Direction Technology Services, BDO, Finland

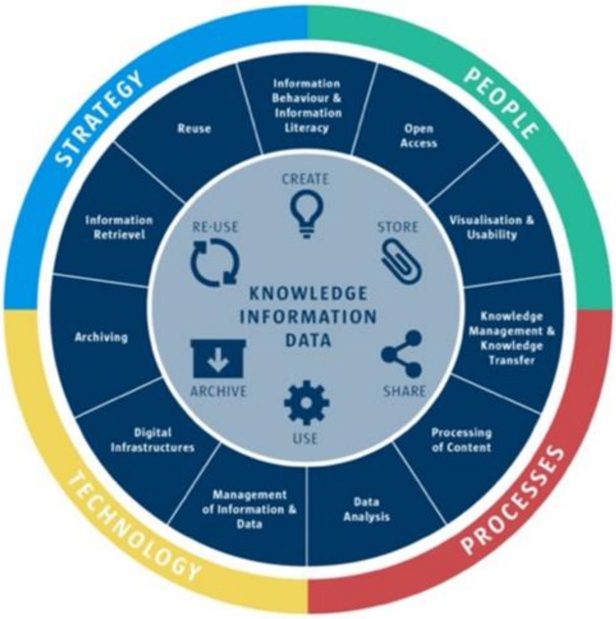
How can Knowledge and Information management professionals contribute to co-creating sustainable climate resilience in society? What are the sustainable climate-resilient futures we want to see take place in a global, interdependent world, and what is the Knowledge and Information management professionals' role in making these scenarios happen?
K&I Management professionals have essential roles to play in addressing the crises, risks, and urgency that our environment is facing. Their work as teachers, researchers, and experts in processing data and knowledge are crucial to how our world will develop in the incoming decades.
In this interactive workshop called knowledge Huddle – developed by GRASP Network – participants are stimulated in a creative way with the use of art to reflect on the above-mentioned questions. Art helps us to imagine alternative futures, and new scenarios, but not only. It also connects to our feelings and emotions and speeds up creative processes.
With this creative approach, the workshop tries to grasp what actions are needed. From individual behavior to organizational, national, and international climate actions, all levels are interwoven and interdependent. In this scene starting from the personal/individual is essential. What can each of us individually do but also how can we collectively make a difference?

The Information Sciences Department, University of Applied Sciences Potsdam

Since its foundation in 1992, the Information Sciences Department (FB5) has made a name for itself among information experts in Germany. It is known for its study concept of a semi-integrative curriculum of archivists, librarians and documentalists - called the "Potsdam Model".

Currently, almost 500 students are taught by 12+ full-professors at the department. In addition to the degree programs, the department also offers adult and part-time education programs for employees in archives and libraries.

The Department hosts the Coordination Center for Archives in the State of Brandenburg and Open Access Brandenburg Current research projects are around the topics of Research Data Management, Digital Literacy and Knowledge Management.



Annex

Prof. Ulrich Schmitt

USB University of Stellenbosch Business School

researchgate.net/profile/Ulrich-Schmitt/research
scholar.google.com/citations?user=CJhwj1oAAAAJ
scopus.com/authid/detail.uri?authorId=7006974138
linkedin.com/in/ulli-schmitt-56842a51

Recent Academic Publishing & Design Science Research Phase

Focus on Actual Real World (incl. fixations & unsustainabilities)	
Digital Scholarship/Curation & Traceability	(JTKS:2015i)
Hierarchy of Needs & Kano Model	(Procedia-ICKM:2016h)
Design Science Research Guidelines	(InformSciJ:2016j)
Experience Management Concepts	(ProWM:2017a)
Webs of Documents and Data	(IEEE-NextComp:2017e)
Digital Threats Assessment	(Sustainability:2018b)
Horizontal Projectability: theoretical-to-theoretical Synergies	
Informing Science Methodologies	(InformSciJ:2015d)
KM Models & Methodologies	(JIKM:2015f)
Memetics	(LNCS/AISC:2016a)
Schools of KM & Knowledge Assets	(ICKCCM:2016d)
Generativity & Fitness-Utility-Models	(Kybernetes:2019e)
Entropy & Generativity Models	(Entropy:2020c)
Vertical Projectability: theoretical-to-empirical/practical Effectiveness	
Network Communities & Social Platforms	(InformSciJ:2017d)
SMEs & Stage-Growth Models	(JEIEE:2018a)
General-Purpose-Technologies & Innovation	(ECKM:2019d)
Sensemaking Loop for Intelligence Analysis Model	(Entropy:2020c)
CK-Theory & Scalable Innovation	(Kybernetes:2020f)
Worldmaking, Scenario Building, and Visioneering	
Twelve Dynamic Knowledge Creation Models	(EJKM:2019c)
Systems Dynamics & Activity-Based Modeling	(IJMO:2020b)
DPE: PKMS-OKMS-LMS Co-evolutions (Gifts...)	(InformingSciJ:2020e)
Visioneering, Vision Quality Criteria	(Sustainability:2021a)
DSR Projectabiliy, Heritage & Domain Evolution	(Sustainability:2021b)
ISO 30401:2018-KMS Standard	(KMRP:2022b in-press)

Knowledge Management System Innovation:

From Nonaka's Conventional SECI/Ba Cycle to SICEE Model...

...from Academic Design Science Research...

LAPLAGE FACTORY
Africa Tech Business Incubator

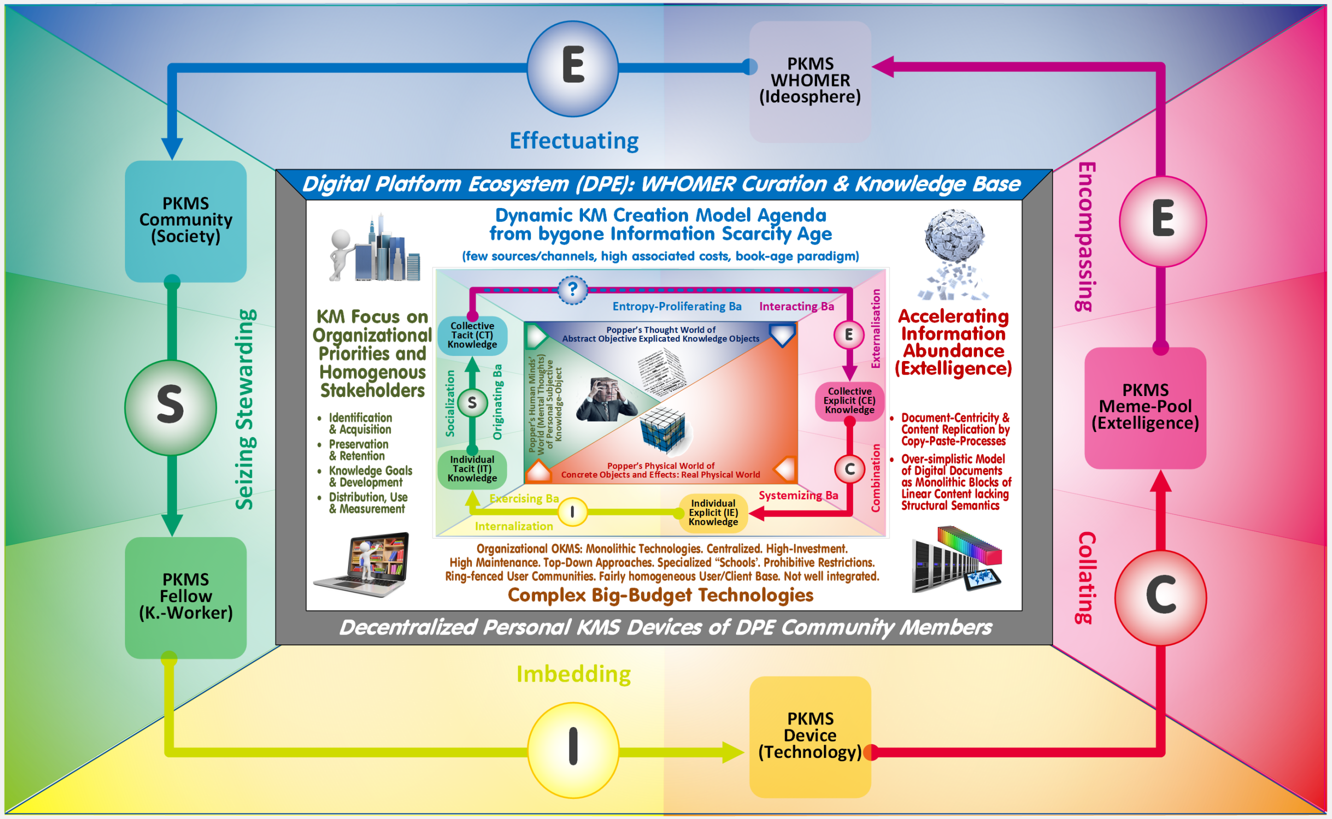
32 St Georges Street
Port-Louis 11324, Mauritius
linkedin.com/company/knowcations

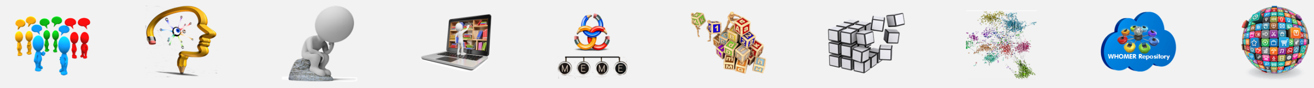
...to 2022 Start-up's Entrepreneurial Phases...

Further Academic Research & Entrepreneurial Start-up Phase

Focus on Start-up Stage	***(Entrepreneurial Toolset)***
Economic Development Board: 5-YR-Business Plan, Investment Application	
Research & Innovation Council: Incubator-Incubatee-Mentor Agreement	
Minimum Viable Product (MVP), Proof of Concept (POC), Lighthouse Project	
Funding/Crowdsourcing Schemes, Strategic LMS/OKMS Partner Agreements	
At the Crossroads: A Platform Ecosystem Roadmap	(InformingSciJ:2022c)
Focus on Incipient Stage	***(Digital Platform Scope Choice)***
Value Appreciation for PKMS Platform Adoption	(in-progress)
Sustainable Development Goals (SDG), Digital Intelligence	(planned)
Non-linear Personal eLearning Environments	(planned)
Contexts of Social Entrepreneurship and Corporate Social Responsibility.	
Focus on Growth Stage	***(Digital Platform Ecosystem Design)***
Appropriation vs. Participation	(in-progress)
Promise and Trust Engineering Methodologies	(planned)
Contexts of Knowledge Workers' Motivations and Absorptive Capacities, Confidence Building that Co-creating Affordances are beneficial and fair.	
Focus on Maturity Stage	***(Digital Platform Ecosystem Dynamics)***
Interdisciplinary Knowledge Organization	(planned)
Decontextualized Boundary Objects Creation	(planned)
Heritage Management and Thought Leadership	(planned)
ISO 56000:2020 Innovation Management Standard	(planned)
Contexts of wider entrepreneurial (Triple Helix) and technological Spaces.	

...to Digital Platform Ecosystem & Community





Stocks & Flows

Shared & Encultured	Seizing or Stewarding	Embrained or Embodied	Imbedding & Immersing	Captured or Transferred	Collating & Codifying	Reposited or Repurposed	Encompassing Edifying & Sharing	Shared Uploaded Enclouded Curatable	Effectuating Easing & Facilitating
Members of DPE Community	Sensemaking & Selecting	Knowledge Worker Community Member	Personalized Extelligence	Decentralized Personal Device	Codification & Meme-Pools	Container & Knowledge Assets	Contexts & Classifications	WHOMER KB & Ideosphere	Enabling Creative Conversations

SICEE 1: Seizing. Tacit or explicit memes representing atomic ideas or content from external sources, desk or field research, and creative conversations via KMS-DPE succeed in competing for an individual user's limited attention span to be located, accessed, and contemplated.

SICEE 2: Imbedding. Memes found useful, might be subjected to collection, comprehension, (re-)composition, capturing and secure storing in an individual's KMS-DPE device as original or mutated versions to facilitate personal sensemaking.

SICEE 3: Collating. Captured memes may be related to other stored memes to form symbiotic relationships to mutually support each other's fitness and to replicate together as memeplexes and knowledge assets for creative work, authorship, citation, classifi-cation, voluntary sharing.

SICEE 4: Encompassing. Memes or knowledge assets and their relationships voluntarily shared are aggregated and curated in a 'World Heritage of Memes Repository' (WHOMER) to eliminate redundancies and consolidate traceabilities for managing entropy and assuring associative integrity.

SICEE 5: Effectuating. WHOMER's curation provides support for creating e-learning assets and citation/reputation metrics to ease collective content access, understanding, retention, re-use, incl. value-added services (e.g., notifications, decontextuali-zed templates/heuristics). *Iteration completed, back to 1.*

Knowcations®

Personal KMS-DPE: Decentralized. Distributed. Affordable. Bottom-Up. Knowledge-Worker-centric Approach focusing on Personalization, Mobility, Generativity, Entropy Reduction. Serving Community with highly diverse Skills (Gifts), Ambitions (Ends), Potential (Means), Settings (Contexts). Integrated.

WHOMER: (World Heritage of Memes Repository – concrete Instantiation of Popper's abstract World:3 as curated, steadily growing (expanding community sharing existent & novel content), single (cloud-based), unified (transdisciplinary), negentropic (redundancy-reducing), tangible, accessible, interrogatable, continuously updated archive of knowledge heritage.

SECI 1: Socializing
For survival, memes may need to be spread by the spoken word from one world:2 host's mind to other world:2 hosts' brains/minds.

SECI 4: Internalizing
For survival, memes may then need to succeed in competing for a living host's world:2 attention span (such as people, teams, corporations, or economies) to be subjectively and tacitly memorized until forgotten or to: see SECI 1 & 2.

SECI 3: Combining
Memes may potentially mutate into new variants or form symbiotic relationships with other memes (memeplexes or knowledge assets) to mutually support each other's fitness and to replicate together. They may be re-edited for fitting particular needs.

SECI 2: Externalizing
Codifying (or encapsulating) memes via objective abstract world:3 objects in inanimate durable world:1 vectors (e.g.; building, machine, products, software, storage devices, books, great art, myths) to spread at times unchanged for millennia.

Blindspot of SECI Model: Entropy-Proliferating Space.

E?

Entropy: Rising stakes of massive duplications of original content (redundancy), partial (fragmentations) or erroneous (inconsistencies) replications or deletions of records, non-disclosure or subsequent erasure of sources (untraceabilities), unsuitable alterations of content (corruptions), lacking curation and maintenance (decay), as well as outdated (obsolescence) and falsified statements (fake facts).

17th International Conference on Knowledge Management 2022

Knowledge, Uncertainty and Risks: From individual to global scale

Potsdam, Germany – June 23–24, 2022

HEALTH DISPARITIES ACROSS NATIVE AMERICAN COMMUNITIES: A KNOWLEDGE SHARING PERSPECTIVE

Motivation: A 2015 National Health Interview Survey in the United States found that Native Americans (NAs) experienced poorer physical health than all other racial groups. Compared to non-Hispanic White adults, NAs have a 50% higher obesity rate and a 58% higher diabetes rate. While existing studies reported NA health as the worst across all ethnicities, the size and economic status of a community impact this statistic. More specifically, if a community is very small with a high proportion of NA population, this NA community is more likely to practice their traditional diets. The assumption of this research is that if they share knowledge on diets within the community, they are likely to have similar health issues. Therefore, the purpose of this study is to investigate two NA communities that differ in size and income levels and explore how their health issues differ.

Literature Review: Literature shows that ethnicity is remarkably correlated to one's health. NA communities living in Arizona have been found to have the poorest health with similar health problems such as obesity and diabetes. These health issues are directly related to diet and the high consumption of flour, sugar, and oil. However, before the 1900s, these NAs supplied and harvested their own food and had few health issues. When they became poor, they depend their food distribution on the government. As such, ethnicity and economic status interact and result in different health outcomes (Rashid et al., 2018).

Data: The research data are patients' electronic health records from a healthcare provider that serves two NA communities in the US with a different proportion of NA concentration and income levels. Community A is economically rapidly growing with a high employment and household income rates. Community A is also less ethnically homogenous compared to Community B. Note that although NA homogeneity is different between the two communities, their ratios in this database are very similar (6.00% for Comm A and 5.45% for Comm B), which makes this dataset ideal for the purpose of this study. As such there is no over-representative sample from either site.

Research Methods: Association rule mining, a data mining technique, is employed to discover knowledge hidden in the data. Association rule mining will discover similar or dissimilar health problems within group.

Results: The data mining technique discovers that NAs in Community B have very high blood pressure and problems associated with obesity. Although the people in Community A have health issues, they are generally healthier and no distinctive pattern are discovered, indicating that the diets among the individuals in Community B are not as similar as those among Community A.

Conclusion and Limitations: The finding shows that NAs in Community B are impoverished and demonstrate a greater rate of homogeneity, and their health problems are consistent with those of other NAs, such as higher rates of co-occurrent blood pressure crisis, obesity, and poor control of A1C problems. On the other hand, NAs in a wealthier community with a lower NA population have some health issues, but they are not as critical as those of Community B. While data mining techniques are powerful to discover knowledge hidden in the massive data and are used for evidence-based recommendation, like other research methods using a secondary data, the finding lacks the qualitive aspect of the analysis. In the future, researchers can visit the sites and interview those patients to fill this gap.

Autor(s)

Health disparities across native American communities: a knowledge sharing perspective (12)

Yong-Mi Kim, University of Oklahoma, United States of America David Steinmann, Access Family Care, United States of America Don McBride, Access Family Care, United States of America

CONSTELADEV: AN INNOVATIVE PLATFORM BASED ON SYSTEMIC METHODOLOGIES TO SUPPORT TEAMS AND PROJECT MANAGEMENT

Simone Perazzoli[1]*, Wagner Sousa Santos[2], José Pedro de Santana Neto[3]
[1]Constela Digital, Videira, BRAZIL
[2]Federation of Industries of the Santa Catarina, Florianópolis, BRAZIL
[3]University of Geneva, Geneva, SWITZERLAND
* Corresponding author: consteladigital@consteladigtal.com

INTRODUCTION

The technological and societal advances of the last decades are demanding the adoption of a new curriculum strategy to foster the development of innovative professionals. Thus, researchers have supported the implementation of more integrated, systemic, and learning-based techniques.

This study presents the project-oriented bootcamp ConstelaDev, designed to promote a high-performance learning and development environment, allowing sharing of ideas and knowledge in practical and active fashioning. Its foundations are in modern and innovative technologies for project development and management (Scrum, Kanban) coupled with artificial intelligence and systemic constellation, which we call the Digital Constellation® framework.

MATERIALS AND METHODS

- The study was carried out with 11 volunteer students from the second and third years of a High School and Technical Institution located in the municipality of Florianópolis, Brazil.
- The participants, belonging to the Technician course in Systems Development, were divided in 3 groups and challenged to design and implement a digital software application from conception to deployment.
- The bootcamp was run in four months and was split into 6 Sprints. Each Sprint consisted of the following sessions: Planning, Delivery, Review, Retrospective, and Checkpoint meetings.
- An intelligent system was developed to monitor and analyze the activities performed, named Digital Constellation® framework (Figure 1).

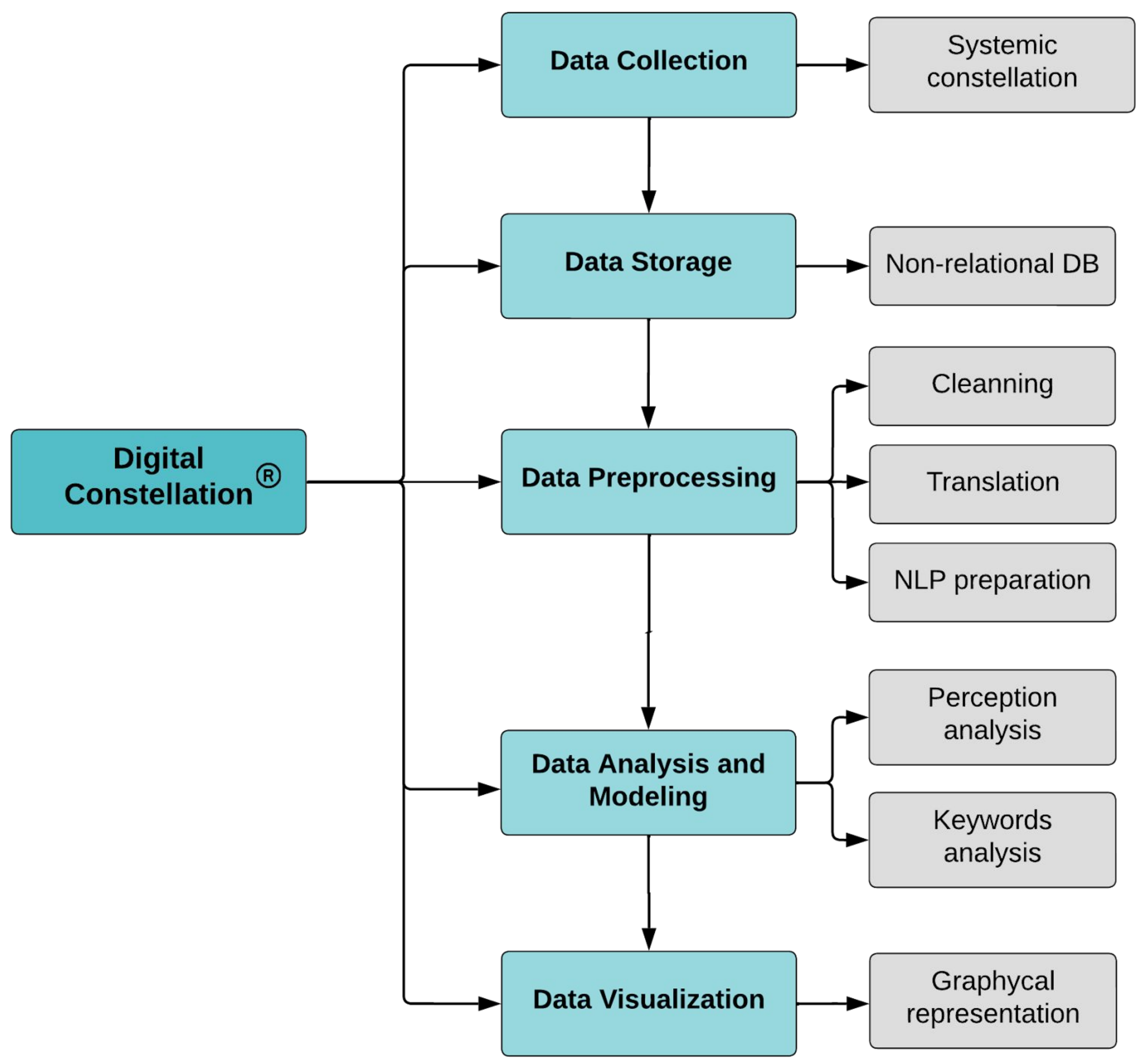


Figure 1 - Diagram of Digital Constellation® framework.

RESULTS AND DISCUSSION

- The perception of participants of their own expectations, experience with the working team, and the project (app) under development is presented in Figure 2.

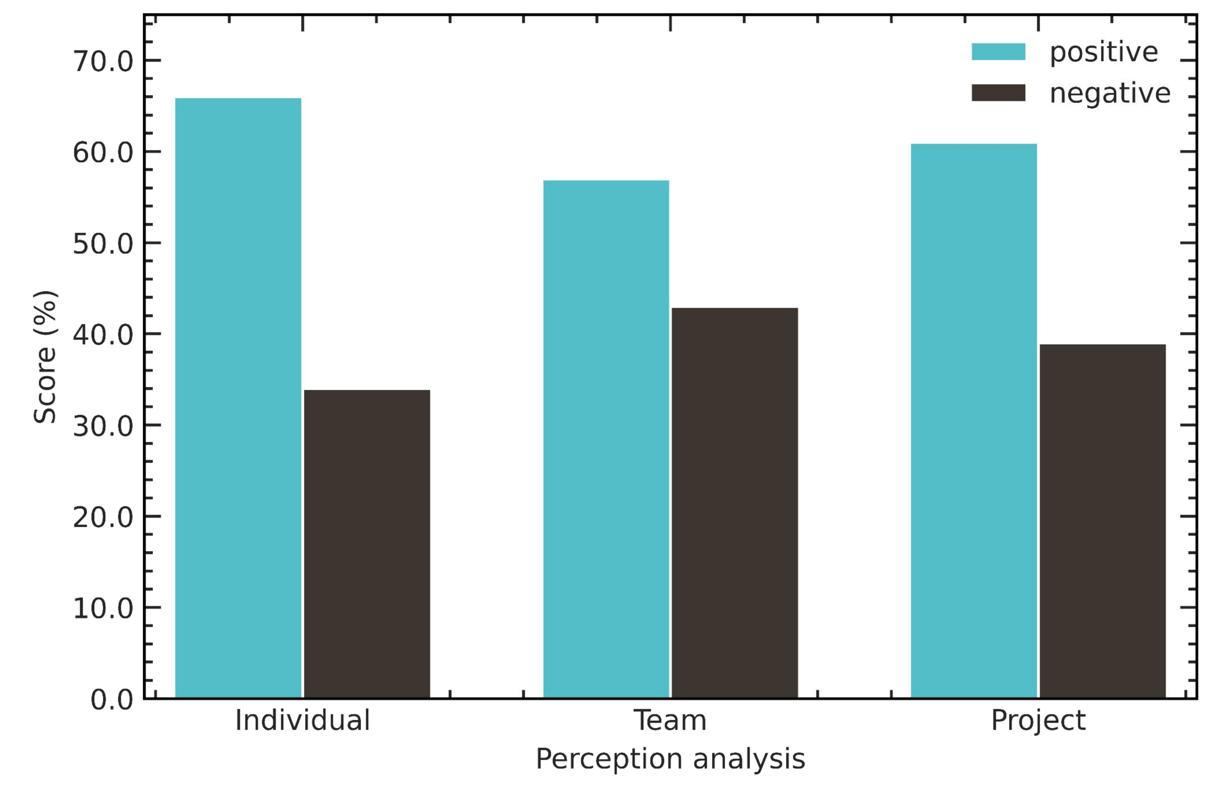


Figure 2 - Perception of participants of their own expectations, experience with the working team, and the project (app) under development, respectively.

- An expressive reduction of anxiety-related terms was observed, with a Pearson's coefficient of 0.85 (Figure 3).
- It can be associated with several factors, such as the gain of confidence of the participants throughout the program, the promotion of a collaborative and psychologically safe environment (Edmondson, 2018), among others.

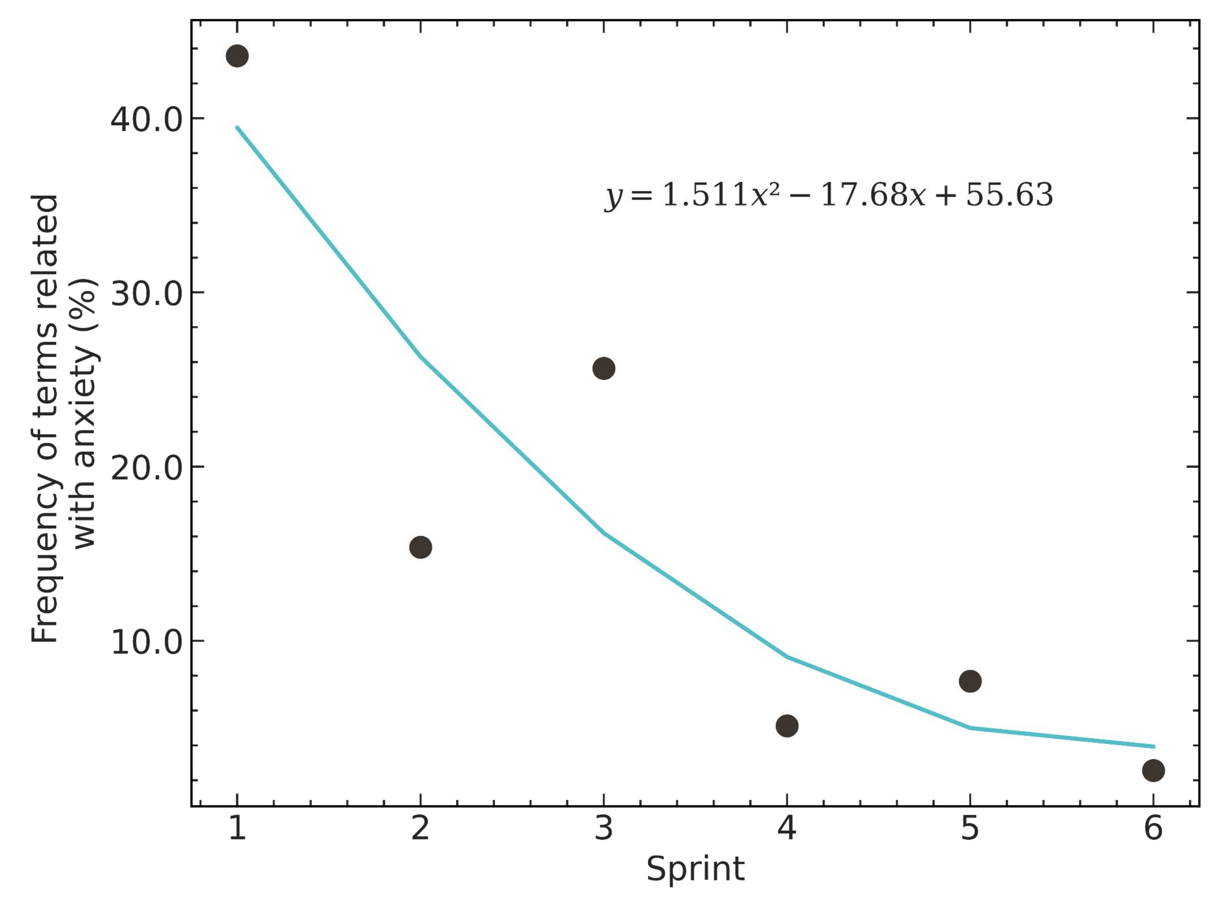


Figure 3 - Frequency of anxiety terms identified along with the program development.

- Participants reported positive feedback, in which the majority of them evaluated their experience as excellent (72.72%) or good (27.27%).
- Practical learning experience, team working, gain of confidence throughout the program, the presence of a collaborative and psychologically safe environment, the exchange of experience and knowledge, and empowerment through the use of technological tools are the key points highlighted.

CONCLUSIONS

- A project-oriented platform that promotes sharing ideas and knowledge in practical and active fashioning driven by an automatic monitoring system was developed.
- The framework bundles a set of capabilities to approach this new emerging reality systematically and sustainably, providing a collaborative and co-responsibility learning environment, fostering dynamism, and cohesion.

ACKNOWLEDGEMENTS

The authors thank the coordinators, teachers, and students of the SESI/SENAI Education Center for their efforts in carrying out this study.

REFERENCES

Edmondson, A. C. (2018). The Fearless Organization: Creating Psychological Safety in the Workplace for Learning, Innovation, and Growth. John Wiley & Sons.

17th International Conference on Knowledge Management 2022

Knowledge, Uncertainty and Risks: From individual to global scale

Potsdam, Germany – June 23–24, 2022

Uncovering disruption of knowledge dissemination in the context of global scholarly communication

An analysis of scientific editorials using a Text Mining approach

Background

Scholarly Communication

- central for distribution of knowledge and research outcomes
- affected by developments in research, scholarly communication and society (e.g. Big Data, Open Access, Covid-19)
- "the foundations of academic publishing are in a state of large-scale disruption" (Barbour 2019, p.151)[1]

Disruption & ANT

- **Disruption:** irritation/disturbance that triggers transformation or adjustment
- **Disruptor:** actor(s) causing disruption
- **Disruptiveness:** potential for/of disruption

Possibilities for disruption in Actor-Network Theory:

- opening the black box
- failed translation processes
- arise of a new actor

Methodological Approach

Sampling

Journal criteria

- Journal Citation Index >1
- 33 categories from SSCI
- combination of 3 search strings

Article criteria

- document type: editorial
- published since 2018

- **database**: Web of Science Social Science Citation Index (SSCI)

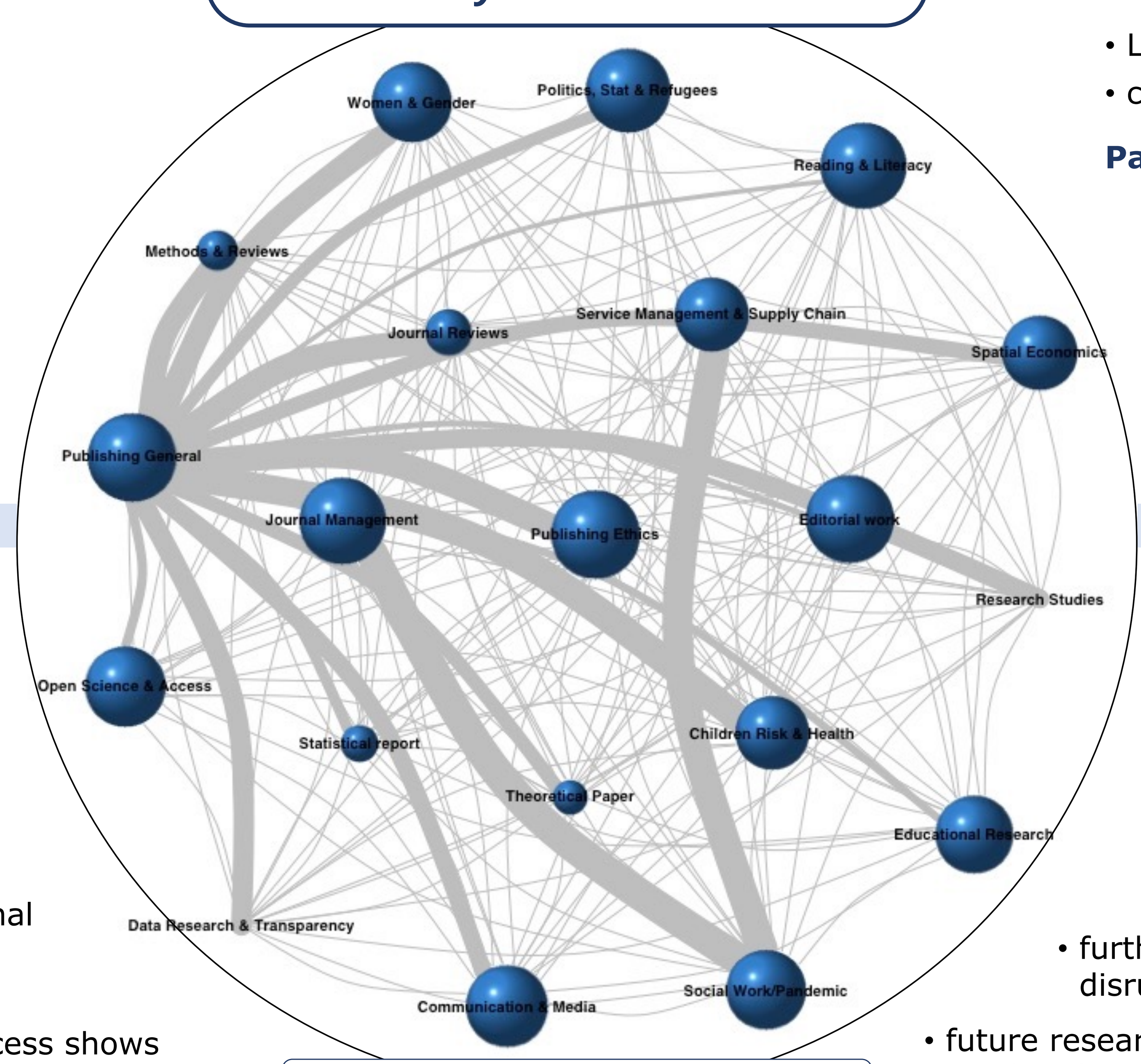
Text Mining

- application via R (quanteda & topicmodels)
- bag of words approach
- LDA topic modeling with k = 20
- computing of topic correlation

Parameter estimation:

- Kullbach-Leibler divergence
- perplexity

Which disruptions can be mapped and identified in scholarly communication?



topic correlation network plot

Results

Networkplot Topic Description Cluster

a) general scientific publication topics

b) discipline-specific topics

c) societal topics in academic discourse

- strong correlation between journal management and pandemic
- surprisingly, Open Science & Access shows relatively few strong links
- publication formats are smaller topics, which show a strong correlation to publishing general

Working Theses

H1: Editors possess shaping power and disruptive potential due to impact → editorials as a communicative medium.

H2: Strong correlation between journal management and pandemic could indicate disruption of the former by the latter.

H3: Discipline specific topics as disruptions in the respective discipline.

H4: Publication formats are mostly smaller topics, but due to the influence of editors expressed through editorials, a change in formats can be assumed.

Discussion

Limitations & Implications

- correlations between each topic due to the statistical detection of them
- disruption potential only statistically elaborated, content analysis in progress
- further elaboration of the identified disruptions necessary
- future research questions: Which disruptors can be identified? Which disruptiveness is present?

Discussion

- based on the editors' shaping power, it can be assumed that all topics have a disruptive character in the meaning of change
- expected disruptions such as climate change or Open Access are not or are less present, which could be due to a normalization over time
- change of publication formats is central to scholarly communication and can partly be explained by changes in research methods due to Big Data (see Lagoze et al. 2015)[2]

1) Barbour, V. (2019). The future of academic publishing: disruption, opportunity and a new ecosystem. *The Medical Journal of Australia, 211*(4), pp. 151-152.e1
2) Lagoze, C., Edwards, P., Sandvig, C., Plantin, J.-C. (2015). *Should I Stay of Should I Go?* Alternative Infrastructures in Scholarly Publishing. *International Journal of Communication, 9,* p. 1052-1071.

TECHNISCHE UNIVERSITÄT DRESDEN

Authors
Anna Cunningham Anna.Cunningham@mailbox.tu-dresden.de
Alina Praun Alina.Praun@tu-dresden.de
Martina Pieperhoff Martina.Pieperhoff@tu-dresden.de

TUDiSC Disruption and Societal Change

Acknowledgement
"Disrupt!Research" is part of the Excellence measure "Disruption and Societal Change" at TU Dresden (TUDiSC) and is funded by the Federal Ministry of Education and Research (BMBF) and the Freestate of Saxony within the Excellence Stategy of the German Federal and State Governments.

The Fifth Law of Library Science: Knowledge Organism
by Patrick Roughen. PhD., JD
International Conference on Knowledge Management
June 23-24, 2022
Potsdam, Germany

School of Library and Information Sciences

School of Library and Information Sciences

Abstract

In 1931 S. R. Ranganathan described his theory of five laws of library science. This research in-progress is an exploration of Ranganathan's fifth law that "the library is a growing organism." Library collection development policies and brands determine the type of information sources to be selected for a collection and organically mediate library growth, and the adaptation and development of associated technologies. Library brands are symbolic representations of the organic identity, collection, and services of a library. This research examines three Library Brands: (1) ImaginOn (Charlotte, NC), (2) Live Oak Public Libraries, (Savannah, Ga,) and (3) the Wild Things a brand theme of the Children's Room at Richland Library, Columbia, SC.

Below is the Logo (part of core brand) of ImaginOn

IMAGINON
THE JOE & JOAN MARTIN CENTER

In 2005 a joint use facility was opened by Charlotte Mecklenberg Library and the Children's Theatre of Charlotte. The core brand above represents the colocation of two educational/entertainment entities which "bring stories to life" A peripheral brand is a sign, symbol, word, or thing representing the values of the institution, such as the Story Jar, (shown in picture to right), an eclectic tree-like structure located in the center of the facility.

"Partying at ImaginOn" Alma Ramos-Mcdermott i CC BY-NC-ND 2.0.
In a 2005 edition of the School Library Journal, Brian Kenney described a group of children using the amazing learning machines at ImaginOn such as the Story Jar pictured above:
"The students are busy at work, using software that guides them through the storywriting process, challenging them to expand their imaginations. When they're done, they can "add" their stories to the Story Jar or, if they have a library card, they can revise them later at home or school."

The Wild Place Courtesy of Richland Library (Columbia, S.C.) The character of Moishe from The Wild Things was used to help thematically organize online information children's resources. A monumental Mural of the Wild Things in a forest is located in the Children's Room, the only public art authorized by Maurice Sendak, author of Where the Wild Things Are.

Southwest Chatham Library Branch of Live Oak Public Libraries (Savannah, Ga,), Courtesy DesignGroup; in this case, "Live Oak" is the core brand, with trees representing the values of this building, built amongst living oaks, and incorporating a life-size sculptural Live Oak. This is the library system's first LEED (Leadership in Energy and Environmental Design) registered building.

Purpose of the Study

The research goal is to explore selected aspects of knowledge transformation. The research uses as reference and as the basis for methodological framework: "A Library is a Growing Organism" (Ranganathan's 5th Law) within the context of library brands on collection behaviors.

Research Methods

By specifically looking at the symbolic branding and development of a collection, the library's knowledge transformation is being explored. Ranganathan's 5th Law can be measured as the brand's growth and prove that "a library is a growing organism."

Data Collection and Analysis

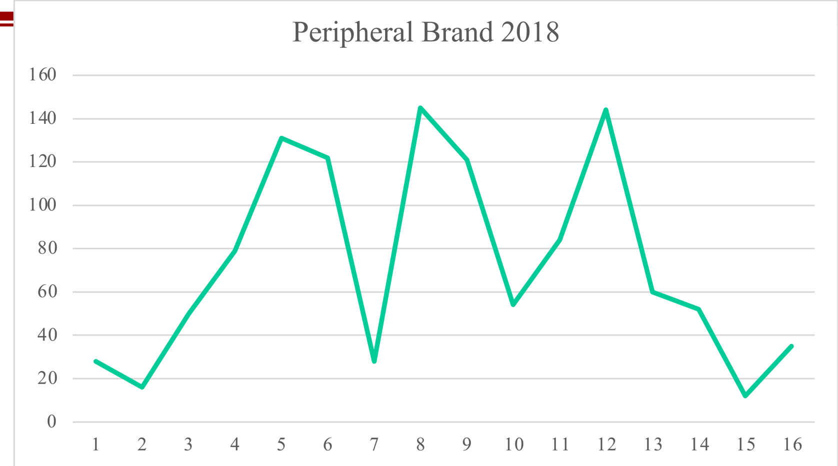
The collection development and branding practices of the *three public libraries* are examined from multiple perspectives. Initial results show that both branding and collection development may mediate library growth, including associated innovation technologies.

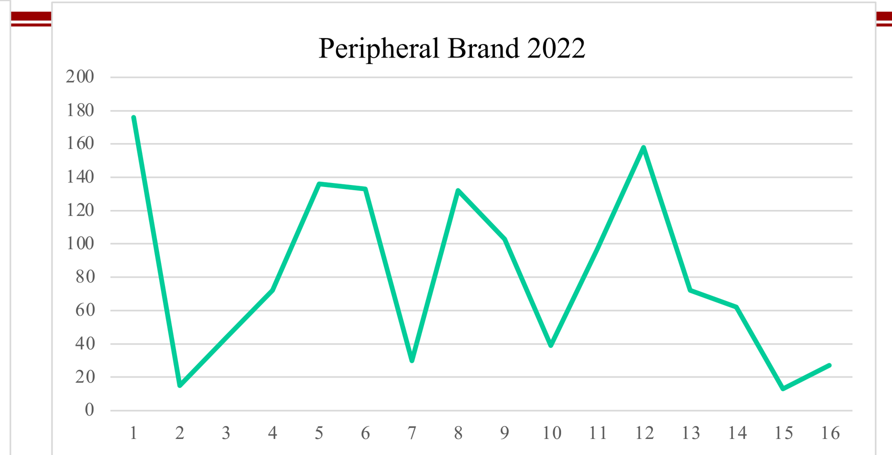
CASE 1: For the Live Oak Public Library, information sources associated with the peripheral brand "tree" grew at a smaller rate than the core brand "Live Oak" and peripheral brands closely related to the core brand, such as "oak". "Live Oak" represents the core brand, which grew by 15%. The following chart represents the growth of relative aspects of the brand:

"Oak"	"Live Oak"	Subject:"oak"	Subject: "tree"	Title: 'oak"	Overall
32.40%	15%	5.40%	3.31%	-9.10%	5.80%

DISCOVERY**:** In some cases, branding encourages collection development; at other times it serves to help organize it; and in some cases, branding moderates a collection and its growth.

Data Graphics or Charts





The above charts show a moderating influence on the peripheral brand "tree" in 16 different library branches from 2018-2022.

CASE 2: Interviews with librarians and library directors showed that when the ImaginOn library and children's theater initially opened, information resources related to the brand "ImaginOn" and its unique function as a hybrid children's library and theater were selected to support and reflect the new brand. This brand identity seemed to decrease over time.

Mecklenberg Library (ImaginOn): QUALITATIVE DATA - Comment from Executive Director at Richland Library (Columbia, SC) and former Youth and Outreach Services Director at ImaginOn (Charlotte, NC):

"I don't recall the strategy at ImaginOn related to collections other than that the focus on folk/fairy tales, poetry and monologues was increased to reflect the joint nature of the facility/partnership.
"Our focus on collections at Richland Library is like everything we do, an extension of our brand promises." .-Melanie Huggins.

Tentative Conclusion

The tentative conclusion that branding mediates and organizes collection development may provide insights and best practices about how a library functions and grows as a learning system. Changes based on the 5th Law can impact how a collection is managed and how it provides useful resources to patrons.

Recommendations and Future Studies

Although it is logical to conclude that collection development policies are the main source of control of the growth of library collections and related technologies, other factors might be considered, such as financial support, personnel policies associated with library collections, and the political environment in which a library functions. Branding is a potentially rich resource for future exploration as it deals with identity and meaning. Branding also appears to have a complex, organic affect on the growth of information resources and associated knowledge learning and innovation technologies, depending on how closely related to the core brand the information technologies and resources sources may be based on quantitative and qualitative data analysis.

References

D. Swain and Roughen, P. (2021). Applying knowledge management to planning joint-use library facilities: The 'ImaginOn,' case study of a library and a theater, *Journal of Knowledge Management*

P. Roughen and Swain, D. (2020) Building Innovative Brands: Cases of Tacit and Explicit Knowledge Conversions in Children's Libraries, *Public Library Quarterly*

Acknowledgements

Dr. Deborah Swain (editor) and Dr. Christopher Cunningham (presenter) – NC Central University

17th International Conference on Knowledge Management 2022

Knowledge, Uncertainty and Risks: From individual to global scale

Potsdam, Germany – June 23–24, 2022

Elicitation of tacit knowledge inside a clinical process using the FRAM

Motivation and Background

Delivering and improving patient safety is a major objective in healthcare since on average 1 in 10 patients experience an adverse event during hospital treatment in high income countries. These adverse events contribute to around 2.6 million deaths every year due to unsafe care (WHO, 2021). Considering Germany, out of 20 million hospitalised patients, 20 000 avoidable death occur annually (Schrappe, 2018).

Health Care - complex organisations

During the past decades, complexity of the health system has evolved significantly impacting patient safety due to:

- limited resources and not ideal work conditions of hospital personnel (Patriarca et al., 2020)
- continuous changes as a result of health care modernisations with constantly growing knowledge and an increased use of modern technology (Goyen & Debatin, 2008; Kälble & Reschauer, 2002)
- an aging population (Statistisches Bundesamt, 2022) with an increase in severe diseases, multiple pre-existing conditions and numerous medical treatments

A great amount of patient safety issues cannot be addressed through simplistic approaches alone anymore in this era (Sujan et al., 2022). Therefore, it is crucial to develop and establish innovative approaches that are suited for complex sociotechnical systems in order to build high-reliability systems in health care. The WHO emphasized in the Global Patient Safety Action Plan 2021-2030 that one characteristic of high reliability organizations is the commitment to resilience (Weick & Sutcliffe, 2007). Resilience can be defined as the ability of a system to adjust its functioning even in the presence of stress. Resilience engineering focuses on the systems capability to constantly cope with the complexity of balancing productivity with safety in daily work (Patriarca et al., 2020).

In complex sociotechnical systems reality often looks different from Work–as-Imagined (WAI) represented by task descriptions, manuals and standard operating procedures. Instead, people constantly adapt to changes in demand, resources and conditions in order to maintain the functioning of the system. This is called Work-as-Done (WAD). This distinction resembles Michael Polanyis (1967) notion of tacit knowledge as something that goes beyond what can be readily expressed by words. The Functional Resonance Analysis Method (FRAM) developed by Hollnagel (2004, 2012) targets eliciting WAD.

FRAM

- Research using the FRAM has developed successfully in scientific terms, has been widely used in several sectors in the industry and has evolved by several researchers who complemented and integrated the original methodological steps proposed by Hollnagel.
- The FRAM remains to be a promising approach in the field of complex dynamic socio-technical systems and has been applied in health for the past 10 years (Patriarca et al., 2020).
- The application of this method has several benefits such as the identification of differences between work as imagined (WAI) and work as done (WAD), the potential to highlight essential system functions and to gain input regarding rethinking clinical processes and procedures that are for instance due to limited resources or not ideal preconditions (Patriarca et al., 2020).

Aim

This qualitative study was conducted in a major regional hospital in Germany, using the FRAM to analyse the recurrent and complex process of surgical care preparation in the morning (Unger et al., 2022). The aim was to offer the hospital a detailed qualitative analysis to promote resilience and maintain a safe state of operating during this process. The FRAM illustrated relevant system functions and unveiled sources as well as effects or variabilities that are now in the process of being evaluated according to appropriate goals and values. In this way, the FRAM illuminates hidden but relevant knowledge, so to speak tacit knowledge, within a complex sociotechnical system that is difficult or impossible to identify otherwise.

Method and Procedure

- For the first time, we applied FRAM to the complex process of surgical care preparation at a major regional hospital in Germany with around 450 beds, 1300 staff members and 75000 patients per year.
- The hospital is an affiliate of the research project GALA within the region of Aachen, Germany (www.gala-regioninnovativ.de). Researcher of the FOM University of Applied Sciences performed the field study.
- The process of surgery preparation starts at 6.30 am and ends at 8.15 am when the surgeons perform their first cut. Figure 1 shows the course of action of patients being prepared for surgery either as an outpatient or as an inpatient departing from the wards.

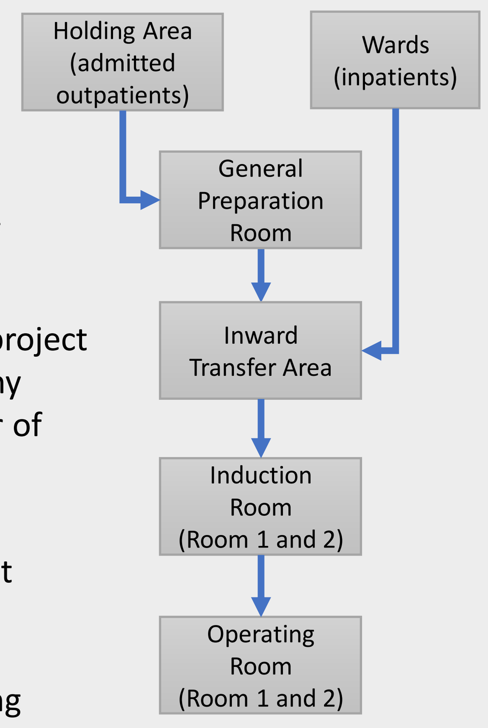


Figure 1. Sequence of the process "Surgical Care Preparation"

- For the collection of qualitative data, researchers conducted semi-structured interviews considering six aspects of each function to accumulate information provided by the medical staff who were involved in this process.
- The duration of each interview was 45-90 minutes, performed by two psychologists using audio recording and transcription. The hexagon in Figure 2 depicts the 6 relevant aspects of each single function or activity.

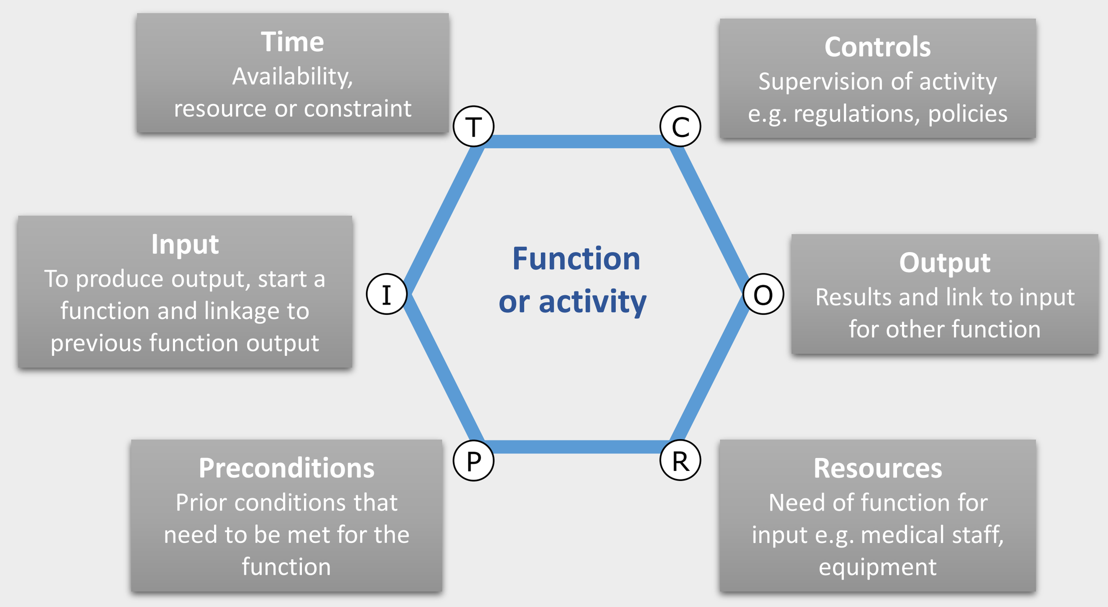


Figure 2. Aspects of Functions within the FRAM (based on Hollnagel, & Goteman, 2004)

- The FRAM views the work processes as a network of functions; these can be either of human, technical or organisational nature. 20 relevant functions, such as for example the function "insert vein access and check patients' identity and confirm that patient did not eat or drink beforehand" were identified and at least two staff members were interviewed for each function. Specific interview questions defined by Hollnagel were used for each function such as e.g., "can you start the function even if the preconditions have not been met?" regarding the aspect "precondition" to identify each aspect of each function.
- Following, the data was inserted into the FRAM-Visualizer Software and analysed.

Results

Figure 3 demonstrates the process of functional resonance and shows that the structured process seen in figure 1 is not visible any longer due to functions with several linkages or couplings to other functions.

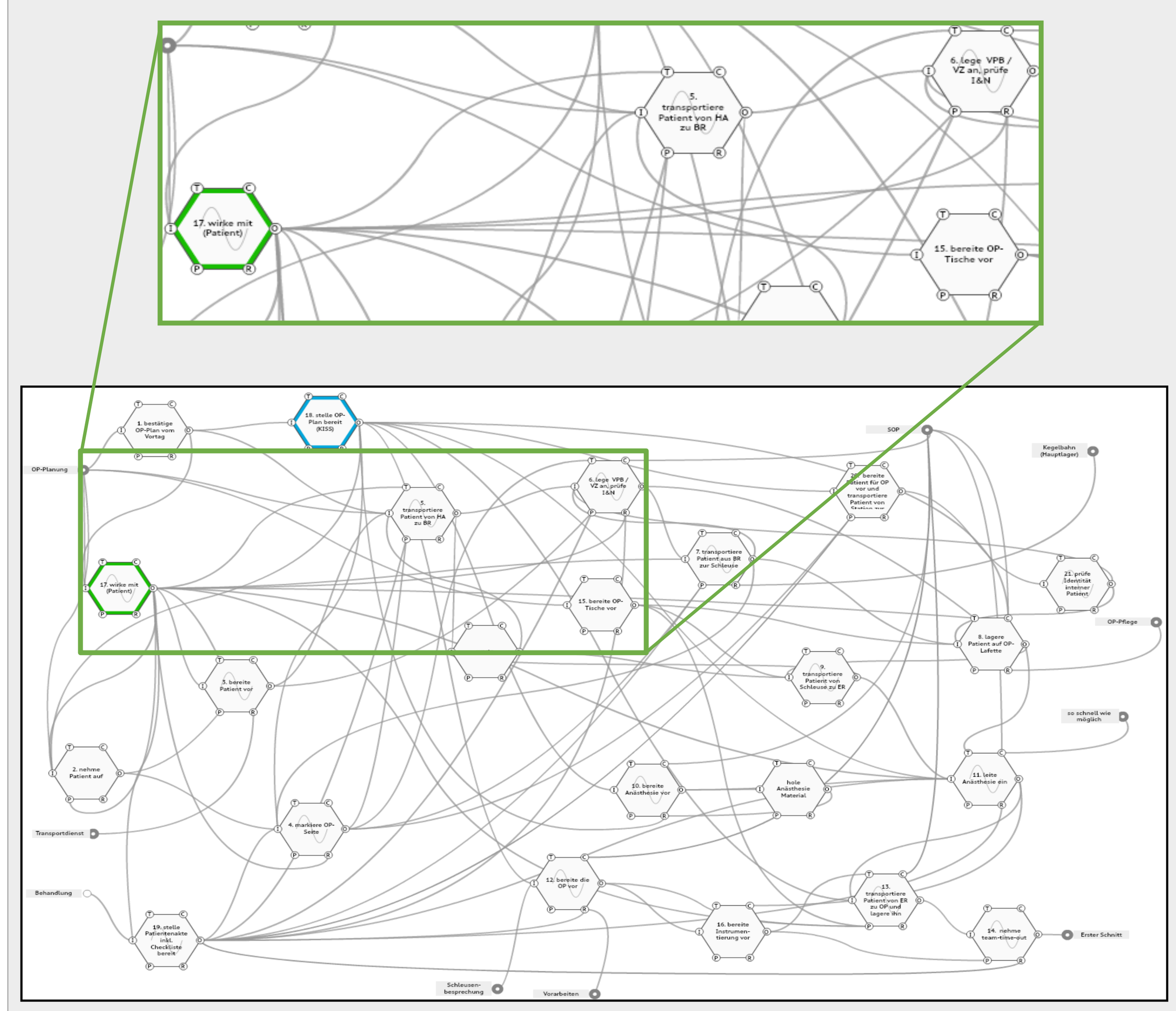


Figure 3. FRAM model of the process "Surgical Care Preparation"

Keeping the aim of redesigning this process in order to increase resilience in this hospital in mind, Hollnagel´s method proposed that looking for functions with multiple couplings can be a first step in establishing whether functional resonance could take place (Hollnagel, 2004).

Therefore, the multidisciplinary team validated the extent of the variabilities through communication and discussed plausible reasons for variances such as human motivation of performing a function.

Positive variances or strengths should be transferred to other processes in the hospital and variances with a high-risk potential were selected to improve clinical processes and procedures.

Considering the concept of resilience, Hollnagel proposes that four aspects are crucial to build high-reliability systems: those are to monitor, respond, anticipate and learn (Hollnagel, 2017). Although the FRAM produces valuable information concerning the potentials to anticipate and to learn as well, its focus is certainly upon the monitor-respond couple as these potentials are fully situated in the here and now.

We identified 14 deviations from the WAI in our study that were further described and evaluated using the WAD-Worksheet we developed as an add-on to the FRAM. The main purpose of this instrument is to systematically check deviations on their validity, origin and consequences. The WAD-Worksheet can be used individually or in teams.

Conclusion

In a world of complex sociotechnical systems new approaches are needed to promote patient safety as well as efficiency of processes. This applies in particular for the health care sector. Resilience Engineering and its associated tools like FRAM offer new insights. From the point of view of knowledge management the FRAM helps considerably in eliciting non-codified work practices, making them explicit and thus open to reflection, evaluation and sharing.

References

Goyen, M., Debatin, J.F. Healthcare costs for new technologies. *Eur J Nucl Med Mol Imaging* 36, 139–143 (2009). https://doi.org/10.1007/s00259-008-0975-y

Hollnagel, E., & Goteman, O. (2004). The functional resonance accident model. *Proceedings of cognitive system engineering in process plant*, 155-161.

Hollnagel, E. (2012). *An Application of the Functional Resonance Analysis Method (FRAM) to Risk Assessment of Organisational Change.* https://inis.iaea.org/collection/NCLCollectionStore/_Public/44/057/44057156.pdf

Hollnagel, E. (2017). *Safety-II in Practice: Developing the Resilience Potentials*. Routledge. https://doi.org/10.4324/9781315201023

Kälble, K. & Reschauer, G. (2002). Wandel der Berufsbilder und Qualifikationsanforderungen in den Gesundheitsberufen. *Public Health Forum*, *10*(1), 2-4. https://doi.org/10.1515/pubhef-2002-1838

Patriarca, R., Di Gravio, G., Woltjer, R., Costantino, F., Praetorius, G., Ferreira, P., & Hollnagel, E. (2020). Framing the FRAM: A literature review on the functional resonance analysis method. *Safety Science*, 129. https://doi.org/10.1016/j.ssci.2020.104827

Polanyi, M. (1967). *The Tacit Dimension*. London : Routledge and Kegan Paul.

Schrappe, M. (2018). *APS-Weißbuch Patientensicherheit* - Sicherheit in der Gesundheitsversorgung: neu denken, gezielt verbessern. Aktionsbündnis Patientensicherheit (Eds.). Medizinisch Wissenschaftliche Verlagsgesellschaft.

Statistisches Bundesamt. (2022). https://www.destatis.de/DE/Themen/Querschnitt/Demografischer-Wandel/Aspekte/demografie-lebenserwartung.html

Sujan, M., Bilbro, N., Ross, A., Earl, L., Ibrahim, M., Bond-Smith, G., Ghaferi, A., Pickup, L., McCulloch, P. (2022). Failure to rescue following emergency surgery: A FRAM analysis of the management of the deteriorating patient. *Applied Ergonomics*, 98. https://doi.org/10.1016/j.apergo.2021.103608

Unger, H., Schröder, S., Mühlbradt, T., Speer, T., & Fastner, C. (2022, March). *Analyse eines komplexen medizinischen Prozesses in einem Krankenhaus der Schwerpunktversorgung mit der Funktionalen Resonanz-Analysemethode (FRAM)*. Paper presented at the: Frühjahrskonferenz der Gesellschaft für Arbeitswissenschaft 2022 "Technologie und Bildung in hybriden Arbeitswelten", Magdeburg, Germany. https://www.researchgate.net/publication/357152369

Weick, K., & Sutcliffe, K. (2015). *Managing the Unexpected*. John Wiley & Sons, Inc.

WHO (2021). Global patient safety action plan 2021–2030: Towards eliminating avoidable harm in health care. Geneva: World Health Organization. https://apps.who.int/iris/rest/bitstreams/1360307/retrieve

Authors: Carmen Aschwer[1], Thomas Mühlbradt[1], Helga Unger[1], Christian Fastner[4], Stefan Schröder[2], Tillmann Speer[3]

[1] FOM University of Applied Science, Dennewartstraße 25-27, 52068 Aachen, Deutschland

[2] Klinik für Anästhesiologie, Intensivmedizin, Notfallmedizin und Schmerztherapie, Krankenhaus Düren gem. GmbH Roonstraße 30, 52351 Düren, Deutschland

[3] Klinik für Anästhesiologie, Klinikum Itzehoe Robert-Koch-Straße 2, 25524 Itzehoe, Deutschland

[4] I. Medizinische Klinik, Universitätsmedizin Mannheim (UMM), Medizinische Fakultät Mannheim, Ruprecht-Karls-Universität Heidelberg, Theodor-Kutzer-Ufer 1-3, 68167 Mannheim, Deutschland

FOM University of Applied Science.

GALA

REGION. innovativ

COMMISSIONED BY

Federal Ministry of Education and Research

This research project is funded by the German Federal Ministry of Education and Research (BMBF) within the Programs "The Future of Value Creation – Research on Production, Services and Work" and "Innovation and Structural Transformation" (funding number 02L20B103) and implemented by the Project Management Agency Karlsruhe (PTKA). The author is responsible for the content of this publication.

ICKM INTERNATIONAL COUNCIL ON KNOWLEDGE MANAGEMENT

Information Sciences Department

FH;P Fachhochschule Potsdam University of Applied Sciences

17th International Conference on Knowledge Management 2022

Knowledge, Uncertainty and Risks: From individual to global scale

Potsdam, Germany – June 23–24, 2022

Knowledge surfacing by Expert created knowledge connections

GuruScan

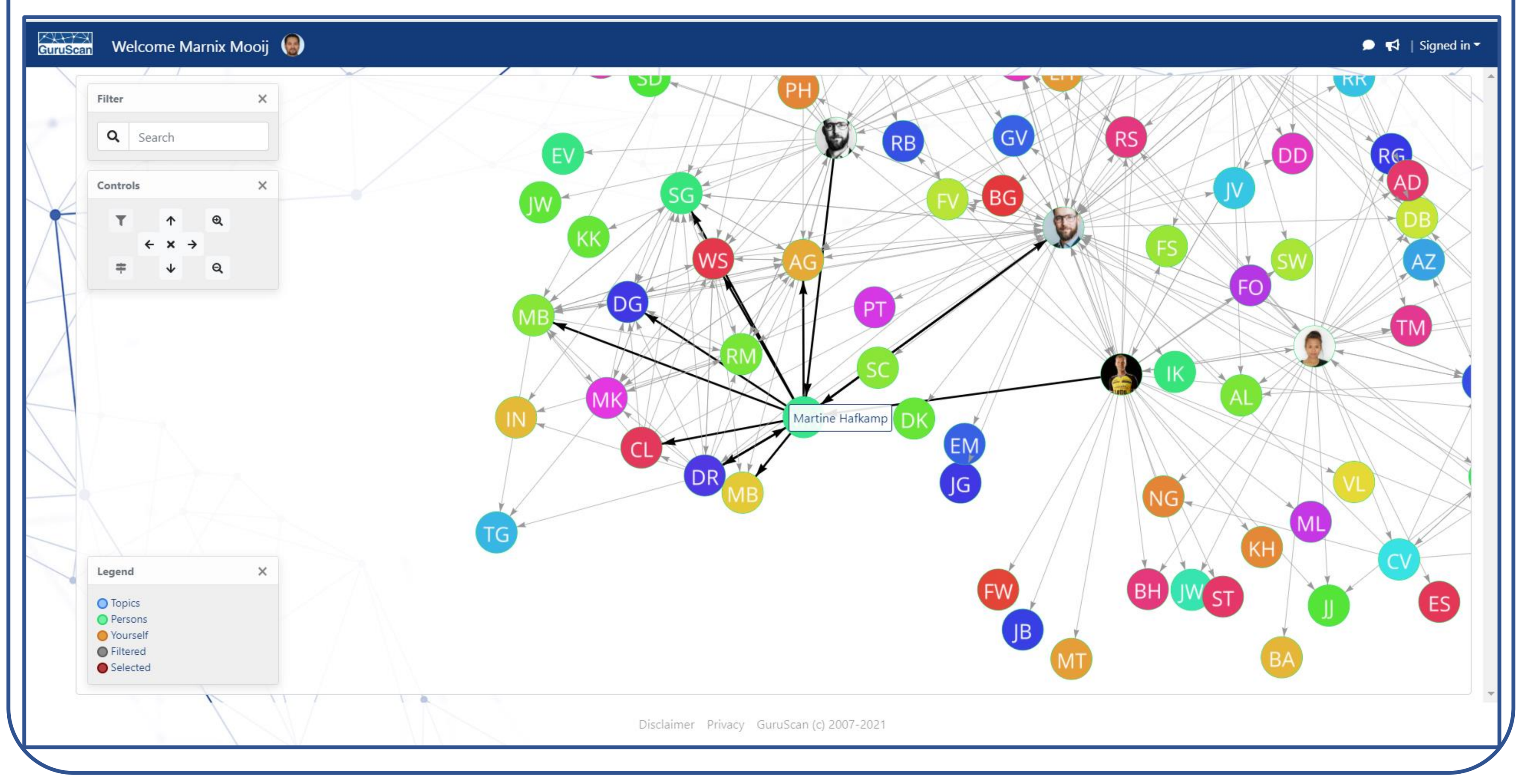
Challenge

Knowledge level self-assessment is biased, mainly by:

- **Competence**
- **Self Perception**
- **Social Comparison**

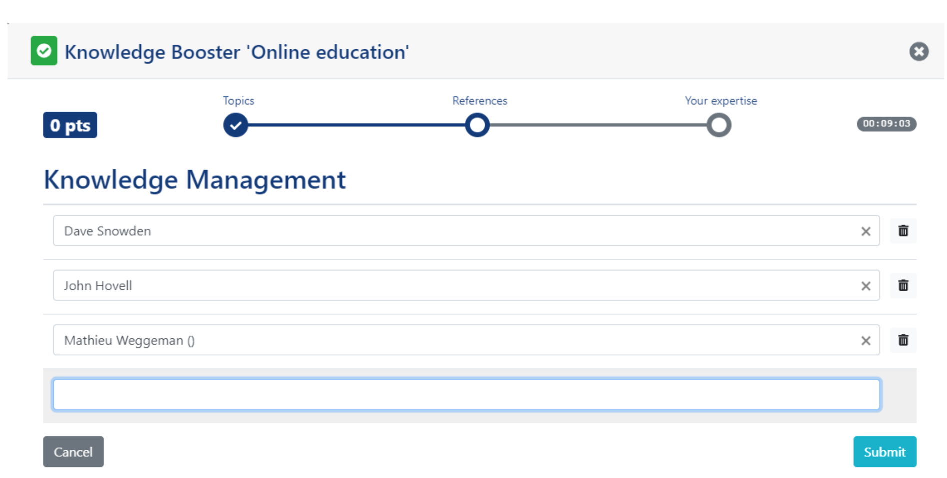
On top of that, non-response is a tough challenge to overcome.

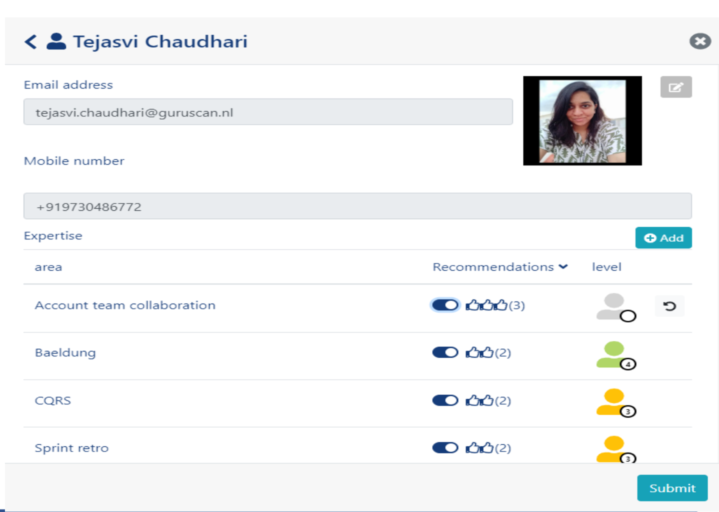
People network



Synchronous and a-synchronous

Synchronous knowledge inventories result in an extremely fast profiling.





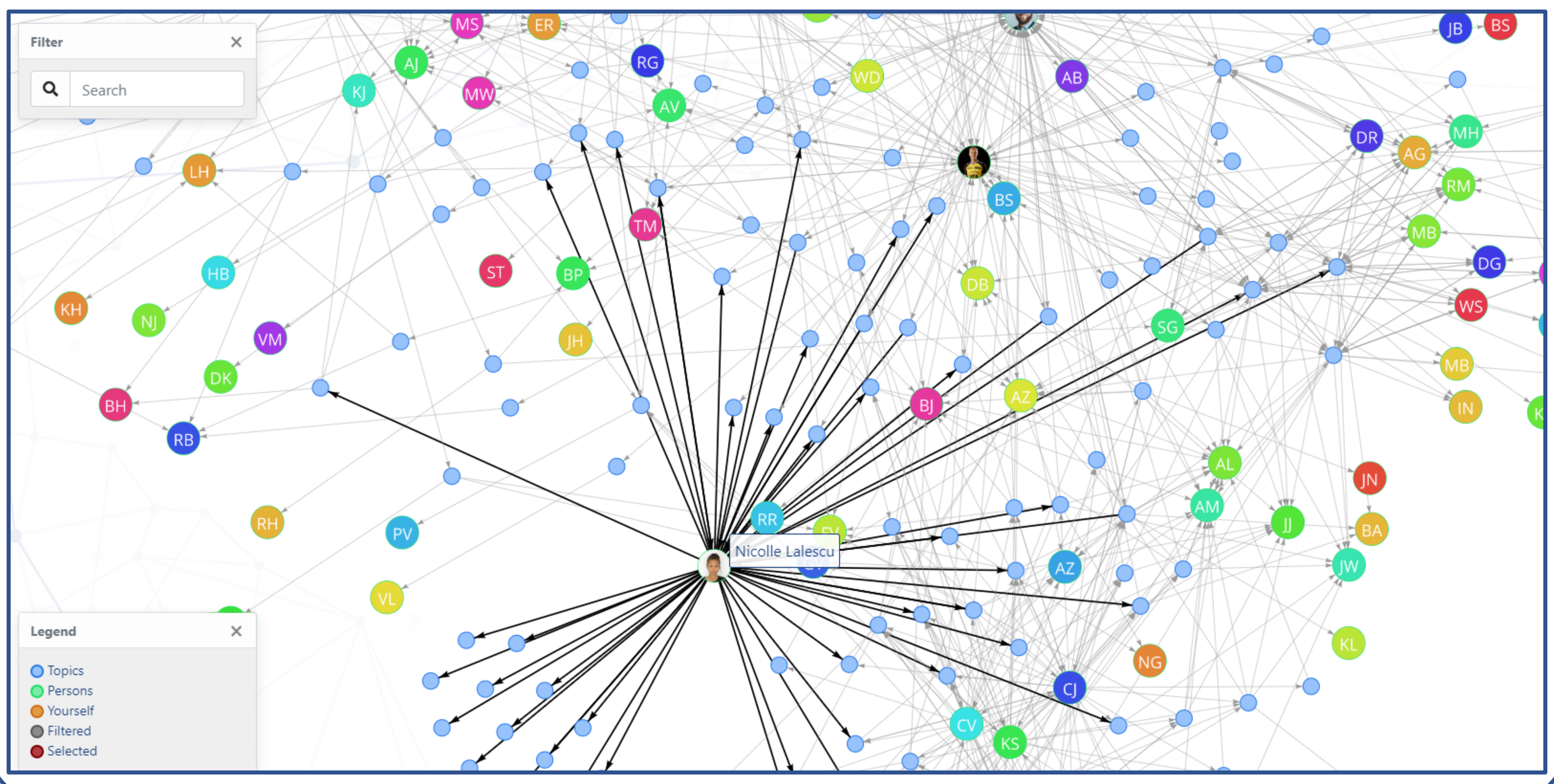
Author

Bart Verheijen MSc.
GuruScan BV (Founder)
bart.verheijen@guruscan.nl
www.guruscan.nl

Knowledge

Knowledge is the most important production factor for any organization.
Knowledge resides in people and is mostly tacit.

Knowledge network



Knowledge connections

Experts in the organization **actively** point out other **experts**. If done collectively, this leads to a subtle but powerful way to overcome the self-assessed knowledge level biases and non-response.

Experience opportunity

Experience the **power** and speed of **Expert created knowledge connections** yourself!
You need:

- **4** or more **people**
  - with laptop and email
- **1** shared **challenge**
- **30 minutes** of your time

We will setup an online GuruScan Knowledge Booster for your experience session.

INTERNATIONAL COUNCIL ON KNOWLEDGE MANAGEMENT

Information Sciences Department

Fachhochschule Potsdam
University of Applied Sciences