AGILE PRACTICES IN DATA SCIENCE AND DATA ANALYTICS PROJECTS: A RESEARCH AGENDA

ABIDEMI ATOLAGBE-OLAOYE

Department of Information Science, University of North Texas,
Denton, Texas, USA
E-mail: abidemiatolagbe-olaoye@my.unt.edu
ORCID: https://orcid.org/0000-0001-7255-1491

Abstract

Goal: The digital age comes with transformational activities (also referred to as digital transformation) triggered by emerging fields and technologies, such as data science and analytics, cybersecurity, cloud computing, blockchain, cryptocurrency, and nanotechnology; helping organizations stay current and competitive. This paper focuses on agile frameworks that support the delivery of data science/analytics projects to ensure organizations rapidly deliver analytics products and services to their customers.
Approach: The study used a systematic literature review, following two stages to collect data from agile-specific literature, then, agile and data science/analytics related literature. The study was conducted using three research questions to understand the agile frameworks used to implement data science/analytics projects, the industries and regions using agile frameworks to implement data science/analytics projects, and the models and frameworks supporting agile practices in data science/analytics projects implementation.
Results: The results show that the popular agile frameworks (Scrum and Kanban) used in software development projects have been experimented on data science projects, mainly in the academic environment among students; however, there is limited research in this area, and there is no evidence that organizations have adopted any particular agile framework or principles for data science projects.
Limitation: There is an absence of empirical data to show the industrial experiences of information professionals in the contexts of this study. Therefore, this study could not analyze any industrial experiences in agile practices in data science/analytics.
Originality/value: There are currently a few studies on agile practices in data science/analytics projects. The literature examined shows no evidence of any consensus or industry-standard agile framework for data science/analytics projects implementation. Hence, this study identifies the gaps and potential research opportunities for information professionals and researchers, such that appropriate agile frameworks and practices that support the rapid delivery of data science/analytics products and services could be established.

Keywords: agile. data analytics. data science. digital transformation. systematic literature review

1. Introduction

Agile originated as an approach to software development in the late 1990s. It was formalized through thought leaders in the software industry with the publication of the agile manifesto in 2001 (PMI, 2017, p. 8). According to PMI, being agile is a mindset embodied in its four values (p. 8), twelve principles (p. 9), and the practices (see the literature review section). The agile approach is suitable when projects “require research and development,” “have high rates of changes,” “have unclear or unknown requirements, uncertainty, or risk,” or “have a final goal that is hard to describe” (PMI, 2017, p. 16). While some organizations have fully adopted these principles for software development projects, others consider how they fit into their business processes.

The digital age comes with many emerging fields and technologies are emerging, such as data science and analytics, cybersecurity, cloud computing, blockchain and cryptocurrency, and nanotechnology, requiring innovative efforts and competencies to transform initiates into valuable outcomes. Prikladnicki et al. (2016) discuss the guiding principles of modern agile as *(1) making people awesome, (2) making safety a prerequisite, (3) experimenting and learning rapidly*, and (4) *delivering value continuously* (p. 21). For organizations to remain competitive in the digital age, they need to embrace digital transformation projects delivered using agile or iterative approaches to increase time-to-market and meet customers’ expectations.

Since the agile approach is suitable for research and development or complex projects and endeavors with some degrees of uncertainties, data science projects fit into this category. It is also essential for information professionals and researchers to understand the concepts of agile practices to support data science/analytics project teams in delivering quality products and services rapidly. The importance of project management is evident in data science competence and competency frameworks such as the EDISON Data Science competence framework and Data to Research CRC's Data Science competency framework. In the EDISON Data Science competence framework, Demchenko et al. (2018) include research methods and project management as part of the five data science competence groups for research-oriented work. Similarly, Data to Decisions Cooperative Research Centres (2017) considers project management as one of the core competencies in their framework. In addition, a recent study shows that employers are interested in employing analytics project managers to manage data science/analytics specific projects (Atolagbe and Floyd, 2020). Therefore, this paper focuses on agile and data science/analytics projects. The research questions that this study aims to address are:

1. Which agile frameworks and principles are used to deliver data science/analytics projects?
2. Which industries and regions use agile frameworks and principles to deliver data science/analytics projects?
3. What models and frameworks (agile and non-agile) support data science/analytics project implementation?

The remainder of the paper is arranged as follows: Section 2, literature review highlighting agile practices and methods used in other fields; Section 3, the methodology used to conduct this study; Section 4 discusses the findings, Section 5 is the discussion that supports the findings, and Section 6 is the summary and conclusion of this study.

2. Literature Review

Agile practice includes activities that revolve around using agile principles to achieve the benefits of being agile (Vallon et al., 2018). The growing complexity around software development and challenges of software engineering (S.E.) birthed a "set principles, methods, practices, and tools to assist the engineering" and implementation of effective systems (Hoda et al., 2018, p. 2). The concept of agile emerged in the late 1990s, offering "disciplined yet lightweight processes while placing human efforts and experience at the core of software development, through its central focus on people and interaction" (Hoda et al., 2018, p. 2). Jovanovic et al. (2020) state that agile adoption depends on the organizational setting, and they (the authors) explore the current agile frameworks to understand the "agile adoption and transition process" (p. 15711).

Curcio et al. (2018) examine the agile concepts used in requirements engineering (RE). The authors defined RE as "a structural set of activities followed to derive, validate, and maintain a system requirements document" (p. 33). Just as project management approaches, RE has the "waterfall life cycle model," also known as traditional RE, and recently, agile RE from agile software development concepts (p. 33). Prikladnicki et al. (2016) discuss the guiding principles of modern agile, a new approach to enterprise agility that enables innovations.

Inayat et al. (2015) focus on how agile concepts resolve the issues found in traditional RE. The authors argue that learning about the traditional RE processes and practices offers an opportunity to understanding and addressing the challenges therein, using agile RE (p. 916). The increasing agile practices in software development necessitates looking at agile practices in RE (Inayat et al., p. 917). In the same light, and with the awareness of agile in many organizations, it is necessary to look at agile practices beyond software development and RE to harness the benefits in other areas. Hence, this study focused on using agile practices and principles in data science/analytics projects.

Other emerging fields and technologies could also explore and apply agile concepts. Al Hashimi and Gravell (2019) describe agile as a philosophy and method used to build and release products. They identify domains such as agile development for big data, agile development for data warehousing, agile development for computer science education, agile development for blockchain, agile user experience design (UX), agile development in cloud computing, and agile development Internet of Things (IoT), and emphasizes the benefits of agile development. However, no specific agile development frameworks or principles were recommended.

3. Study methodology

Research in agile concepts has increased "since the agile manifesto in 2001," with most publications in conference proceedings, workshops, and journals (Curcio et al., 2018). This study was approached in two stages of article selection. Firstly, recent literature focusing on agile practices and principles were selected, and a tertiary study on previous systematic literature review and systematic mapping studies was conducted. The articles were searched and selected from top-rated information science and information systems journals published recently (in the last five years). The goal of exploring journals from the agile domain was to investigate the origin of agile practice. Eventually, the relevant journals (six) that met these requirements were selected. The Agile Practice Guide, published by PMI, was also used. Table 1a below shows the selected articles.

Table 1a. List of literature - Stage 1.

Code	Title	Author(s)	Year	Paper type	Keywords	Domain
A1	A systematic literature review on agile requirements engineering practices and challenges	Inayat, I., Salim, S., Marczak, S., Daneva, M., & Shamshirband, S.	2015	Journal	Agile software development methods, Agile requirements engineering, Collaboration, Traditional requirements engineering, Systematic review	Agile
A2	Trends in agile: Perspectives from the practitioners.	Prikladnicki, R., Lassenius, C., Tian, E., & Carver, J.	2016	Magazine / Journal	N/A	Agile
A3	Agile Practice Guide	Project Management Institute (PMI).	2017	Professional Guide	N/A	Agile
A4	Requirements engineering: A systematic mapping study in agile software development	Curcio, K., Navarro, T., Malucelli, A., & Reinehr, S.	2018	Journal	Agile software development, Requirements engineering, Systematic mapping study	Agile
A5	The rise and evolution of agile software development	Hoda, R., Salleh, N., & Grundy, J.	2018	Magazine / Journal	N/A	Agile
A6	Systematic literature review on agile practices in global software development.	Vallon, R., da Silva Estácio, Bernardo José, Prikladnicki, R., & Grechenig, T.	2018	Journal	Global software development, Global software engineering, Distributed software development, Agile software development, Agile practices, Scrum, Extreme programming, XP, Systematic literature review	Agile
A7	Agile transition and adoption frameworks, issues and factors: A systematic mapping.	Jovanovic, M., Mesquida, A., Mas, A., & Colomo-Palacios, R.	2020	Journal	Agile software development, agile transition and adoption, systematic mapping.	Agile

Secondly, a systematic review of Agile and Data Science/Analytics literature was conducted. Cerdeiral and Santos (2018) describe a systematic literature review as “a means of identifying, evaluating, and interpreting the available research related to a research question, topic area, or phenomenon” (p. 57), to understand research trends. Cerdeiral and Santos and many other researchers have followed the guidelines by Kitchenham and Charters (2007) and Petersen et al. (2015) to search, select, evaluate, and interpret relevant literature to address research questions. This study used the “Publish or Perish” software via the “Google Search” option to retrieve relevant literature. A total of 364 articles were retrieved with the search strings “Agile,” “Data science,” “Data analytics,” and “Systematic literature review.” An initial screening of these articles reduced the relevant articles to fifty-six (56). After reviewing the abstract and introduction sections, the final selection of sixteen (16) articles was made for this study. The articles are mainly from information systems journals (10) and conferences (6). Table 1b shows the list of the selected literature.

Table 1b. List of literature - Stage 2.

Code	Title	Author(s)	Year	Paper type	Keywords	Domain
A8	Context-Aware Enterprise Modelling towards Agile Models Development	Bilinkis, J., Zueva, A., & Zaytseva, E.	2017	Conference	Agility, BPM, Context-awareness, Process modelling	Agile
A9	A Critical Review of the Use of Spikes in Agile Software Development	Al Hashimi, H., & Gravell, A. M.	2019	Conference	Agile, Spikes, Risk management, Uncertainty.	Agile
B1	Editorial—Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research	Agarwal, R., & Dhar, V.	2014	Journal	N/A	Data Science/ Data Analytics
B2	Unicorn data scientist: the rarest of breeds	Baškarada, S., & Koronios, A.	2016	Journal	Data analytics, Skills, Definition, Framework, Data science, Business analytics	Data Science/ Data Analytics
B3	Systematic Literature Review of Big Data Analytics	Eachempati, P., & Srivastava, P. R.	2017	Conference	N/A	Data Science/ Data Analytics
B4	Big data and business analytics ecosystems: paving the way towards digital transformation and sustainable societies	Pappas, I. O., Mikalef, P., Giannakos, M. N., & Krogstie, J.	2018	Journal	Analytics, Big data, Digital ecosystems, Digital transformation, Sustainable societies	Data Science/ Data Analytics
B5	'Big time': An examination of temporal complexity and business value in analytics	Conboy, K., Dennehy, & D., O'Connor, M.	2018	Journal	Temporality, Business analytics, Business value, Information systems development	Data Science/ Data Analytics
B6	Big data analytics capability and co-innovation: An empirical study	Lozada, N., Arias-Pérez, J., & Perdomo-Charry, G.	2019	Journal	Business, Economics, Information science, Big data analytics capabilities, Co-innovation, Big data, Co-creation	Data Science/ Data Analytics

Table 1b. (continued).

Code	Title	Author(s)	Year	Paper type	Keywords	Domain
C1	Big data team process methodologies: A literature review and the identification of key factors for a project's success	Saltz, J. S., & Shamshurin, I.	2016	Conference	Analytics Process, Big Data, Data Science, Project Management, Process Methodology	Agile & Data Science/Analytics
C2	Comparing Data Science Project Management Methodologies via a Controlled Experiment	Saltz, J. S., Shamshurin, I., & Crowston, K.	2017	Conference	N/A	Agile & Data Science/Analytics
C3	Exploring How Different Project Management Methodologies Impact Data Science Students	Saltz, J. S., Heckman, R., & Shamshurin, I.	2017	Conference	Data Science Education, Big Data Education, Project Management, Agile Development	Agile & Data Science/Analytics
C4	Data science as an innovation challenge: from big data to value proposition	Kayser, V., Nehrke, B, & Zubovic, D.	2018	Journal	N/A	Agile & Data Science/Analytics
C5	Agile-DAS to Use a Data Science Process Methodology	Saltz, J. S.	2018	Conference	Data Science, Big Data, Project Management.	Agile & Data Science/Analytics
C6	Achieving Agile Big Data Science: The Evolution of a Team's Agile Process Methodology	Saltz, J. S., & Shamshurin, I.	2019	Conference	Big Data Science, Agile, Process Methodology.	Agile & Data Science/Analytics
C7	Application of Methodologies and Process Models in Big Data Projects	Quelal, R., Mendoza, L. E., & Villavicencio, M.	2019	Conference	Agile Methodologies, Big Data, Systematic Literature Review, Text Mining	Agile & Data Science/Analytics
C8	Managing Big Data Analytics Projects: The Challenges of Realizing Value	Jensen, M. H., Nielsen, P. A., & Persson, J. S.	2019	Conference	Big data analytics, Benefits realization management, Value realization, Case study	Agile & Data Science/Analytics

4. Findings

4.1. Agile Mindset and Frameworks

Many organizations see the adoption of agile practice as a strategic move to rapidly create and release products that meet customers' specifications. PMI Agile Practice guide indicates two factors that can help organizations determine their project management approach. The factors are requirement uncertainty and technical degree of uncertainty. Figure 1 shows four classes of projects: (1) simple, (2) complicated, (3) complex, and (4) chaotic (PMI, 2017, p. 14). Among these classes, complicated and complex projects are the ideal candidates for agile. Another way to determine candidates for agile projects is the life cycle characteristics shown in Figure 2.

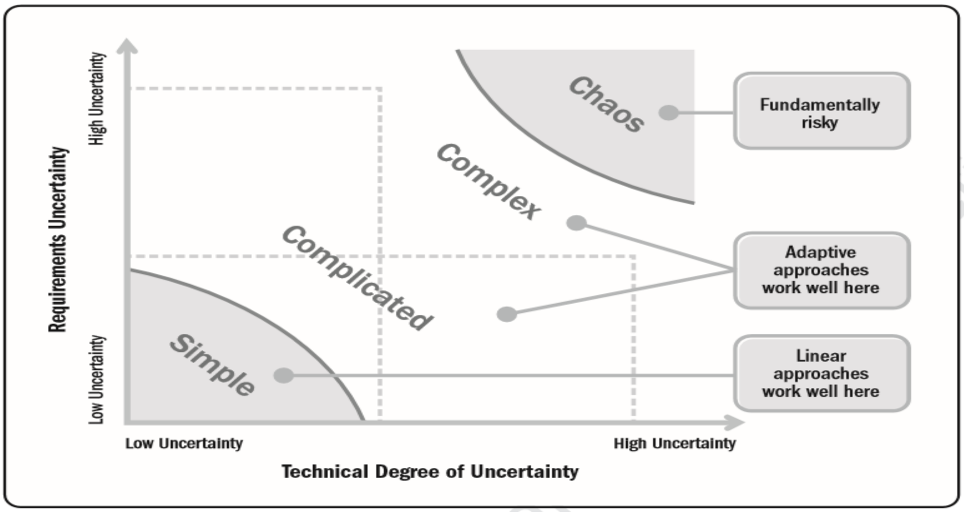

Figure 1. Uncertainty and Complexity Model inspired by the Stacey Complexity Model. Reprinted from "Agile Practice Guide," by the Project Management Institute, p. 14. Copyright 2017 by the Project Management Institute.

Characteristics				
Approach	**Requirements**	**Activities**	**Delivery**	**Goal**
Predictive	Fixed	Performed once for the entire project	Single delivery	Manage cost
Iterative	Dynamic	Repeated until correct	Single delivery	Correctness of solution
Incremental	Dynamic	Performed once for a given increment	Frequent smaller deliveries	Speed
Agile	Dynamic	Repeated until correct	Frequent small deliveries	Customer value via frequent deliveries and feedback

Figure 2. Characteristics of the four categories of Life Cycles. Reprinted from "Agile Practice Guide," by the Project Management Institute, p. 18. Copyright 2017 by the Project Management Institute.

Agile practices involve pairing, customer collaboration, daily stand-ups, review meetings, retrospectives, planning games, and other activities. Agile frameworks are *Scrum, extreme programming (XP), feature-driven development (FDD), disciplined agile delivery (DAD),* and *dynamic system development method (DSDM)* (Vallon et al., 2018; PMI, 2017; Griffiths, 2015). Other frameworks and methods are *Scaled Agile Framework (SAFe), Crystal, Lean, Kanban*, and *DevOps*. Agile teams practice *face-to-face communication, customer interaction, user stories development and estimation, stories prioritization, response to changes*; which resolves the challenges of the waterfall approach, such as *communication issues, over scoping, requirements validation, requirements documentation*, and *rare customer involvement* (Inayat et al., 2015). Some of the tools that support agile are *experience reports, lessons learned, user stories, Archinotes, Agile Workbench*, and *Scrumpy* (Vallon et al., 2018).

Agile is now a major S.E. discipline in practice and research, which was "introduced through a set of four core values and twelve principles laid out in the Agile manifesto" (Hoda et al., 2018, p. 2). While acknowledging the various agile frameworks, Al Hashimi and Gravell (2019) focused their discussion on (1) Extreme programming (XP), "a disciplined approach to producing high-quality software," (2) Scrum, a framework that identifies and resolves complex adaptive problems, and (3) Kanban, "a Japanese word for "just in time" created by Toyota, used in the manufacturing processes and now adopted in software development. Figures 3 & 4 below depict the SAFe approach to agile practice where a blend of agile practices such as Scrum, XP, and Kanban is typically used.

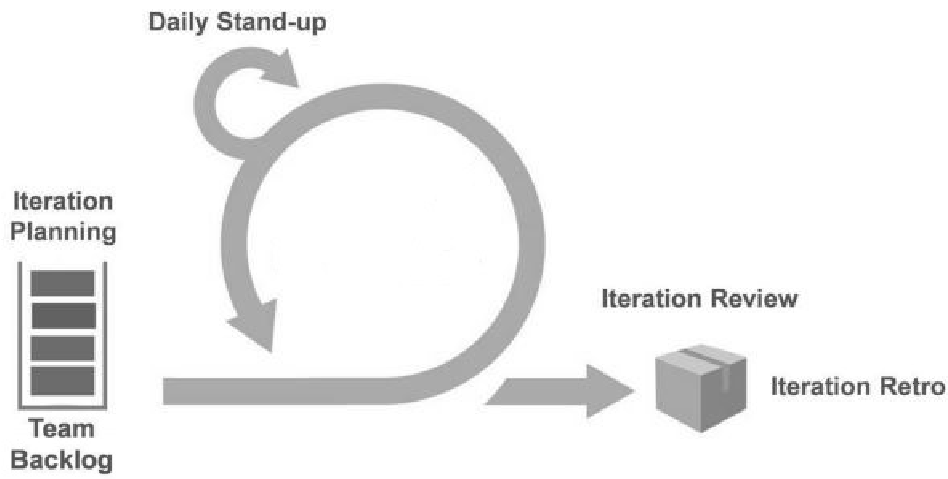

Figure 3. Scrum-XP blend in SAFe. Reprinted from "SAFe distilled: Achieving business agility with the Scaled Agile Framework," by R. Knaster, & D. Leffingwell, p. 79. Copyright 2020 by Scaled Agile, Inc.

Kanban method comes with five core principles – visualize the workflow, limited work in progress, manage flow, make process policies explicit, and improve collaboratively (Griffiths, 2017, p. 59). Table 4 shows an example of a Kanban board.

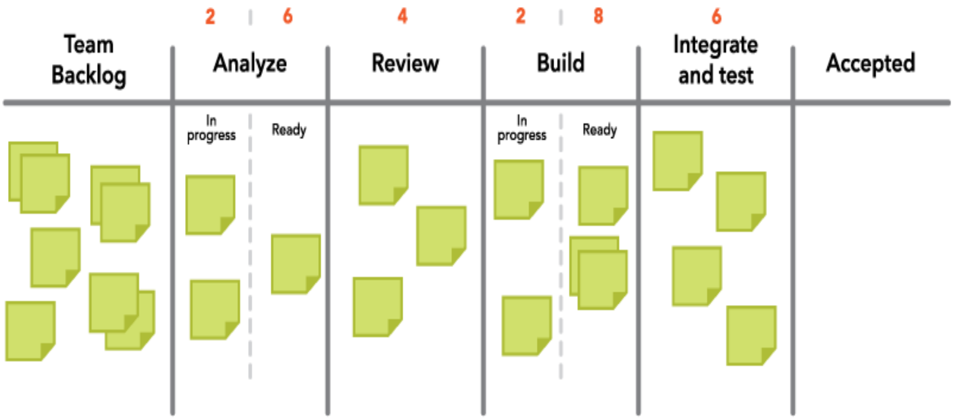

Figure 4. Kanban. Reprinted from "SAFe distilled: Achieving business agility with the Scaled Agile Framework," by R. Knaster, & D. Leffingwell, p. 79. Copyright 2020 by Scaled Agile, Inc.

Although agility does not currently have a uniform definition, there are still some common characteristics in the existing literature (Bilinkis et al., 2017). The authors proposed a "context-awareness process modeling framework" to address the challenges through the annotation of models with experts and artifacts. Bilinkis et al. (2017) also propose a "new agile approach for business process modeling with a self-organizing expert team," offering a tool and foundation for agile and business process automation. Teams and organizations have the liberty to adopt agile framework(s) that aligns with their culture and strategic goals.

Table 2a shows a summary of agile specific literature used for this study. The data collected from the articles offer information on agile principles, practices, and trends.

Table 2a. Summary description of articles - Stage 1.

Code	Method	Focus/Purpose of study	Findings/Authors' contributions
A1	Systematic literature review (SLR)	To identify the agile practice in RE	1. Agile RE needs more attention and empirical studies. 2. Lack of agile research from Africa and South America.
A2	Keynote discussions	To gather information about agile principles and practices in organizations	Agile is changing business processes in organizations.
A3	Book	Agile professional guide to aid agile practices	Agile history, frameworks, and principles.
A4	Systematic mapping (SM)	To identify the agile practice and challenges in RE	Practitioners face environment, people, and resources challenges when dealing with agile RE.
A5	Tertiary study of SLR and SM studies	To understand agile practice in software development	"Agile has become a major software engineering discipline" (p. 2).
A6	SLR	To identify agile principles and practice used in GSD settings	1. Agile is a maturing field with greater variety of research publication from 2010-2016 than 1999-2009. 2. More empirical studies is needed to improve the generalization of results and creation of stronger agile frameworks.
A7	SM	To identify agile transition and adoption frameworks and related issues	Agile transformation frameworks such as agile software solution framework (ASSF), agile adoption and improvement model (AAIM), Sidky agile measurement index (SAMI), the 4-stage process of the adoption framework, etc. (pp. 15728-15729).

4.2. Agile and Data Science/Analytics

Agarwal and Dhar (2014) suggest that the information system (IS) discipline has formulated concepts from the intersection of technology, data, business, and society for a couple of decades, and IS scholars should also consider science in order to harness the benefits of big data, which includes "delivering the right information to the right person at the right time in the right form" (p. 447). Baškarada and Koronios (2016) describe data science as a discipline that extracts knowledge from raw data through statistical methods to support organizations' decision-making. Data science roles require skills such as business domain knowledge and qualitative skills, among others. For effectiveness, *domain expert, data engineer, statistician, computer scientist, communicator*, and *team leader* roles should be included on a data science team (Baškarada and Koronios, 2016). Eachempati and Srivastava (2017) argue that big data analytics concepts originated after magnifying the 4Vs of big data – volume, variety, velocity, and veracity. The authors opined that the banking and finance domain has not seen enough research on big data and mentioned research interests like credit card analytics and fraud analytics.

Utilizing a case study approach, Conboy et al. (2018) describe seven temporal factors emerging from the 4Vs of big data. These factors are based upon time and the business value of data analytics. Pappas et al. (2018) focus their study on big data and digital transformation. They state that "big data analytics capability includes basic resources and technology (tangible), technical and managerial skills (human skills), and data-driven culture and organizational learning." Pappas et al. argue that the data science "ecosystem comprises of data actions" in *academia, public sectors, private sectors*, and *business corporations*. They propose a digital transformation and sustainability (DTS) model, depicting how data actors in various environments can develop big data analytics capability and create business value (pp. 483-484). Lozada et al. (2019) discuss big data capability and co-innovation. The authors describe co-innovation as the outcome of various stakeholders' collaboration to create new products and services.

Table 2b. Summary description of articles - Stage 2.

Code	Method	Focus/Purpose of study	Findings/Authors' contributions
A8	Literature review and case study	To develop an agile enterprise framework	1. Factors triggering framework changes based on enterprise information assets. 2. Agile enterprise modelling framework.
A9	Literature review	To critically review spikes in software development projects	Limited studies on the use of spikes in various software development domains.
B1	Literature review and feedback from experts	To address the exploding interest in the emerging field of data science & data analytics	The authors discussed the novelty of the fields, the strengths that IS community brings to this discourse, the role of predictive and explanatory modeling, and how research in this emerging area should be evaluated for contribution and significance.
B2	Semi-structured interviews	To describe the elusive myth & notion that data scientists "can do it all"	1. The authors reported six key roles considered to be requirements of an effective data science team. 2. The Primary and secondary skills for each of the roles were identified and a framework was created.
B3	SLR	To explore and empirically analyze the extent and quality of research work in business analytics worldwide	1. Identifiied the domain BDA as generic, operations, & healthcare on a global lanscape. 2. States that the publication BDA research emerges from the following institutions in india: Private, IITS, IITS, & IIMS.
B4	Literature review	To investigate the the big data and business analytics ecosystem and its interdependencies	1. The authors argues that there is a need to improve our understanding of their interactions and interrelations that lead to knowledge, innovation, and value creation. 2. Discovered the capabilities that organizations need to develop to harness the potential of big data analytics.
B5	Case study	To develop the temporal factors that affect the business value of analytics	The authors used temporality theory to identify and develop temporal factors to examine the values of analytics.
B6	Sampling & statistical analysis	To study and analyze the existing relationship between BDA Capabilities and Co-innovation	The key findings allow to positively relate BDA capabilities with better and more agile processes of product and service co-creation and with more robust collaboration networks with stakeholders.

4.3. Agile and Data Science projects

The 4Vs of big data, data science team and skill requirements, organizational structure, and team readiness all point toward big data capability at both the team and corporate levels. Achieving this capability involves learning and understanding business processes, data science methodologies, project management competencies, complexities, and risks. These factors reverberate agile as a realistic approach for data science/analytics projects.

Kayser et al. (2018) posit that an analytics value chain is created from data infrastructure and analytics, driven by business needs. Data science offers the capabilities to extract knowledge (insights) from data, which is valuable for an organization's competitive advantage. However, organizations need to be innovative by designing a structure and analytics process that aligns with the organizational goals. Kayser et al. identified *budget, testing complexity, standards*, and *compliance* as barriers that need to be addressed for an effective analytics process integration. Some complexities surrounding analytics projects, such as *stakeholders' interests, competencies*, and *viewpoints* (Kayser et al., 2018), should also be considered. These challenges require an innovative approach or an agile mindset, especially in the digital transformation age, where organizations need to become agile to respond to changes rapidly.

Table 2b. (continued).

Code	Method	Focus/Purpose of study	Findings/Authors' contributions
C1	SLR	To understand how teams work together to execute big data (BD) projects	1. The authors argue that there is no consensus on standard for executing big data projects. 2. The study suggests a list of 33 success factors for executing BD projects.
C2	Experiment	To investigate the outcomes of various project management methodologies on data science projects	Findings show significant differences based on the methodology used. Agile Kanban methodology was identified as most effective.
C3	Experiment	To investigate the impact of different project management methodologies on students data science projects	The results indicate that the project methodology used in the classroom has a significant difference in student outcomes. Agile Kanban was found to be more effective than Agile Scrum methodology.
C4	Practical experiences	To establish a process for analytics projects from first ideas to realization	Classification of analytics projects and discussion of common innovation barriers.
C5	Case study	To explore the key acceptance factors for teams to implement a data science process methodology	The author identified 10 factors that can influence data science process methodology selection. Seven positive factors are associated with compatibility, & three negative factors are associated with complexity.
C6	Ethnography (observaion, interview, & survey)	To clarify the concept of agility within a BD science project and the key process challenges that teams encounter when executing a BD science project	The authors idenfied task duration estimation, accounting for team member selection to other short burst tasks, and coordination of data science project team as the key issues. Additionally, the study suggests that Agile Kanban is beneficial in BD science teams.
C7	SLR	To identify the methodologies used in BD projects	The study revealed that agile practices in BD projects has increased since 2016, with Scrum framework being applied the most.
C8	Case study	To identify the processes and challenges of realizing values in BDA development projects	Identifies challenges in eight activities for incorporating BRM in BDA development projects.

Saltz and Shamshurin (2016) identify the need for an iterative process approach and effective team communication. They also identify project execution insights based on frameworks such as the Cross-industry standard process for data mining (CRISP-DM). In addition, Saltz and Shamshurin discuss the success factors for big data projects such as *data, governance, process, objectives, team,* and *tools*.

Saltz, Shamshurin, and Crowston (2017) define data science as “an emerging discipline” that combines expertise across “software development, data management, and statistics.” The authors proposed a process evaluation model to measure a team’s effectiveness and investigated the outcome of various methodologies such as “Agile Scrum,” “Agile Kanban,” and CRISP in data science projects. Findings show that an “Agile Kanban” method is the most effective, surprisingly over the popular “Agile Scrum” method, and the same discovery was made by Saltz, Heckman, and Shamshurin (2017). However, more studies need to be conducted in this area before establishing “Agile Kanban” as the best methodology or framework for data science projects.

In a recent study, Saltz (2018) argues that data science process methodology is not yet a popular phenomenon in practice and discusses data science methodology adoption factors, namely *compatibility, complexity*, and *relative advantage*. Saltz and Shamshurin (2019) affirm that “Agile Kanban” is beneficial for many data science teams and identifies the key issues of a data science project team as *task duration estimation, accounting for team member selection to other short burst tasks*, and *coordination*. Big data projects require a methodology that drives benefits realization from big data analytics (BDA) (Jensen et al., 2019). Quelal et al. (2019) identify “big data project methodologies” such as *Scrum, XP, Kanban, Agile data science, CRISP-DM, SEMMA, and KDD*. See table 3 below, showing models and methods used in data science and analytics projects.

Table 3. Models/frameworks from agile, data science and analytics related literature

Code	Authors' description	Model/Framework
A7	Agile transformation frameworks	ASSF, AAIM SAMI, the four stage adoption framework, etc. (pp. 15728-15729)
C1	Success factors for big data projects	CRISP-DM, SEMMA
C2	Process evaluation model & agile methodologies	CRISP-DM, SEMMA, Agile Scrum, Agile Kanban
C3	Data project framework & agile methodologies	CRISP-DM, Agile Scrum, Agile Kanban
C4	Framework of data, infrastructure, analytics, and business need	Suggested shifting to agile and iterative working model (p. 23)
C5	Data science process methodology adoption model	Agile Kanban
C6	Process methodology & agile methodologies	CRISP-DM, Scrum, Kanban
C7	Agile & non-agile methodologies	Scrum, XP, Kanban, CRISP-DM, SEMMA, KDDM

5. Discussion

According to Hoda et al. (2018), the adoption of agile practices in organizations has increased from 84% to 97% between 2007 and 2018, and 84% of organizations are still maturing in their agile practices. The favored agile frameworks used in organizations are *Scrum, XP, Scrumban* (a Scrum-Kanban hybrid), and now, the growing adoption of DevOps (Hoda et al., 2018). The challenges around agile adoption are *resistance to change* and *inadequate management support* (Hoda et al., 2018). Some literature *identified pair programming, customer collaboration, daily stand-up, continuous exploration, continuous integration*, and *continuous deployment* as the most used agile practices. With the increasing awareness of agile, many information professionals across organizations are training through Scrum Alliance, Scrum.org, Scaled Agile, Inc., Kanban University, PMI, and other accredited agents, to learn agile methods and practices, and earn certifications in Scrum, SAFe, Kanban, DevOps, and Lean quality management.

While research in agile is growing, the application of agile principles and practices is not grounded in emerging fields such as data science/analytics. The specific goals of this study are to identify the types of agile frameworks used in data science/analytics projects, the industries and regions where the practices are used, and the models/frameworks for agile practices in data science/analytics projects. Hence, this study discovered that while Scrum, XP, and Kanban frameworks are popular in agile software development, and “Agile Kanban” methodology is the effective methodologies for implementing data science/analytics projects (Saltz and Shamshurin, 2016; Saltz, Heckman, and Shamshurin, 2017; Saltz, 2018). However, an abundance of caution is necessary here since (1) there is only a handful of studies in this area; and only a few of them arrived at this conclusion, (2) the existing studies were mostly conducted as experiments involving college students in the higher education settings. Therefore, empirical studies in industrial settings are needed to establish the agile frameworks suitable for data science/analytics projects. In addressing the second research question, the literature explored shows that research in agile data science/analytics is still early. However, agile-specific literature indicates that many organizations are adopting agile principles and practices, and agile is still maturing across organizations. Although agile practices are popular in North America (the US & Canada) and Europe (the UK, Finland, Germany, Denmark, Netherlands, & Russia), literature shows little or no agile activities in Africa and South America (Vallon et al., 2018).

The third research question was addressed and summarized in “Table 3” above, showing the models and frameworks proposed by various researchers. The models and frameworks are – the agile transformation frameworks (A7), Success factors for big data projects (C1), Process evaluation model (C2), and others, as shown in the table.

6. Summary and Conclusion

Agile awareness is increasing across organizations, and time-to-market is a competitive advantage that makes a business sustainable. To remain relevant in today’s ever-changing environment, adopting agile and tailoring agile practices effectively could help organizations succeed when handling complex and high-risk projects. This study shows that agile practices in data science/analytics are either new or not yet explored in organizations. This provides opportunities to look at agile practices that support the delivery of analytics products and services in various industries across the globe. Many digital transformation projects today require an agile mindset. For instance, emerging fields and technologies (e.g., data science/analytics, cybersecurity, cloud computing, blockchain, cryptocurrency, nanotechnology, virtual reality, augmented reality, artificial intelligence) can adopt agile practices into their processes to harness business values. Researchers and graduate students could also adopt agile principles and practices in conducting their research since the characteristics of this approach include “very short feedback loops,” a “frequent adaptation of process,” “reprioritization,” “regularly updated plans,” and frequent/incremental delivery. (PMI, 2017, p. 16).

Knowledge workers, including information professionals across all industries, have sought training in agile frameworks such as *Scrum, SAFe, DAD*, and *Kanban* to learn the principles, tools, and techniques of managing complex projects and product development. With the emergence of new fields and technologies and the growing interests in data science and data analytics across industries and academic institutions, there is a need to investigate and understand effective methods and models for data science/analytics projects. Although recent research shows that “Agile Kanban” could be a way to go, it is too quick to establish a particular methodology for data science/analytics projects because there is limited research to support this claim. The limitation of this study is the absence of empirical data to show the practical experiences of information professionals in the contexts of this study. A recent study shows that employers now seek analytics project managers to manage data science/analytics projects, which indicates that empirical studies in organizations are possible. Therefore, this paper concludes that there is an urgent need for information professionals and researchers to collaborate and conduct empirical studies to establish the appropriate agile frameworks that support the rapid delivery of data science/analytics products and services.

References

Agarwal, R., and Dhar, V. (2014) "Editorial—Big Data, Data Science, and Analytics: The opportunity and challenge for IS Research", *Information Systems Research*, 25(3): 443-448.

Al Hashimi, H., and Gravell, A. M. (2019) "A critical review of the use of spikes in agile software development", In *Proceedings of the 14th International Conference on Software Engineering Advances. ICSEA 2019* (pp. 154-162).

Atolagbe, A., and Floyd, S. (2020) "Diversifying the Next Generation of Project Managers: Skills Project Managers Must Have in the Digital Age", In *International Conference on Information* (pp. 665-676). Springer, Cham.

Baškarada, S., and Koronios, A. (2017) "Unicorn data scientist: the rarest of breeds", *Program electronic library and information systems*, 51(1): 65-74.

Bilinkis, J., Zueva, A., and Zaytseva, E. (2017) "Context-aware enterprise modelling towards agile models development", In *Proceedings of the 5th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW)* (pp. 79-87).

Cerdeiral, C. T., and Santos, G. (2018) "Software project management in high maturity: A systematic literature mapping", *The Journal of Systems and Software*, 148(1): 56-87.

Curcio, K., Navarro, T., Malucelli, A., and Reinehr, S. (2018) "Requirements engineering: A systematic mapping study in agile software development", *The Journal of Systems & Software,* 139: 32-50.

Data to Decisions Cooperative Research Centres (2017) "Data Science Competency Framework", *Department of Industry, Innovation and Science*, Australian Government.

Demchenko, Y., Belloum, A., Manieri, A., Wiktorski, T., Los, W., Brewer, S., and Spekschoor, E. (2018). "EDISON Data Science Framework: Part 1", *Data Science Competence Framework (CFDS)*, Release 3 v10.

Eachempati, P., and Srivastava, P. R. (2017) "Systematic literature review of big data analytics", In *Proceedings of the 2017 ACM SIGMIS Conference on Computers and People Research (SIGMIS-CPR '17)*, Association for Computing Machinery, New York, NY, USA, 177-178.

Griffiths, M. (2015) *PMI-ACP Exam Prep (2nd ed),* Minnetonka, MN.: RMC Publications, Inc.

Hoda, R., Salleh, N., and Grundy, J. (2018) "The rise and evolution of agile software development", *IEEE Software*, 35(5): 58-63.

Inayat, I., Salim, S. S., Marczak, S., Daneva, M., and Shamshirband, S. (2015) "A systematic literature review on agile requirements engineering practices and challenges", *Computers in Human Behavior*, 51(Part B): 915-929.

Jensen, M. H., Nielsen, P. A., and Persson, J. S. (2019) "Managing big data analytics projects: The challenges of realizing value", In *Proceedings of the 27th European Conference on Information Systems (ECIS)*, Association for Information Systems, AIS. Proceedings/European Conference on Information Systems (ECIS).

Jovanovic, M., Mesquida, A., Mas, A., and Colomo-Palacios, R. (2020) "Agile transition and adoption frameworks, issues, and factors: A systematic mapping", *IEEE Access*, 8: 15711-15735.

Kayser, V., Nehrke, B., and Zubovic, D. (2018) "Data science as an innovation challenge: from big data to value proposition", *Technology Innovation Management Review*, 8(3): 16-25.

Kitchenham, B. A., and Charters, S. (2007) "Guidelines for performing systematic literature reviews in software engineering", *Technical Report EBSE-2007-01*, School of Computer Science and Mathematics, Keele University.

Knaster, R., and Leffingwell, D. (2020) *SAFe distilled: Achieving business agility with the Scaled Agile Framework*, Boston, MA.: Addison-Wesley.

Lozada, N., Arias-Pérez, J., and Perdomo-Charry, G. (2019) "Big data analytics capability and co-innovation: An empirical study, *Heliyon*, 5(1): 1-7.

Pappas, I.O., Mikalef, P., Giannakos, M. N., and Krogstie, J. (2018) "Big data and business analytics ecosystems: paving the way towards digital transformation and sustainable societies", *Information Systems and e-Business Management*, 16: 479-491.

Petersen, K., Vakkalanka, S., and Kuzniarz, L. (2015) "Guidelines for conducting systematic mapping studies in software engineering: An update", *Information and Software Technology*, 64(1): 1-18.

Prikladnicki, R., Lassenius, C., Tian, E., and Carver, J. C. (2016) "Trends in agile: Perspectives from the practitioners", *IEEE Software*, 33(6): 20-22.

Project Management Institute (PMI). (2017). *Agile Practice Guide*, Newtown Square, PA.: Project Management Institute, Inc.

Quelal, R., Mendoza, L. E., and Villavicencio, M. (2019)" Application of methodologies and process models in big data projects", In *Proceedings of the 21st International Conference on Enterprise Information Systems, ICEIS 2019* (pp. 277-284).

Saltz, J. (2017) "Acceptance factors for using a big data capability and maturity model", In *Proceedings of the 25th European Conference on Information Systems, ECIS 2017* (pp. 2602-2612), Association for Information Systems.

Saltz, J. (2018) "Identifying the key drivers for teams to use a data science process methodology", In *Proceedings of the 26th European Conference on Information Systems, ECIS 2018* (paper 58).

Saltz, J., Heckman, R., and Shamshurin, I. (2017) "Exploring how different project management methodologies impact data science students", In *Proceedings of the 25th European Conference on Information Systems, ECIS 2017* (pp. 2939-2948), Association for Information Systems.

Saltz, J., Shamshurin, I., and Crowston, K. (2017) "Comparing data science project management methodologies via a controlled experiment", In *Proceedings of the 50th Hawaii International Conference on System Sciences* (pp. 1013-1022).

Saltz, J., and Shamshurin, I. (2016) "Big data team process methodologies: A literature review and the identification of key factors for a project's success", In *Proceedings of IEEE International Conference on Big Data, Big Data 2016* (pp. 2872-2879), [7840936] Institute of Electrical and Electronics Engineers Inc.

Saltz, J., and Shamshurin, I. (2019) "Achieving agile big data Science: The evolution of a team's agile process methodology", In *Proceedings of 2019 IEEE International Conference on Big Data, Big Data 2019* (pp. 3477-3485), [9005493. Institute of Electrical and Electronics Engineers Inc.

Vallon, R., da Silva Estácio, Bernardo José, Prikladnicki, R., and Grechenig, T. (2018) "Systematic literature review on agile practices in global software development", *Information and Software Technology*, 96: 161-180.