THE STATUS OF THE ORGANIZATION OF KNOWLEDGE IN CULTURAL HERITAGE INSTITUTIONS IN ARABIAN GULF COUNTRIES

Saleh H. Aljalahmah

Dissertation Prepared for the Degree of

DOCTOR OF PHILOSOPHY

UNIVERSITY OF NORTH TEXAS

August 2021

APPROVED:

Oksana L. Zavalina, Committee Chair
Miksa, Shawne Committee Member
Daniel Alemneh, Committee Member
Jiangping Chen, Chair of the Department of Information Science
Kinshuk, Dean of the College of Information
Victor Prybutok, Dean of the Toulouse Graduate School

Aljalahmah, Saleh H. *The Status of the Organization of Knowledge in Cultural Heritage Institutions in Arabian Gulf Countries*. Doctor of Philosophy (Information Science), August 2021, 180 pp., 6 tables, 25 figures, 6 appendices, references, 129 titles.

No published studies to date examined the practices in creation or adoption of metadata in cultural heritage institutions or evaluated metadata in bibliographic databases in the Arabian Gulf counties and assessed its potential interoperability in the aggregation that would provide a central point of access to bibliographic databases of cultural heritage institutions. This exploratory study aimed to address this gap with the goal of: (1) developing understanding of the current state of information representation and knowledge organization in cultural heritage collections in Arabian Gulf countries, and (2) exploring perspectives for future developments such as creating regional large-scale portals similar to Digital Public Library of America, Europeana etc. that facilitate discovery by aggregating metadata and possible barriers to these developments.. The study is focused on a Kuwaiti, Qatari, and Omani libraries, museums, and archives. The mixed-methods research combined semi-structured interviews of the bibliographic database managers at 15 cultural heritage institutions and in-depth content analysis of a sample of 412 metadata records that represent items in these bibliographic databases for accuracy, completeness, consistency, use of knowledge organization systems, etc. This study findings make a research contribution important for evaluating the feasibility and planning of future aggregations of cultural heritage bibliographic databases. Results provided insights into possible ways to achieve interoperability in metadata for such digital portals in the Arabian Gulf region.

Copyright 2021

By

Saleh H. Aljalahmah

ACKNOWLEDGMENTS

Throughout the writing of this dissertation I have felt the ultimate support from my mother's last words which was "I want you to be a professor." My mother Hussah Alfozan passed away in 2002 after fighting cancer for almost a year. Her last wish was the true motivation for me since; I would like to dedicate this achievement to her soul. May she rest in peace.

I would like to acknowledge my wife Fedaa Boodai for her support and standing next to me during this journey. You supported me greatly and helped me a lot through those years. Thank you so much and I am glad to have you as my wife.

I would also like to thank my wonderful advisor Dr. Oksana Zavalina who expertise was invaluable in the formulating of the research topic and methodology in particular. The same applies to the committee members Dr. Miksa Shawne and Dr. Daniel Alemneh who offered a lot of assistance and wonderful collaborations during my PhD journey.

In addition, I would like to thank my father, friends and my former mentors from University of North Texas, University of Denver, Kuwait University and the Public Authority for Applied Education and Training who supported me the whole time especially during the COVID-19 pandemic. You provided peace as well as a happy distraction to rest my mind outside my research.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

1.1 Research Problem

Information and knowledge are powerful elements that help to organize and develop a person's life. From the beginning of time, people have developed ways of establishing and sharing information with others. It has recently become difficult to manage and store information because it has been increasing over time. Although the internet has helped in the storage and organization of information in some ways, it has also created new challenges: libraries and museums have had to develop new ways to present information and interact with their collection to meet users' needs (virtual touring, for example); users demand more access to information in multiple digital libraries; users expect results within a short period time and with less effort. Libraries and museums in Arabian Gulf countries have had to obtain and use systems such as bibliographic databases, which serve as digital content managements tools that help them to provide satisfying experiences for their users.

As a student at University of Denver in 2016, I was impressed by a personal story shared by one of the guest speakers in my class. In 1980s an Indian woman was walking with her daughter in New Delhi when her daughter saw attractive children's books in the library windows. The woman decided to take her daughter to that library to borrow or use some of those books. However, when they entered, a librarian asked to check her ID and told the woman that she is not allowed to check out or use in the reading room any of the library's books because she was not classified as a "worthy/trusted" person due to her low-profile job and neighborhood she lived in. Libraries in India at that time were accessible/usable by only those that were believed to be worthy of access to information. Thousands of miles away, in the Middle East, in 1990 the

Gulf War started when Iraq attacked Kuwait and invaded it for around seven months. During that time, almost all libraries, museums and archives were either destroyed or robbed. Kuwait cultural heritage institutions lost a lot of their valued print collections due to that war. This increased the realization of the importance of preservation and creating digital collections. Those two stories illustrate that access to knowledge in other parts of the world was in a totally different state than in the United States or Europe. The regional disparities are also commonly observed in how the information is organized to facilitate access.

Libraries and museums have their own systems for organizing in formation, including metadata and cataloging system. Often these systems are developed for the needs of a particular institution, without considering broader contents of information preservation and access but it would be beneficial to consider matching their systems to the professional standards. Doing so would allow to provide information users worldwide more effectively and efficiently with better services. In this study, I will examine the current status of the organization of knowledge in cultural heritage institutions in Arabian Gulf countries. It is important to know what tools and approaches for knowledge organization—i.e. digital content management tools, metadata schemas, classification systems—are being used and in what ways they are similar or different to each other. The answers to these questions will help to establish the idea of creating one large central digital portal, or central digital library, to allow users to search for any item in all digital libraries in one place with les effort. I will further discuss this information in the literature review.

The problem that I would like to solve in this research is how to curb the excessive time and effort required by users in the Arabian Gulf countries to find items in libraries and museums, which stems from the current need to search each institution separately. This was a problem that

the United States and Europe faced in the past years, however, they created large scale portals (e.g., American Memory, Digital Public Library of America, Europeana, etc.) that contain useful sources from multiple institutions. These portals helped to provide more sources to their users effectively and efficiently. The problem in the Arabian Gulf countries is that there is a clear gap between the process of creating or adopting metadata about items and understanding the needs of users. Although these users want to find information easily and in the shortest time possible, Arabian Gulf countries currently do not provide a tool that allows one to search for an item through all libraries and museum systems. To develop large-scale aggregations, metadata brought together from participating institutions needs to be interoperable. Interoperability is the process where two or more different systems exchange information and data to be used without the need to modify it (Taylor, 2004). There is a need for research that explores the current state of metadata practices and awareness regarding that problem with the goal to develop an understanding of possible ways to achieve this interoperability in metadata used by different cultural heritage institutions in Arabian Gulf countries. I will examine the status of the current organization of knowledge in cultural heritage in Arabian Gulf countries to understand how this gap can be filled. The study will also explore the difficulties that might prevent the solution to that problem. In order to address this research problem, I will look for answers to the following questions:

1.2 Research Questions

1. How is information organized in bibliographic database at cultural heritage institutions in Arabian Gulf countries?
    a. What tools and techniques in knowledge organization (data content standards, data value standards, metadata schemas, search options, database management tools) are used?

b. What are the patterns of the metadata quality (completeness, accuracy, consistency) of metadata records?

c. What levels and types of training, best practice documentation and policies are provided to metadata creators and/or managers?

d. What role does user feedback play in decision-making?

2. What are the similarities and differences in information organization within bibliographic databases among cultural heritage institutions in Arabian Gulf countries?

3. What are the views of metadata managers at cultural heritage institutions in Arabian Gulf countries on the creation of large-scale aggregations of digital content?

1.3 Purpose Statement

The purpose of this study is to examine the status of the organization of knowledge in cultural heritage institutions in Arabian Gulf countries to suggest an idea of creating a large-scale portal that aggregates digital collections from different institutions (similar to Digital Public Library of America and The European Library) to help users access information more easily. As part of this study, I will be analyzing similarities and differences in how information is organized and presented to the users in cultural heritage institutions in Arabian Gulf countries. Results of this examination are expected to provide a basis for mappings and crosswalks to facilitate interoperability and create a large central digital portal or library for the Arabian Gulf countries, which will harvest metadata from participating institutions.

CHAPTER 2

LITERATURE REVIEW

2.1 Historical Background about the Organization of Knowledge and Its Main Tools

Information flow has become one of the most critical challenges for libraries and museums today. Ranganathan, an Indian expert in library and information science, established five rules that guideline the relationship between the library and the users. His rules were adopted by many cultural heritage institutions because they prioritize user needs in library services. Today, institutions are working hard to provide users with quality metadata for information sources—one of their rights—so that they may easily access them. While metadata, as a term, is considered new in some ways, using bibliographic records that describe information resources is not a new idea; the organization of knowledge is a very old concept that people have been focusing on for centuries. The idea of cataloguing or organizing information has been around since 235BC in the Alexandria library in Egypt (Haynes, 2004). Callimachus of Cyrene (305-235 BC) was a librarian in the Alexandria library in Egypt with a collection of over 500,000 scrolls (Haynes, 2004). This library's catalog itself contained roughly 120 scrolls, with titles, which were grouped by genre and subject (Haynes, 2004). People used to maintain and organize their information and knowledge in similar ways until the printed catalog was created. It was not until 1605 that the Bodleian library in Oxford began to use a printed catalog (Gartner, 2016).

The history of bibliographic records has changed over time: in the Panizzi Era handwritten entries were used; in the Card-Catalog Era; lastly, in the Electronic Era (Panizzi, 1850; Svenonius, 2000). Harleian Library's catalog was created by Samuel Johnson in 1743 (Bade, 2004). In 1873, Melvil Dewey started working on a plan to create a classification system

to organize information in a way that made it united in all libraries (Dewey, 1876). The Dewey Decimal Classification (DDC) system divides knowledge into nine classes, each one then being separated once more into nine special subdivisions, which are then numbered from one to nine (Dewey, 1876). The idea of Dewey's system was to provide an easy access and retrieval system that both users and librarians could use.

The Dewey Decimal Classification system has been in use by many libraries over the years and it is still used by libraries in the Middle East area today. In the 1980's, DDC was represented in an electronic form, which made it even more useful for those still using it (Beal & Mitchell, 2010). However, other places of the world still relied upon an alternate classification system known as the Library of Congress Classification System (Lund, Agbaji, Tijani & Omame, 2019). Both classification systems are used today in different academic libraries in United States, however, the numbers show that 25% of academic libraries use DDC (Lund, Agbaji, Tijani & Omame, 2019). The Universal Decimal Classification (UDC), created in 1896, is another classification system used very commonly in Europe. In 2006, UDC was acknowledged as the second-most used classification system in the world when it comes to the number of libraries using it (Slavic, 2008). However, although the UDC system was based on the DDC system, over the years it has continued to develop in structure, vocabulary, and syntax (Slavic, 2008). These are the most popular classification systems that libraries are using in organizing their collections across the world.

Another big development in the history of knowledge organization came with the creation of the Machine-Readable Catalog. In 1964, the Council on Libraries Resources Inc. (CLR) established the first conference that supported and examined the idea of having a Machine-Readable Catalog with the hopes of solving problems that libraries used to face at that

time (Avram, 1969). Additional conferences were organized in 1965 and 1966, which resulted in a pilot machine readable catalog project called MARC (Avram, 1969). Establishing Machine-Readable Cataloging standard (MARC) at that time was a big improvement for cataloging and organizing information for many libraries. While museums similarly benefitted from the system, MARC was created primarily to implement the card catalog system that libraries had been using for a long time in the electronic environment. MARC had many developments and versions that maintained its use throughout history.

It is important to know that in the 1900s, cooperation efforts between libraries (libraries-libraries), museums and libraries (museums-libraries) and between museums (museums-museums) were very rare. Today, libraries and museums are much more apt to share, cooperate, and maintain their bibliographic information about their collection (Stegaeva, 2016). There are many current programs that institutions got involved in as a cooperation project to share and maintain their records. Such projects include BIBCO (the Monographic Bibliographic Record Program), SACO (the Subject Authority Cooperative Program), NACO (The Name Authority Cooperative Program) and CONSER (the Cooperative Online Serials Program; before 1986, the Conversion of Serials) (Stegaeva, 2016). All these programs and cooperative efforts helped to create large central digital portal, harvest data, provide more access for users, and maintain the collection. This helped libraries and museums to provide users with an effective and efficient way to access their collection.

Another important development that affected the shape of libraries and museums came about with the creation of digital content management tools. Content management systems were created and used often in the 1990s along with the development of the web and digital content (Chen, Li & Chen, 2014). Content management tools were then created to provide better

organization and management to the institutions' collections (Mescan, 2004). Having a content management system helps to improve information accuracy, support preservation, and manage access to information (Han, 2004). The main task of content management systems is to centralize the content in one repository to provide better organizing, sharing, and tracking (Mescan, 2004). For example, DSpace, ContentDM, Fedora, Islandora, Hydra, etc. are open-source digital content management tools that are being used widely in libraries today. According to Kawano, Greenstone, Drupal, Joomla! Wordpress, XOOPS and Zikula are some commonly used content management systems in both libraries and museums because of their flexibility and interoperability (2012). All these content management systems helped to improve the performance of libraries and museums for years and increase the quality of their user services.

2.2 Metadata and Access

Because one of the primary goals of libraries and museums is to serve their users, it is important for them to consider users' needs. To get a clear sense of needs, libraries should ask their users questions about their hobbies and interests, as well as what types of alerts they might like to receive in their mailbox. Asking these questions would help libraries build user profiles and help them to better understand users' needs (LaRue, 2010). Accessibility is one need that libraries and museums should consider when building and organizing their collection. Using longer and more detailed metadata records is one of the many ways libraries can help users understand the data of an object more meaningfully (O'Dell, 2013). Metadata is key to interoperability and should therefore enhance retrieval performance, provide a way of managing electronic objects, and help to determine the authenticity of data. It is a large part of the future of libraries' services (Haynes, 2004). It is also important to know that each type of library may have different concepts that affect the creation of their metadata. For example, privacy is one of the

most important characteristics of metadata policy in the health industry (Sweet & Moulaison, 2013). Because patient information is very private, doctors will have more access to a patient's records than nurses do.

While this literature review focuses on organization of knowledge, it is also important to first explain the need for digital environments and digital collections. The digital library was created as an idea to serve and provide users with easy access to resources; it is a special library that provides access to different kinds of digital resources such as e-books, e-journals, videos, audios and images (Chowdhury & Chowdhury, 2002). The digital library provides users the ability to organize, store, and retrieve resources from the digital collection in a way that is easier for most people (Calhoun, 2014; Lesk, 2005).

Graduate and undergraduate students appreciate the value of metadata in digital libraries (Nemeth, 2009). However, the recent changes in publishing formats from traditional printed formats to digital forms have created challenges for libraries, which has led many of them to rethink their services and collections (Allen, Cole, & Schmitz, 2002). This means users now have different preferences of items' format that they are interested in using. Libraries must consider new issues when they create their digital library such as the cost, electronic copyright and the possibility of archiving their documents (Allen, Cole, & Schmitz, 2002). Still, metadata professionals show high interest in continuing to improve their knowledge and skills through continuing education, because technology skills and knowledge are needed in the digitizing process ((Park, Tosaka, Maszaros & Lu, 2010; Borbinha, 2004). The majority of participants (35 out of 59 librarians) in a study by McKenna, Debruyne and O'Sullivan mentioned that they do not need technical support or help to use electronic cataloging tools (McKenna, Debruyne & O'Sullivan, 2018). Thus, libraries should think about accessibility as an important factor that

affects the digital library's performance (Joint, 2006).

Libraries carefully choose the hardware, software and electronic resources management system they will apply in their digital library (Maurya, 2011). This can be different according to the libraries' goal, users, geographical region and mission. Many of the academic digital libraries in the United States, for example, use the Dublin Core metadata standard for their collection (Park & Childress, 2009). MARC is the predominant metadata standard for libraries and is not widely used outside of the library community (Machovec,1998). Libraries are using traditional tools and standards in organizing their collection, such as the Dewey Decimal Classification scheme and Library of Congress subject headings (LCSH) to provide subject access to their non-print materials (El-Sherbini, 2000).

Libraries have helped in providing more accessibility of information when they develop the physical card catalog to an online catalog with descriptive metadata (Thiele, 2016). That development increased the user services quality as it allows the users to use the catalog from distance. Developing metadata schemas for libraries should go through participating in local and international discussions and standards development (Nelson, Belden, Reis, Gelaw, Phillips, & Dunlop, 2005). Librarians are the architects of their collection and they should provide and organize a good metadata of it (Weaver, 2007).

Today's media is providing direct and easy access to information, which increases the importance of librarians, documentarists and archivers to organize and manage said information (Gutiérrez & Ávila, 2014). Not only that, but many users today search for all options related to any news they hear to make sure they can make the accurate judgments (Gutiérrez & Ávila, 2014). In other words, users are demanding access to more versions of information in order to get the whole idea from all perspectives. This shows the increase in both users' intelligence as

well as their expectation to have access to the information they demand. One of interesting application of offering access to the library in the past was requiring a deposit of cash for browsing the library's book or having book in high shelves that cannot be reached by the users without the staff help (Lemos, 1981). All those applications indicate the fact that libraries at that time were trying to protect the books from the reader and not for the reader (Lemos, 1981). Libraries and museums should consider this when striving to grant users access to their collection without problems. Development regarding accessibility did not only come from the users' perspective, but also from librarians and professionals. Ranganathan, an Indian expert in the library and information field, created five laws that guarantee users' right to access and use collections (Bhatt, 2011). Those rules were respected and applied by many libraries all over the world. Ranganathan's five rules are:

1. Books are for use.
2. Every reader his/her book.
3. Every book its reader.
4. Save the time of the reader.
5. The library is a growing organism.

Many of the libraries' services were inspired by these five rules. Some services include the library tour, newsletters, posters, advertising and many others that aim to educate users of their accessibility limit and motivate them to use the library's collection (Bhatt, 2011). The same response occurred within museums when they applied a multitude of alternate options of accessibility for their users such as the online tour, which allows users to take a tour in the museums from a distance. All these services and modifications to libraries and museums show that these institutions are aware of the importance of accessibility for users, as well as their

desire to meet users' needs.

2.3 Theoretical Cataloging Literature

The important part of providing access to information for users through metadata is data content standards: Anglo-American Cataloguing Rules (AACR), Resource Description and Access (RDA), International Standard Bibliographic Description (ISBD), Cataloging Cultural Objects (CCO), Describing Archives: A Content Standard (DACS), etc. These standards were adopted widely by cultural heritage institutions (Gilliland, 2016). Data content standards worked as guidelines and rules for the format and syntax of the data value that was usually entered into metadata fields by catalogers (Gilliland, 2016). Cataloging rules such as AACR help libraries to create their metadata, indexes, abstracts, and bibliographic records (Gilliland, 2016).

Another important component of organizing information is data value standards. The data value standards mostly control the use of terms in order to provide better recall and precision. Examples of these standards include Library of Congress Subject Headings (LCSH), Library of Congress Name Authority File (LCNAF), LC Thesaurus for Graphic Materials (TGM), Medical Subject Headings (MeSH), Art & Architecture Thesaurus (AAT), Union List of Artist Names (ULAN), Getty Thesaurus of Geographic Names (TGN), etc. (Gilliland, 2016). Maknaz, an Arabic-made thesaurus, was created mainly to meet the needs of information centers and libraries in the Arab World in order to keep up with the technology development in the rest of the world (http://en.maknaz.org/). Maknaz is "an expanded thesaurus that can be described as an authority list of descriptors or indexing terms integrated into an information system application" (http://en.maknaz.org/). It is also a general thesaurus that contains a controlled vocabulary of subject terms, personal corporate and geographic names, and genre terms (El-Sherbini, 2015). Maknaz has a total of 53979 authority records and is accessible in three languages: Arabic,

English and French (http://en.maknaz.org/). Maknaz was created by three Arabic organizations: The Abdul Hameed Shoman Foundation (AHSF), Juma AI-Majid Center for Culture and Heritage and The Dubai Culture & Arts Authority (http://en.maknaz.org/).

Information is also organized with the help of data structure standards such as Resource Description Framework, MARC (Machine-Readable Cataloging format) fields, MARC21, MARCXML Encoded Archival Description (EAD), Dublin Core (Simple of DCMES 1.1. and Qualified or DCTERMS), Categories for the Description of Works of Art (CDWA), VRA Core. These standards are used for structuring, encoding, and exchanging metadata records between cultural heritage institutions. Finally, data format/technical interchange standards (metadata standards expressed in machine-readable form) were created. These standards often occur as a symptom of a specific data structure standard encoded or labeled for machine processing (Gilliland, 2016).

The practice of cataloging, as it is used today, was established when MARC was created. MARC was established in the mid-1960s as a new standard for data format of catalogue records (Haynes, 2004). MARC was quickly adopted among the libraries' community, with forty libraries expressing a willingness to participate in the MARC project in 1964 (Avram, 1968). A total of seven million MARC records were processed by OCLC in 1987 (Gorman, 1990). In 1999, the first version of MARC XML was established, which allowed MARC to enter the Internet (Tennant, 2002). Libraries used MARC, as well as classification systems, to organize and catalog their collection for years. 132 out of 230 librarians who participated in a study mentioned that they were working in metadata cataloguing processes in their workplace; 73% of them said MARC 21 was the metadata format they used in that process (McKenna, Debruyne & O'Sullivan, 2018). While there are many classification systems, the most popular two included

the DDC and Library of Congress classification systems.

In the mid-1990s, experts in the fields started exploring the idea of having a core set of semantics for web-based resources, which eventually became known as the Dublin Core (DC) metadata (Haynes, 2004). Experts also began working on a tool that would help in organizing the access of digital items on the internet. Resource Description and Access (RDA) is a new bibliographic content that is replacing Anglo-American Cataloguing Rule (Mansor & Ramdzan, 2014). RDA was created based on two conceptual models: functional requirements for bibliographic records (FRBR) and functional requirements for authority data (FRAD) (Mansor & Ramdzan, 2014). RDA helps users to find, identify, select, and obtain resources that meet their needs not only in libraries, but also in museums, archives and digital libraries (Mansor & Ramdzan, 2014).

2.4 Practical Cataloging Literature

According to Dr. Taghreed Alqudsi from Kuwait University, cataloging has always been a tough job to do because it needs full attention and focus to details (personal communication, September 23, 2012). Catalogers invest most of their time in creating bibliographic records for the entire collection, and yet there remain mistakes in their catalog (Budanović & Žumer, 2018). Those mistakes could be the result of misjudgments, lack of provided information, or simply human error (Budanović & Žumer, 2018). With that being said, catalogers and the practical work of cataloging has been a field of study that researchers have conducted many studies on. Catalogers' judgments are one of the biggest issues that shape the description of an item in the catalog. Catalogers' judgments are not related to the cataloging decisions they make, but also decision making can involve evaluating cataloging work done by others (Diao, 2018). This means catalogers may criticize others' work because of their own judgments. For example, an

LGBT book or novel could be described as a romance book in a library, while the same text can be described as a sin book, or something similarly, in another library. These varying results can be misleading to users and thus create difficulty in finding the book. Another example of how cataloger judgments can affect cataloguing is seen when entering data textual fields in MARC (5XX notes, subject headings); the cataloger's judgments will indicate what text should be entered and whether such textual entries are necessary (Diao, 2018). An actual example is when I worked on cataloging the old image collection in the Denver Museum of Nature and Science and was told to enter in metadata records only the names of people in some images, while entering an extra description for other images. It is clear that catalogers have their own judgments on what images need more description and what ones need only the names of the people who were in those images (Diao, 2018).

One challenging situation that libraries faced came about when users began demanding non-book materials, such as video games. Catalogers struggled to adequately apply cataloging rules and practices when describing these materials. Computer materials were not mentioned in either the Anglo-American Cataloging Rules, North American Edition (1967) nor the ALA Cataloging Rules (1941); It was not until 1983 that a manual of AACR2 Examples for Microcomputer Software and Video Games was produced by Nancy Olson. Cataloging professionals at that time were developing policies that could help to describe those types of materials. Chapter nine of AACR2 documents, released in 1988 focused on the rules for description of computer files. It defined computer files as "files that are encoded for manipulation by computer. These files comprise data and programs. Computer files may be stored on, or contained in, carriers available for direct access or remote access." (De Groat, 2015). Currently, RDA and the MARC 21 provide better and more controlled vocabulary for the

described electronic resources and data files (De Groat, 2015).

2.5 Theoretical Metadata Literature

Originally from the computer science field, the term metadata had been adopted by information science in the 1990s to refer to bibliographic records in traditional library catalogs and digital library databases. Metadata as a concept has become very popular beyond the information science community in the last few years when the general public was introduced to it: "Metadata became a commonly used term in the media during the Snowden affair" (Gartner, 2016 p. 1). There are several different definitions of metadata depending on the type, however, it is quite simply data about data (Baca, 2016). It provides identifications and descriptions of all objects (Machovec,1998). The definition may change in terms of its purpose. Christel (2009) writes:

> [Metadata] might be used to help us to find the resource (resource discovery metadata) or might tell us what it is (descriptive metadata). It might tell us where the resource has come from, who owns it and how it can be used (provenance and rights metadata). It might describe how the digital resource was created (technical metadata), how it is managed (administrative metadata), and how it can be kept into the future (preservation metadata). Or it might, as mentioned earlier, help us to relate this digital resource with other resources (structural metadata). (p. 7)

This shows the need of using a metadata schema to organize the digital collection for libraries. Metadata plays an important role in the information society and understanding its importance and how it works will help the next generation to deliver more effective services and systems (Haynes, 2004). For years, libraries have thought carefully about the importance of the metadata schema used in their collections, because it is required to provide access to items electronically in digital libraries (Wallace, 2001). Using metadata in an effective way will make an institution's collection more accessible and open to the Web environment (Bowen, 2008).

Good metadata, as described by the national information standards organization (NISO),

not only provides easy access and retrieval for users, but also provides a very well-organized method for those who manage the collection (Moulaison & Dykas, 2016). Frameworks and criteria for evaluating the quality of metadata have been developed, such as that shown in Figure 2.1, which shows that metadata should be assessed based on seven criteria: accessibility, conformance to expectations, logical consistency and coherence, timeliness, completeness, accuracy and provenance (Bruce & Hillmann, 2004). Several levels of metadata must be addressed when defining quality such as the semantic structure, the syntactic structure and the data values themselves (Bruce & Hillmann, 2004).

Quality Measure	**Quality Criteria**	**Compliance indicators**
Completeness	Does the element set completely describe the objects?	Application profile; documentation
	Are all relevant elements used for each object?	Visual view;*; sample
Provenance	Who is responsible for creating, extracting, or transforming the metadata?	OAI server info,† File info, TEI Header‡
	How was the metadata created or extracted?	OAI Provenance; colophon or file description
	What transformations have been done on the data since its creation?	OAI About
Accuracy	Have accepted methods been used for creation or extraction?	OAI About; documentation
	What has been done to ensure valid values and structure?	OAI About; visual view; sample; knowledge of source provider practices; documentation for creator-provided metadata; known-item search tests
	Are default values appropriate, and have they been appropriately used?	Known-item search tests; visual view
Conformance to expectations	Does metadata describe what it claims to?	Visual view; external documentation; high ratio of populated elements per record

Quality Measure	Quality Criteria	Compliance indicators
	Are controlled vocabularies aligned with audience characteristics and understanding of the objects?	Visual view, sample, documentation; expert review
	Are compromises documented and in line with community expectations?	Documentation; user assessment studies
Logical consistency and coherence	Is data in elements consistent throughout?	Visual view
	How does it compare with other data within the community?	Research or knowledge of other community data; documentation
Timeliness	Is metadata regularly updated as the resources change?	Sample or date sort of administrative information
	Are controlled vocabularies updated when relevant?	Test against known changes in relevant vocabularies
Accessibility	Is an appropriate element set for audience and community being used?	Research or knowledge of other community data; documentation
	Is it affordable to use and maintain?	Experience of other implementers; evidence of licensing or other costs.
	Does it permit further value-adds?	Standard format; extensible schema

Figure 2.1: Measures for evaluation of metadata quality suggested by Bruce and Hillmann (from Bruce & Hillmann, 2004)

The six aspects that reflect the value of the metadata in any database are as such: it should be accurate, prompt, well-directed, brief, rare and complete (Bade, 2004). Accuracy, consistency and sufficiency are three characteristics that determine the quality of any metadata schema (Robertson, 2005). Two main levels of quality assessments can be applied in evaluating metadata records: compliance, utility and appropriateness of elements in terms that support the goal of metadata schema used (Moen, Stewart & McClure, 1997). One of the common ways to ensure the metadata quality is its ability to be modified in order to correct gaps in quality (Zavalina, Kizhakkethil, Alemneh, Phillips, & Tarver, 2015). Good metadata schema also

provides a guiding framework for building a great digital collection (Park & Childress, 2009). However, metadata quality criteria for other areas, beyond cultural heritage institutions (e.g., in business, etc.) have not been researched, and those developed in the information science field may or may not apply.

The quality of metadata in digital collections needs to be studied more (e.g., Moen, Stewart & McClure, 1997). For example, new digital objects, such as video games, need new metadata schema that can better describe their nature (Lee, Tennis, Clarke & Carpenter, 2013). The quality of metadata plays a role in linking various types of metadata that describe the same object (Jones, 2017).

2.6 Practical Metadata Literature

Many of the major metadata standards created in the last few years aim to solve problems related to classifying and cataloging digital collections (El-Sherbini, 2000). However, there is no tradition in sharing metadata between institutions such as museums and libraries (Farneth, 2016). Having more than one metadata schema that describes the same item is common in large databases (Thornburg & Oskins, 2007). The metadata schemas' creation process involves many specialists such as catalogers, indexers, webmasters and data in-putters (Greenberg, 2002). Using a specific metadata schema helps in creating large scale indexing or harvesting projects.

There are many metadata formats that can offer either simple or rich descriptions of the digital collection (Machovec,1998). The Colorado digitization program was a challenge when dealing with the digitizing process, because there was no single common metadata schema being used in all the participating institutions (Bailey-Hainer & Urban, 2004). The digitization process is a sensitive process for museums as they deal with unique and valuable items (Ruthven & Chowdhury, 2015). This process increases the need for cultural heritage institutions to create and

follow polices and guidelines to digitize their collection. The digitization process created challenges for museums to consider the best tools for creating, accessing, managing and storing their collection, especially since having more than one metadata format can create challenges (Ruthven & Chowdhury, 2015). It is the librarians' responsibility to maintain a recognizable metadata schema for their collection in order to provide better service to the users: "Librarians should become filters between their users and the huge growing volume of information "(Bade, 2004, p. 149). While most libraries use metadata schema and metadata guidelines, they ought to consider that their metadata schema should match professional standards such as the Dublin Core standard, which was developed as an international and interdisciplinary schema for a variety of resources across many domains (Miller, 2011; Zeng & Qin, 2008). The most used Dublin Core standard (Metadata Element Set 1.1. or DCMES 1.1) has 15 metadata elements. Using this metadata standard helps to make web documents more visible to search engines (Coyle, 2012). The metadata standard is one of the issues that must be considered with digital content and its preservation and creation (Copeland, 2014). When creating the Dublin Core metadata standard, the intention was to identify a set of elements that described both digital and physical objects (Gartner, 2016). There are also some institutions which used the Dublin-Core-based application profile, which is basically their modified copy of Dublin Core standards. For example, the Australasian Virtual Engineering Library's (AVEL) Metadata Set contains 19 elements (Chan & Zeng, 2006). Creating a Dublin-Core-based application profile is a process that in most cases contains the following steps: select a base metadata namespace; select elements from another metadata namespace; define the local metadata elements; enforce the application of the elements (Zhang, 2004; Duval et al., 2002). A great way to look at Dublin Core is as a small language that presents a particular level of statements regarding resources (Hillman, 2005). A modification has

been applied to Dublin Core standards called qualified Dublin Core (DCTERMS) which includes refinement terms and extensions to the simpler Dublin Core 15 elements (Hillman, 2005). In other words, qualified Dublin Core (DCTERMS) provides more details regarding the simple Dublin Core elements, which thus increases the benefit of using it for different type of information objects. Subject elements can be considered as indicators of connectedness for metadata records, as they can be unique (Phillips, Zavalina & Tarver, 2020). Some institutions required to use some elements and let other elements remain optional. The University of North Texas digital library requires using the elements of title, language, description, subject, resource type, and format, while making the other elements optional (Phillips, Zavalina & Tarver, 2019).

Companies, governments and international organizations started using Dublin Core metadata in the 2000s to describe the resources they held, owned or produced (Hillman, 2005). However, some institutions decided to use and develop their own version of metadata standards to better meet their needs. Many of the language archives use their own developed metadata schema that meet their needs (Burke & Zavalina, 2019).

Museums also use metadata to maintain their collection. One of the problematic obstructions that museums' collection management records faced in the past was the lack of standardized controlled vocabularies. This issue led to using terms that were unfamiliar to the general public (Bailey-Hainer & Urban, 2004). Before Categories for the Description of Works of Art (CDWA) and the Visual Resources Association (VRA) Core 4.0 metadata standards for museums were developed in the 2000s, the databases in museums had commonalities with library databases. However, their metadata were not standardized using MARC, as was done in libraries (Caplan & Haas, 2004). The situation now is different because several high-quality controlled vocabularies, such as the Getty Research Institute's Thesaurus of Geographic Names

TGN and Art and Architecture Thesaurus AAT etc., have been developed and are now used in museums.

Administration in libraries and museums plays a very important role in motivating their employees to establish high-quality work, especially when it comes to new projects. “The quantity and quality of people working to build a database depends on the attitudes of the administration toward the work they do” (Bade, 2002, p. 26). This means that the institution’s administration is responsible for motivating and appreciating building good metadata for their collection.

Cocciolo examines how a digital asset management system (DAM) was employed by the New York art museum to store and manage their department’s photographic collection (Cocciolo, 2014). However, Cocciolo found that the most challenging part in that process was the social aspect of it; employees were not motivated or prepared to work on digitizing their collection because of the lack of awareness they had about the importance of that process (Cocciolo, 2014).

Functional requirements for bibliographic records (FRBR) is a 1998 recommendation of the International Federation of Library Associations and Institutions (IFLA) to restructure catalog databases to reflect the conceptual structure of information resources. FRBR is one of the conceptual models that libraries previously used to describe their digital collection (Coyle, 2012). The International Consortium of Museums developed the functional requirements for bibliographic records object oriented (FRBROO) standard to implement FRBR concepts in museums and other institutions (Coyle, 2012). Functional Requirements for Authority Data (FRAD) is another functional requirement model that covers the authority data created by libraries and museums (Taniguchi, 2013). Another functional requirement that focuses more on

subjects and authority data is the Functional Requirements for Subject Authority Data (FRSAD) (Taniguchi, 2013). The Library Reference Model (LRM), which was accepted by the International Federation of Library Associations and Institutions (IFLA) in 2017, is another functional requirement that defines five users' tasks: find, Identify, select, obtain and explore (Žumer, 2018). It is important to know that IFLA recognized LRM functional requirements as an IFLA standard (Žumer, 2018). CIDOC Conceptual Reference Model (CRM) is a conceptual model that has been used very often in many cultural heritage institutions, especially archives and museums (Lin, Hong & Doerr, 2008). CIDOC CRM was able to provide a common understanding of cultural heritage information by offering an extensible semantic framework for that cultural heritage information to map (Lin, Hong & Doerr, 2008).

Many museums are aiming now to map their metadata to a conceptual reference model (Farneth, 2016). Museums in Europe are facing problems in delivering effective integration of their images collection (Lewis, Martinez, Abas, Fauzi, Chan, Addis … et al., 2004). However, many cultural heritage institutions found that harvesting metadata—collecting metadata from a data provider—is the solution for that problem. Harvesting metadata is the process where the institution collect metadata from a data provider (Roy, Sutradhar & Das, 2017). Harvesting projects should realize the fact that common metadata standards lack descriptions of the collections within archives, libraries, and museums (Marcum, 2014). In 2006, research was conducted on the metadata standard selection for 160 digital collections of museums and digital libraries, and the results were that they mostly use the Dublin Core standard (Palmer, Zavalina & Mustafoff, 2006). Future museum users need descriptive metadata that offer a connection and relevance within the museum collection to find more relative resources (Ray, 2017). Museums' communities take a different approach to the characteristics that represent the quality of metadata

standard for their collection (Robertson, 2005). Unfortunately, most museums are not funded to provide free services like those offered in libraries, and thus they cannot provide a free metadata of their collection (Roel, 2005).

Group/Count	Cataloger/Cataloging	Metadata	Cataloging or Metadata (Job Description)	Metadata and Cataloging	Metadata + Other
1	Asian Language Cataloger	Assistant/Associate Librarian for Metadata	Digital Archivist	Assistant Law Librarian for Cataloging and Metadata	Database Management and Metadata Librarian
2	Catalog and Acquisitions Librarian	Biblio-Metadata Librarian	Digital Initiatives Librarian	Catalog/Metadata Librarian	Digital and Metadata Services Librarian
3	Catalog Librarian	Digital Initiatives Metadata Librarian	Digital Preservation Librarian	Cataloging and Metadata Librarian	E-Resources and Metadata Services Librarian
4	Catalog Librarian (Music/Media)	Digital Metadata Librarian	Digital Projects Librarian	Chinese Language Cataloging/Metadata Librarian	Metadata and Electronic Resources Management Librarian
5	Cataloger/Taxonomist Librarian	Discovery Metadata Librarian	Discovery Services and Data Curation Librarian	Metadata & Cataloging Services Librarian	Metadata and Systems Librarian
6	Cataloging and Serials Librarian	Electronic Resources Metadata Specialist	Discovery Services Librarian	Metadata and Catalog Librarian	Metadata/Discovery Librarian
7	Cataloging Librarian/Manager	Metadata Analyst Librarian	Resource Access Librarian	Metadata and Catalog Services Librarian	Metadata/Electronic Resources Librarian
8	E-Resources Cataloger	Metadata Librarian	Resource Description Coordinator	Metadata Management / Cataloging Librarian	Metadata Librarian and Coordinator of Monographic Acquisitions
9	East Asian Collections Cataloger	Metadata Librarian for Special Collections	Taxonomy Manager	Metadata/Catalog Librarian	Technical Services and Metadata Librarian
10	Electronic and Continuing Resources Cataloger Librarian	Metadata Management Librarian	Technical Services Librarian		
11	Electronic Resources/Non-Print Cataloger	Metadata Strategist			
12	Librarian 2 (Cataloger)	NextGen Metadata Professional			
13	Librarian I - Technical Services - Cataloger				
14	Monographs Original Cataloger				
15	Original Cataloger & Japanese Studies Bibliographer				
16	Special Collections Cataloger				

Figure 2.2: Ads describing the relationship between jobs of metadata and cataloging from (Geckle & Nelson,2017)

Although metadata was created to serve libraries in the first place, it is now being using widely in many other industries such as publishing, record management and geospatial communities (Haynes, 2004). With that being said, a confusion in who should create and manage metadata has been raised in these other industries. It is important to know the differences in working on metadata between data managers and librarians (Geckle & Nelson, 2017). Data managers usually see element descriptors as metadata, and the text of the title would constitute

data, while librarians view the values in MARC field as metadata (Geckle & Nelson, 2017). There is a need to clarify what metadata librarians ("Metadaticians") are expected to do, because they do not have a long history in the profession like catalogers (Geckle & Nelson, 2017). This need becomes even clearer when an examination of job ads in a Geckle and Nelson study demonstrated that metadata librarians were combined to many other jobs in the job ads, as shown in Figure 2.2 (Geckle & Nelson, 2017). Because of this, metadata professionals must figure out ways to clarify their job; this also should be reflected on school programs and education in library and information science (Geckle & Nelson, 2017).

2.7 Metadata Interoperability

Interoperability became emphasized in library and information science and practice in the 2000s. Interoperability can be defined as "The compatibility of two or more systems such that they can exchange information and data and can use the exchanged information and data without any special manipulation" [Taylor 2004, p. 369]. Interoperability becomes one of the important principles in metadata implementation in the information community (Chan & Zeng, 2006). There are several factors and concepts related to the interoperability process that need to be introduced in this section.

Metadata application profiles are a way of achieving metadata interoperability. When an institution creates their own metadata element set based on other standard metadata schemas, they will have an application profile. An application profile is a set of metadata elements drawn and combined from one or more namespace schemas by implementors for a particular local application (Heery & Patel, 2000). The need for an application profile is to promote interoperability, data sharing and linked process between institutions (Krause, Clary, Ogletree & Greenberg, 2015). There are standards for building application profiles with a specific

framework based on the Dublin Core (DCAP) (Baker & Coyle 2009). DCAP standards can be summarized in four main points: Resource Description Framework (RDF), the Dublin Core™ Abstract Model (DCAM), the Dublin Core™ Description Set Profile (DCSP), and the DCMI guidelines for implementation encodings (DCMI-ENCODINGS) (Baker & Coyle 2009). The metadata application profile is a concept that has been evolving throughout the years, however, a clear definition of that concept provided by (National Information Standards Organization- NISO (2007) is how metadata elements from one or more metadata schemas combine and fit to describe a specific set of items (Andrade & Baptista, 2015).

Mapping is another way to achieve interoperability. Mapping helps to solve the problem of the growth of new bibliographic schemas that are using different vocabularies and creating gaps in the metadata environment (Hillmann, Dunsire & Phipps, 2013). Crosswalk is a result of mapping that explains the process of comparing metadata elements from one metadata schema to another (Woodley, 2016). Metadata managers must be aware of the possibility of losing some depth of information about materials during crosswalk, especially if they are cramming specific metadata from other schemas into a simplified/broad metadata schema such as Dublin Core (Park & Childress, 2009). There are several metadata crosswalk standards that have been used by different institutions and studies such as the NISO document Issues in Crosswalking Content Metadata Standards (Pierre & LaPlant, 1998). This document suggested that common properties of a crosswalk may include a semantic definition of each metadata element and proposed four main Crosswalking issues, including: Whether a field of metadata is compulsory or not; whether a metadata field may occur in the same record many times or not; restrictions due to the arrangement of elements of metadata relative to each other; limitations placed on an element's worth (Zavalina, 2019). Getty Research institute also provides a metadata standard crosswalk

that could minimize the risk of losing valued data during conversion[1]. This crosswalk allows metadata managers to create a clear record conversion plan before starting the process and includes: CDWA, CONA, CDWA lite, MARC, MODS, DC alongside seven other metadata schemas.

Mapping and crosswalks are used in metadata harvesting. It is important to point out the fact of having several protocols for metadata harvesting based on the metadata schema used. For example, Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) that uses simple Dublin Core (DC) metadata schema (Andrade & Baptista, 2015). JSON-API is another protocol that has been used by some data providers and is based on JavaScript Object Notation (JSON) schema (Agocs & Goff, 2018). By using metadata harvesting protocols, metadata is collected from different databases into one single database with links for each individual record back to their home database (Woodley, 2016). In addition, data providers have the option to decide to enrich the metadata record they are planning to harvest by choosing the metadata element set they want to use (Woodley, 2016). Repositories who are using metadata schemas that are richer than Dublin Core—i.e., CDWA Lite, MARC XML, MODS—need to map their content to unqualified DC metadata schema to conform to the harvesting protocol (Woodley, 2016). However, it is important to know there are four potential problems with harvested metadata, identified by Dushay and Hillmann: missing data, incorrect data, confusing data, and insufficient data (Dushay & Hillmann, 2003).

There are many different metadata harvesting projects conducted all over the world using different protocols and metadata schemas. For example, National Science Digital Library

[1] More information can be found here: http://www.getty.edu/research/publications/electronic_publications/intro-metadata/crosswalks.html

(NSDL) is a good example of a project which implemented metadata records that were harvested from various digital collections in different institutions (Zeng & Chan, 2006). Another well-known example is Europeana which is a large-scale aggregation created in 2008 containing millions of metadata records for books, paintings, films, museum objects and archival records from different institutions in Europe (Isaac & Haslhofer 2013). In 2010, the Digital Public Library of America (DPLA) was created with the support of over 150 partners that included public libraries, local museums, and large universities that shared and provided metadata records to users using one-point access (Gregory & Williams, 2014).

Several studies examine the quality of metadata in aggregations that rely on harvesting. Bruce and Hillman's metadata quality framework emerged from the evaluation of metadata harvested into NSDL (Bruce & Hillman, 2004). Data providers are encouraged by researchers to make available written crosswalking and metadata creation guidelines to provide appropriate information for interpreting and analyzing their metadata records. According to Jackson, Han, Groetsch, Mustafoff & Cole (2008) this practice would enhance metadata harvesting for aggregated projects. Development of metadata creation guidelines is an important step in being prepared to participate in aggregation projects. For example, the Portal to Texas History is a statewide aggregation that includes UNT Libraries collections such as Texas Patents[2]. Metadata guidelines for Texas Patents and other UNT digital collections are based on the Portal to Texas History metadata guidelines but provide details specific to those collections. However, institutions do not have to create a local metadata creation guideline if they are using more standard metadata schemes (Dublin Core, MODS, VRA Core 4.0., MARC 21) because the standard metadata creation guidelines for those metadata schemes already exist: user guides,

[2] More information can be found here: https://texashistory.unt.edu/explore/collections/TXPT/

RDA and more. Following those metadata standards and metadata creation guidelines would make the metadata records shareable/harvestable, which could help aggregators better interpret harvested metadata (Shreeves, Knutson, Stvilia, Palmer, Twidale & Cole, 2005).

Unfortunately, there are no such large-scale aggregations of digital content and metadata repositories (such as DPLA or Europeana) for the Arabian Gulf countries or the Arab countries. In July 2020, the international collaborative project that includes Qatar National Library, Stanford Libraries, and other partners, released an open platform of the Digital Library of the Middle East (DLME)[3]. This project started collecting items and metadata in the beginning of 2021 from different institutions all over the world. There are over 141 thousand metadata records representing items in DLME aggregation. Both metadata records and end-user interface of DLME are available in two languages: Arabic and English. Several large international institutions participated in this project: British Library, Cambridge University library, University of Michigan and many others. As for the Arabian contributors, Palestine Poster Project Archive[4] is one of the most active contributors as they shared over 12 thousand metadata record. Another major Arabic contributor to DLME aggregation is Qatar National Library, with more than 17 thousand metadata records.

Union catalogs is a traditional way to aggregate library metadata that has been used since before protocols for digital library metadata harvesting (e.g., OAI-PMH) were developed. A well-known example of a union catalog is WorldCat (the OCLC Online Union Catalog), which was launched in 1998 and aggregated millions of MARC records (Bennett, Lavoie & O'neill, 2003). The WorldCat catalog is currently used by over 15,000 libraries from 107 countries all

[3] More information can be found here: https://www.clir.org/2020/07/clir-and-stanford-libraries-announce-digital-library-of-the-middle-east-platform/ ;https://dlmenetwork.org/library

[4] More information can be found: https://www.palestineposterproject.org/

over the world[5]. MARC records can be aggregated and harvested using Z39.50, which is an international standard communication protocol for retrieving information created in the 1970s (Lynch, 1997). In 2007 Saudi Arabia founded a nonprofit project called the "Arabic Union Catalog," which aims to support the Arabic content by establishing a collaborative environment to collect and share Arabic metadata records from all over the world (Al-Nawas, 2017). This project included non-Arabic countries such as United Kingdom, United States, Argentina, and many other countries that have Arabic metadata records using MARC 21 schema.[6]

2.8 Organization of Knowledge Over the World

It is important to examine how the organization of knowledge takes place all over the world. In Italy, in 1869, the idea of having a general inventory—a title/author catalog and subject catalog for each library—was suggested by the Commissioner Cibrario (Bianchini & Guerrini, 2007). In 1922, the first descriptive catalog was established in Italy that was built based on a cataloging code similar to the Anglo-American code used at that time elsewhere (Bianchini & Guerrini, 2007). This descriptive catalog was created and evaluated by a group of members, including the heads of two honored libraries, a supervisor from the ministry and a university professor (Bianchini & Guerrini, 2007). Creation of the European Union in 1992 affected the shape of Italian institutions, including libraries, with more focus on integration in international efforts. For example, Dewey classification standards and IFLA documentation were translated into Italian in 1997, and thus Italy had taken a part in international meetings working on international cataloguing principles (Bianchini & Guerrini, 2007). This shows that even though Italy created its own cataloging code, they were willing to work on modifying this code to meet

[5] More information can be found here: https://www.worldcat.org/

[6] More formation can be found in the official website of the Arabic Union Catalog at: https://www.aruc.org/home

the international code of cataloging. Another example of such integration in international cataloging processes is in Malaysian libraries, which had used AACR2 for more than 30 years, but eventually replaced it with RDA. A study conducted by Mansor & Ramdzan (2014) to examine the awareness of Malaysian librarians regarding RDA put out a survey among 128 librarians who work in 51 higher education institutions. While only 43 responses were received and only 41 of them were useful for analysis, the study revealed that many of the librarians were aware of RDA but expressed the need for more practice to learn new rules and use RDA (Mansor & Ramdzan, 2014).

Major metadata schemas that were originally developed in North America had been adopted worldwide for the development of digital libraries. For example, metadata object description schema (MODS) is used in harvesting projects in both the United States and Australia (Guenther, 2004).

In the United States, descriptive cataloging rules are governed by an international committee of representatives from the English-speaking cataloging worlds. However, the domain of descriptive cataloging rules' physical terms were defined by the Library of Congress (De Groat, 2015). In Slovenia, the majority of libraries use the same software as the national union catalog COBIB (Budanović & Žumer, 2018). Around 300 libraries in Slovenia use the COBISS/3 cataloging for creating their bibliographic records (Budanović & Žumer, 2018). The study uses both observation and screen recording to examine catalogers while they do their job (Budanović & Žumer, 2018). The study scope was investigating how librarians (specifically catalogers) can describe an item without using cataloging rules, principals, and standards, as shown in Figure 2.3. The study invited 46 librarians from 11 different libraries; 32 librarians participated. An expert interviewed as part of this study stated that Slovenian catalogers are

aware of the changes happening in the organization of knowledge, especially relating to technology, and confirmed that they are ready to experience more changes in the future as the Slovenian cataloging follows international trends (Budanović & Žumer, 2018).

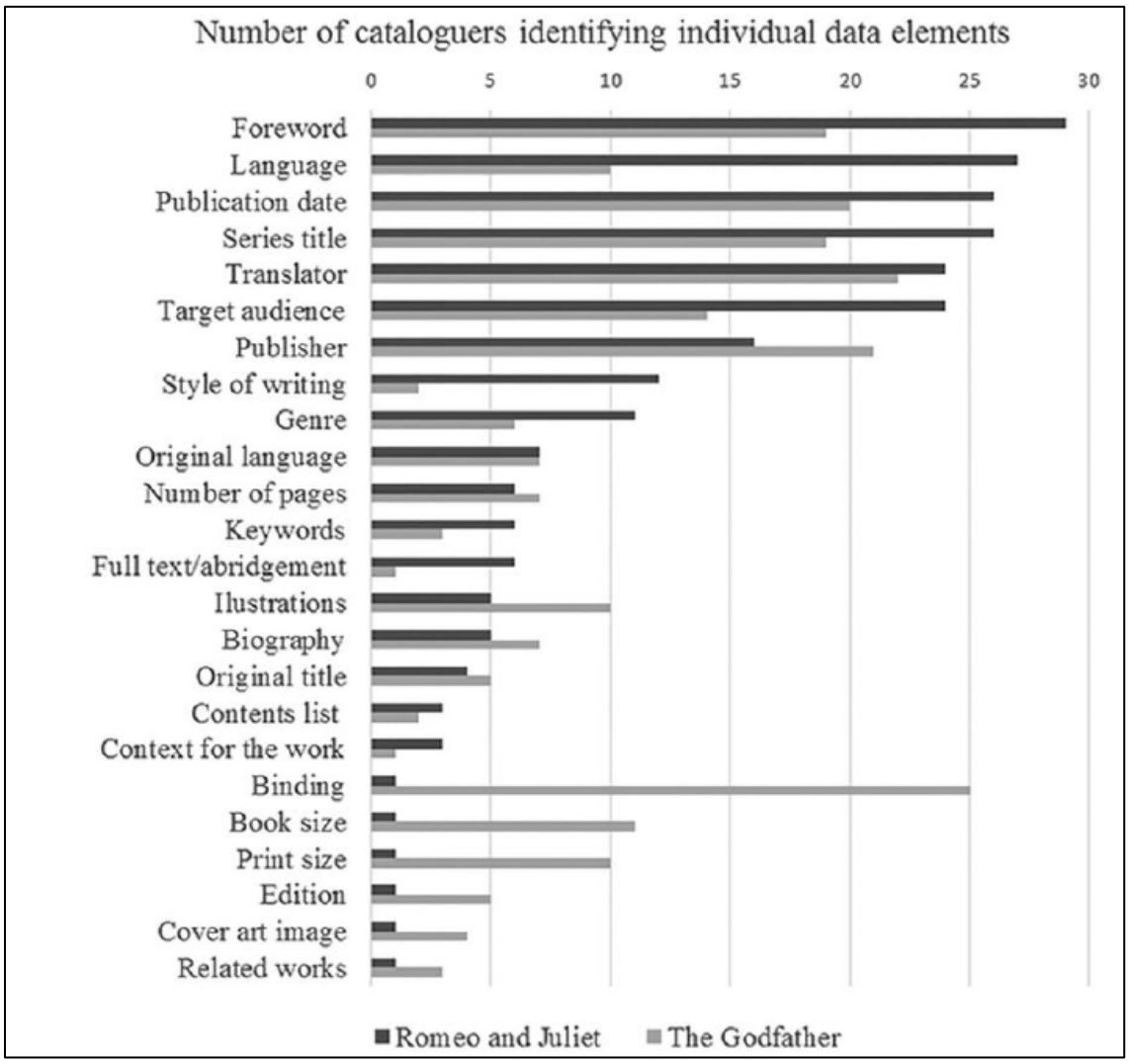


Figure 2.3: Catalogers identifying individual data elements (from Budanović & Žumer, 2018)

Another study conducted by Burke and Zavalina interviewed archivists in order to know more about the use of metadata schema in their archives, and what controlled vocabulary is being used (Burke & Zavalina, 2020). The researchers were able to interview seven archivists using the Zoom video-conferencing tool (Burke & Zavalina, 2020). One of the preliminary findings of this research states that archivists are suffering from the lack of funding they received, which effects

the quality of their service and the maintenance of their archive (Burke & Zavalina, 2020). Another study conducted by Park and Tosaka in 2015 used e-mail interviews to examine the RDA Implementation and Training Issues that academic libraries in United States faced at the time. The researchers were able to reach 12 participants, with an additional ten as a backup. The email interviews helped researchers to gather deep and well-organized information from participants as there were able to write their answers comfortably (Park & Tosaka, 2015).

2.9 Organization of Knowledge in the Arab and Arabian Gulf Countries

According to the Qatar digital website, digital libraries started in the Arabian Gulf countries in 2012 after Qatar launched their original digital library. Kuwait University and the University of Dubai started their digital libraries around that time also. Qatar represents a great example of how the Arabian Gulf countries are interested in creating digital libraries and sharing their knowledge with other countries. Sheikha Mozah Bint Nasser al-Missned, wife of the emir of Qatar and UNESCO special envoy for basic and higher education, believes that different angles and dimensions can be provided via the documents that could make the world understand Arab culture very well (Bade, 2010). Many of the libraries and museums in the Arab countries benefit from using international systems; for example, Jordan Library and Information Association (JLA) adopted Dewey classification system in the1970s and modified it many times to meet their needs (Eid, 2019). The functional requirements for bibliographic records (FRBR) was translated into Arabic by King Fahad National Library (KFNL) in Saudi Arabia (Eid, 2019). Qatar National Library (QNL) considered converting their catalog into Linked Data in 2016 (Eid, 2019). Linked data is a term which describes the best set of techniques for establishing, creating and connecting data on the web (Bizer, Heath & Berners-Lee, 2011). Egyptian Organization for Standardization and Quality Control (EOS) is an organization that helped translate many of ISO

standards in the library field since 1957 (Eid, 2019). The Egyptian national library is the oldest national library in the Middle East area, which was founded in 1870, recently they opened several branches in Cairo and they all use Dewey Decimal classification system (Holloway, 1959)

Despite these modifications over the years, there has been no research done about the how information is organized. This includes metadata in the Arabian Gulf countries' digital libraries. A study conducted in 2002 shows that many of the national libraries use Dublin Core metadata standards (DCMES 1.1) for describing their digital resources (Guinchard, 2002). Using Dublin Core would similarly make it easier for the Arabian Gulf countries to make their digital libraries more international and easier to use for users from around the world.

Metadata can be embedded in the digital object itself (for example a page on a website as <meta> tags) in HTML source code. Metadata records can also be encoded in eXtensible Markup Language (XML). Resource Description Framework (RDF), JSON and other syntaxes. Most of these national libraries use Hypertext Markup Language (HTML) to encode Dublin Core, while a smaller number of libraries use eXtensible Markup Language (XML), and fewer still use MARC as the encoding format (Guinchard, 2002). Iran digital libraries, however, do not use any special metadata standards such as Dublin Core (Ravandi & Zeinolabedini, 2013). Because Iran's digital libraries are very similar to the Arabian Gulf countries' digital libraries, Arabian Gulf countries' digital libraries might experience the same problem. The sample size of that study was eight digital libraries out of 110 sites. Ravandi and Zeinolabedini eliminated the 102 sites for many reasons; they find it difficult to work with because of its lack of accessibility, of adaption with digital library criteria, and of available full text resources, among many other reasons. Ravandi and Zeinolabedini used content analysis methods to conduct their study, where

they considered metatags in digital libraries websites and then determined the adoption rate of these elements with Dublin Core metadata elements using one adapting checklist of "HTML" language meta tags with Dublin Core elements. This is shown in Figures 2.4 and 2.5.

Library Name	HTTP-Equive Metatag	Meta name Attribute	Frequency
Omid-e Iran	Content-Type, Expire, Rating	Description, Author, Copyright, Revisit, Title, Relation Robot, Keyword	11
Yassobeldin	Cache control, Content, Progma, Type	Title, Description, Robot-Revisit	8
Ahlolbayt	Content-Type, Progid	Title, Description, Robot-Orginator	6
Parstak	Language, Content-Type	Title, Keyword, Author, Description	6
Downloadbook	----------	Robot, Author, Description, Keyword, Revisit	5
Tebyan	----------	Tebyan-Keyword-Copyright-Relation-Verify	5
Hawzah	Language-Progid	Title	3
DID	----------	Title	1

Figure 2.4: Metadata elements in digital libraries in Iran (from Ravandi & Zeinolabedini, 2013)

No.	HTML Elements	Dublin Core Elements	Frequency	Percent
1	Title	Title	7	87.5
2	Description	Description	5	62.5
3	Keyword, Classification, Subject	Subject	5	62.5
4	Expire, Rivisit	Date	4	50
5	Creator	Creator	3	37.5
6	Copyright	Rights	2	25
7	Relation	Relation	2	25
8	Language	Language	2	25
9	Sourse Type	Type	-------	-------
10	Source	Source	-------	-------
11	Format	Format	-------	-------
12	Electronic Address	Identifier	-------	-------
13	Contributor	Contributor	-------	-------
14	Coverage	Coverage	-------	-------
15	Publisher	Publisher	-------	-------

Figure 2.5: Dublin Core metadata elements usage in Iran digital libraries (from Ravandi & Zeinolabedini, 2013)

Another related study shows that there are different types of metadata elements used in the central library website of the Ministry of Health in Iran (Zare-Farashbandi, Ramezan-Shirazi, Ashrafi-Rizi & Nouri, 2014). This research uses the data collection tool of a researcher-formed checklist that lists the usage of HTML meta tags in a set of elements such as keywords, description, date and format. Zare-Farashbandi, Ramezan-Shirazi, Ashrafi-Rizi & Nouri, collected data by observing and visiting the websites first, then information objects, and finally

their metadata. Data analysis was prepared by Microsoft Excel software, 2011. The sample size of that study was 98 websites in Iran, out of the total126 websites. The study eliminated 27 websites because they do not have Central Libraries' Websites. Iran is a very close country geographically to Kuwait and Qatar, so it is interesting to see how knowledge is organized in their libraries' websites.

Another study revealed that web page designers (including designers of digital libraries websites) in Iran tend to use the keywords "meta tag" in the HTML source code of web pages more than the description, meta tag, because they are not highly aware of the structure and function of meta tags (Alimohammadi, 2004). This research used several research methods including the documentary method and survey method to review the statistical sample of the research.

Collections of cultural heritage institutions in Kuwait were critically damaged after the Gulf war in 1990. Iraqi troops plundered thousands of valuable items from libraries, museums, and archives in Kuwait during the war (Montgomery, 2015). However, some archives managed to survive these attacks because of their back up systems. While new archives were created after that war, studies of Kuwait archives were critically limited as most of them had been distorted or stolen. The main focus for libraries and archives in the Middle East in the years after the Gulf war was to obtain a disaster management plan to prevent their collections from being lost (Moustafa, 2013). In the Arabian Gulf countries, recording oral history can be a main source of creating new archives, as for example in Qatar (Ahmed, 2018). Because of this, it is interesting to examine the status of Alqabas 'archive, as it is one the most active and valuable archives in both Kuwait and Arabian Gulf countries.

As shown by this literature review, there is a lack of studies on information organization

in Arabian Gulf counties. More specifically, there are few studies focused on the creation or adoption of metadata in cultural heritage institutions, how users' needs are met, and how metadata can be made interoperable and aggregated in a central point of access to collections of cultural heritage institutions. According to Robinson, e-government, Arabic language, and information retrieval are the most common research topics of articles written by Arabs in LIS field and published in top journals (Robinson, 2016).

In this study, I will fill this gap to provide the necessary understanding of the state of information organization and knowledge representation in the cultural heritage institutions in the Arabian Gulf countries. The study will use interviews with the metadata managers in libraries and museums and content analysis of their metadata.

CHAPTER 3

METHODS

The proposed research methods for this study includes a combination of interviews with metadata managers and content analysis of metadata records in these cultural heritage institutions. The semi-structured interview guide was developed in two versions–an English-language version and an Arabic-language version. This was tested in the pilot study and updated based on the pilot study results (Appendix A). As part of the pilot study, the procedure for content analysis of metadata records was developed and tested.

3.1 Pilot Study

The interview guide was tested by a pilot study that I conducted with the employee of Alqabas archive, a Kuwaiti archive that exists since 1970s and contains a large collection of 15-thousand books, four million articles and two million images. The pilot study was conducted in February 2020 and used Zoom software to interview the participant, who is the head of the information and studies department in Alqabas' archive. Their role is to make decisions about information organization within the collection. The interview questions in both languages (Arabic and English) were sent to the participant two hours before the interview was conducted. The interview took around 30 minutes and was in the Arabic language. An advantage of this data collection approach was that Zoom provided a great communication channel with the interviewee that allowed meto ask all interview questions and collect the answers. However, the automatically generated textual transcript of the interview was unclear because of the language; Zoom failed providing acceptable textual transcript of the interview, which resulted in the need for manual transcription of the interview recording by me. In addition to transcribing, I performed translation of the interview in English. The answers for the interview were completed

and the interviewee was able to understand most of the concepts in the questions. However, some of the terms and concepts (e.g., metadata harvesting) were not clear to the interviewee as she needed a little bit of explanation. As a result of that observation made during pilot interview, the decision was made to attach definition to the interview question which contains the term "metadata harvesting" in the main study to avoid confusion of the main study participants.

The findings of the interview can be summarized in a few points: the archive uses a locally developed metadata scheme created to meet the needs of both users and employees; the archive uses Structured Query Language SQL to manage their collection; Maknaz thesaurus is used as a controlled vocabulary that includes name authority file and subject headings; there are no training workshops or outreach programs offered to metadata creators and managers in Alqabas organization. Access to archive items and metadata is mostly restricted to the users who visit the Alqabas building. Users' needs are not considered in the decision-making process, as most of the decisions were made based on the organizations and its employees' needs.

As Alqabas metadata records are not available on the Alqabas website, after conducting the interview, I collected from the interviewee the metadata records for ten items in Alqabas archive for the purposes of testing the study methodology and refining the content analysis instrument. At the time of conducting this study, Alqabas archive did not provide online access to its collection so users can only access their archive locally. However, with the COVID-19 pandemic surge and resulting quarantine measures implemented in March 2020, Alqabas no longer accepts visitors. COVID-19 is a disease that spread across the world in 2020. The metadata records were collected as screenshots sent by interviewee over email to me and following the sampling procedure that I explained to interviewee (described below). The ten records that represent eight textual objects and two images were selected based on a specific

topic: “Arabian Gulf War.”

The initial plan was to obtain more records from interviewee at a later date, for a total of 50., However, after collecting the initial 10 records, the connection between me and the interviewee was lost (possibly due to the interviewee’s organization being closed for quarantine in response to COVID-19 crisis). Despite the very small size of the resulting sample of records, analysis of these ten records revealed interesting findings presented below.

The collected metadata records were analyzed for answers to research questions of this study. Analysis of metadata records as part of this pilot study allowed to develop and refine the measures to evaluate and to configure the metadata record analysis tool (Excel spreadsheet) that will be used in the main study.

The content analysis of the records included:

- Examination of the level of application of metadata fields and the lengths of data values (with central tendency and variability measures)
- In-depth qualitative comparative analysis of each field in the records based on major metadata quality criteria of accuracy, completeness, and consistency.

A total of 15 metadata fields were observed in those ten records: Identifier, Date (Gregorian), Date (Islamic), Source, Type, Format, Relation, Rights, Creator, Title, Description, Descriptors, Coverage and Contributor. Other fields that according to the interviewee are included in the Alqabas system metadata forms and end-user display of metadata records but are never used by metadata creators were excluded from analysis. Contributor field was found to not be included in any of the records in the sample. As shown in Table 3.1, the average percentage of records with the field across these 14 fields was 92%.

Analysis of central tendency measures for the length of the data values (in characters without spaces) revealed the highest median data value length in Descriptors field (91) and the lowest (0) in Coverage, Rights, Relation,and Date (Islamic) as shown in Table 3.1. The highest

mode (70) is also in Descriptors and the lowest (0) is in Description, Creator, Coverage, Rights, Relation and Date (Islamic) as shown in Table 3.1. The highest average of data value length is in Descriptors (110) and the lowest is in Rights, Relation (0) as shown in Table 3.1. The lowest level of application was observed for Topic, Relation and Coverage fields (10% of records), and the highest for Identifier and date (Gregorian) fields (100% records each).

Table 3.1: Pilot study findings: Quantitative Indicators

	Highest	Lowest
% of records with the fields: Presence	Identifier & Date (Gregorian): 100%	Relation, Topic & Coverage: 10%
Data value length in characters without spaces: MEDIAN	Descriptors: 91	Coverage, Rights, Relation & Date (Islamic): 0
Data value length in characters without spaces: MODE	Descriptors: 70	Description, Creator, Coverage, Rights, Relation & Date (Islamic): 0
Data value length in characters without spaces: Average	Descriptors: 110	Coverage 1.56, Rights 0.89 & Relation 0.56
Data value length in characters without spaces: Variance	Description: 25911	Date (Gregorian) 0.19, Format 1.61, Type 2.53, Relation 2.87, Rights 3.36
Data value length in characters without spaces: Standard deviation	Description: 161	Date (Gregorian) 0.44

Variability measures for the length of the data values (in characters without spaces) were also assessed. As shown in Table 3.1, the highest variance was observed for Description field (25911) and the lowest was Date (Gregorian) field (0). The highest standard deviation was the highest in Description (161) and the lowest in Date (Gregorian) field (0).

Content analysis of records showed that there are three metadata element sets and orders for records in Alqabas archive. The eight records that represent textual objects used the same metadata element set and formatting of data values in their metadata fields. However, the metadata records representing images looked differently. The first record that represents an

image (out of two collected by researcher) was entered to the system and had less metadata elements as it did in the records that represent textual objects, and they also appeared in a different order. For example, that record (representing image does not include the following fields: Date (Islamic), Rights, Creator, Type, Format, Coverage and Relation. However, it has a Topic field which was not present in the other records. However, the other metadata record that represented historical image was found to rely on a completely different metadata element set. In this image record, Weekday of Publication, Size of the image, Coverage, Page Number elements that were not present in records for textual objects were included but all these fields were empty expect Coverage field. On the other hand, Title, Description, and Descriptors fields common for other records in this sample were not included in this record representing historical image. One possible explanation for such a difference observed in this record could be that it was created much earlier than the other 9 records in the sample and was a part of an old digitizing project and that relied on a different metadata element set that is no longer used in the archive.

A follow-up interview is needed to provide better understanding of the patterns observed in the analysis of these ten metadata records. However, so far, my multiple attempts to get in contact with the interviewee after the interview have been unsuccessful.

Another interesting finding was the pattern of including the subject headings in the Descriptors field instead of the Subject/Authority headings field in all records in the sample. Ninety percent of records had subject headings that follow the Expanded Thesaurus (Maknaz). Descriptors field in Alqabas Archive contains only subject headings. According to Dr. Yasser Abdul-Mottey from Public Authority for Applied Education and Training (a Kuwaiti University), it is common in the Middle East to use Descriptor's metadata field for subject headings. In Alqabas cause, subject headings were presented in a high number in the eight records that

represent textual content: a total of 36. The highest number of subject headings (8) was observed in one record; the lowest number (3) was observed in one record w.

Analysis of data values in two date fields—Date (Gregorian) and Date (Islamic)—revealed several findings. First, the relatively low level of application of two formats for the date information was observed; it is common for databases and archives in the Middle East to include both Gregorian and Islamic date formats, however, in the small sample of records analyzed as part of this pilot study, only 40% included both dates. The other 60% of records included only the Gregorian dates. It looks like the Gregorian date is a required field, while the Islamic date might be an optional field. Further research is needed (e.g., examination of metadata creation guidelines of Alqabas archive, if any, and/or follow-up interview with metadata manager) to verify this conclusion. Second, the formatting of the Islamic date was not the same in the analyzed records. In 20% of cases, the Date (Islamic) metadata field contained only numbers, but in the other 80% of records it had words for the months instead of numbers. This discrepancy indicates that there is no clear policy on formatting the dates in the Alqabas metadata records.

Another interesting finding of the content analysis concerns the data values in Creator fields. Based on the answers provided by the interviewee, Alqabas archive uses name authority headings from the Expanded Thesaurus (Maknaz) for representing names of authors or creators. However, some name repetitions were observed in the Creator field: the author's name was entered twice instead of once in 10% of records. Also, the creator's names did not always follow a consistent format: one of the creator's names included the previous job (retired minister), which is not the format in which names are formatted in Maknaz (possibly because there was no authority record for this person in Maknaz). In one additional record out of five that included non-empty Creator fields, the name of the creator was repeated twice. Further research is needed

to determine the source of these inconsistencies when entering the author or creator name: is it the metadata creation practices at Alqabas archive changing over time or the lack of metadata creation guidelines?

Another problem observed in this analysis was using the Source field for entering the publisher's name and location. It was unclear why this was done as it would be more appropriate for the records to include both the Source field with an accurate data value and the Publisher field. However, this pattern of application of the Source metadata field might be due to the way this metadata element is defined in the Almaqnaz archive which could potentially be different than in Dublin Core. The Almaqnaz archive does not publish documentation for its metadata scheme so it was not possible to verify that assumption.

Also, Language field was missing in the metadata records which could be explained by the fact that the archive currently provides local access only. However, does not mean all the users in Kuwait can read Arabic. Language field should be included to help users who do not read Arabic to determine the language of the information object it represents. Moreover, the omission of Language field could become a serious obstacle for aggregating metadata from institutions like Almaqnaz into a centralized Arab Gulf countries repository similar to DPLA or Europeana as multiple languages are spoken in the region.

An interesting finding was observing a non-machine-readable record with handwritten data values which possibly indicates the need for retrospective conversion of older analog records into machine-readable records before any aggregation efforts and metadata harvesting to a centralized portal. It is important to know what proportion of metadata records are currently available only in analog form and what plans are in place to convert those records into machine-readable form.

Finally, I observed redundancy in data values in different fields. It was found that 40% of records repeat the title of the information object in the Description field. Also, records for textual objects were found to use more than one format of data values in the Date (Islamic) and Creator fields, as mentioned above.

Overall consideration of the three major metadata quality criteria, including completeness, accuracy and consistency, revealed the following:

- Records in the sample were mostly complete, however, the lack of completeness was observed in the Relation, Rights, and Coverage fields that were included in 0%, 30%, 20%, and 10% of records in the sample respectively. Also, Creator field was skipped in 50% of the records.

- Accuracy was low overall. This was evident in Description filed (50% of records) where half of the data values were repeated and in Source field which used publisher name and country in 100% of records in the sample. Another accuracy problem was inclusion of title as a data value in Description field.

- High consistency was observed among the records using the same metadata scheme (records for textual objects). Most of the data values in the fields of these eight records followed the same formatting. However, lack of consistency was observed in the fields of Creator and Date (Islamic).

The overall conclusion based on the analysis of ten metadata records from Alqabas archive is that they are retrievable through both title search and, more importantly, subject search. All records included Title field data value, and the level of use of subject headings was high. However, users will need to check the actual textual documents or image object to make sure if that is what they are looking for, as the data values in some fields (e.g., Description) are not completely accurate. Metadata quality problems observed in this analysis might stem from several causes, the most important one being, according to the interviewee, that none of the Alqabas archive employees responsible for metadata have a library or information science degree. I believe the lack of metadata training manifested itself in the quality of metadata records. Another possible reason is that the Alqabas archive lacks established metadata creation

guidelines documentation that all the employees can follow when creating metadata records. Lastly, I believe the lack of feedback from users and professionals leads employees to be less careful regarding following specific rules for creating metadata.

The pilot study described above provided useful information that can inform the design of the proposed dissertation research study. For example, a general thesaurus such as Maknaz with a controlled vocabulary of subject terms, names, and genre terms (El-Sherbini, 2015) is widely used in the metadata records in the Middle East. More questions should be asked in the interview about using Maknaz. Another important lesson learned in the pilot study is the need to ask the interviewees about the possibility of conducting a follow-up interview (if needed) after examining the metadata records of their collection. Lastly, it is important to collect a balanced number of records that represent textual objects and images for content analysis: this will help build an understanding of the differences and similarities between the metadata application in description of textual objects and images in the bibliographic database.

This pilot study was conducted before COVID-19 spread worldwide. Contacting and interviewing the Alqabas manager over Zoom was easy at that point, and I was waiting for the Institutional Review Board (IRB) approval to start collecting data and conduct the interviews over Zoom for this dissertation research. However, after a few weeks the disease took control of almost all the countries in the Arabian Gulf. All public places were closed, as well as schools and universities. In Kuwait, the government announced a full ban, meaning no one could leave his/her home. With this big change, conducting the interview via Zoom became inefficient, as potential respondents may not be able to participate in a continued interview for one hour: they might be busy taking care of their children or home. As a result of that, the decision was made to change data collection for the proposed study to online interviews conducted via email. Email

interviews allow participants to find time in their schedule to provide more thoughtful, reflective responses to the interview questions (Meho, 2006). Previous studies by other researchers on metadata-related topics relied on email interviews and found this data collection approach effective and resulting in relatively high response rate (e.g., Park & Tosaka, 2015).

This data collection method will be applied to the participants of this study as the interview to compensate for challenges caused by the COVID-19 pandemic. I believe this is the best way to reach the participants in this situation. More explanation can be found in the next section.

3.2 The Effect of COVID-19 on the Study

COVID-19 reached the Arabian gulf countries in February 2020, starting in Kuwait then moving to all other countries. The governments took many actions to stop this disease from spreading. Kuwait, Qatar and Oman closed all the schools, universities, museums and many other cultural heritage institutions. Kuwait announced that schools and universities would remain closed until October of 2020 (six months from now), and online learning would not be an option. Similar decisions were made among the Arabian gulf countries, but some of them agreed to provide online learning. However, all those decisions are not final as things might change according to COVID-19 developments. In the meantime, social distancing is highly demanded by governments and no one is allowed to go to their job unless they support the fight against COVID-19.

It is clear that COVID-19 affected the study methodology of this research as many of the employees will not be able to be in their offices to participate in interviews. This will affect the rate of responses as many of those employees may not be willing to do the online interview in their home on personal devices, or they might not have good internet communication skills. To

avoid missing some responses, I decided to change the data collection tool into the email interview rather than the recorded interview using Zoom. I believe an email interview provides a fair chance to all participants to respond anytime they prefer from their home without making any special preparations. The approval of conducting email interviews from University of North Texas Institutional Review Board (UNT IRB) was received on May-04-2020 (Appendix B). The participations will be contacted first with an email invitation letter in both languages (Arabic and English) to introduce the study and explain the importance of their participation in it (Appendices E and F). The participants also will receive an informed consent form that includes their rights as participants in this study in both languages (Appendices C and D). The form must be signed and collected by me for his records.

3.3 Methods of Research Proposed: Data Collection

In this study, I focus on three countries in the Arabian Gulf, which are Kuwait, Qatar and Oman. The reasons behind selecting these three countries include the following: information organization has not been researched in them before; good foreign economic relations between these counties and shared language make it possible to collaborate in centralized multinational aggregations; as a citizen of one of these countries and a graduate of the MLIS program in Kuwait, it will be easier for researcher to establish connections with participants of the study. Institution's selection criteria included:

- Has a bibliographic database.
- Must be in Kuwait, Oman or Qatar.
- The headquarter of the institution should be local (in one of those three countries).

I applied qualitative method because it works better with comparative studies. Interviews were used as one of the data collection methods. The assumption behind this was that

interviewing employees of libraries and museums who make decisions about information organization in their collections from those countries would reveal similarities and differences, as well as opportunities and challenges for providing access to their collections via a large-scale centralized portal similar to DPLA or Europeana in the United States and Europe.

These three countries have many libraries, museums and archives, but in this study I only interviewed participants in cultural heritage institutions that have bibliographic database. The collection could contain images (photographs, posters, drawings, paintings, etc.), and/or textual resources (books, articles in journals, magazines, or newspapers, letters, diaries, etc.).

Sampling included the following specific steps. First, researcher selected participants from different institutions in Kuwait, Oman, and Qatar, to be interviewed. The potential respondents were selected from the list of employees on their institutions' websites. I was able to interview 15 participants which ensured reaching the interview sample targeted with the response rate of 50%. I sent invitations to 30 potential respondents. The research sample was stratified by country and by institution type: I invited between 8 and 14 participants from each of the three countries, half of whom were representing libraries (national and academic) while the other half were representing museums. The sampling method that was used in this research is selective as I selected participants from cultural heritage institutions that have bibliographic databases and are responsible for information organization (including metadata). The interview recruitment email (Appendix E) was sent to participants in both languages (English and Arabic) to make sure that participants were able to understand the invitation based on their preferred language.

In the interview I asked specific questions related to information organization tools and approaches. The interview was semi-structured, with a set main questions prepared in advance (see interview guide - Appendix A) and follow-up questions as needed. Interviews were

conducted by email, which allowed participants to respond whenever they can, considering the COVID-19 situation. The interview questions were sent to the participants in both languages (English and Arabic) and the participants had the freedom to choose which of the two languages they want to respond in.

For triangulation purposes, I supplemented interviews with content analysis of a small sample of metadata records from all those cultural heritage institutions in Kuwait, Qatar and Oman, the employees of which participated in the interview. The sample size was 412 records total from 15 databases, 20-28 records which contain both textual works and images from each of the 15 institutions. The records were selected using random, stratified sampling: I searched in each collection by a specific topic, such as Gulf War, War, or something similar, and every fifth of the retrieved metadata records were selected for analysis until the sample was complete from each institution. The comparison process started with the search in the collection by using a specific topic.

3.4 Methods of Research Proposed: Data Analysis

I exported metadata records into an Excel spreadsheet that was used for data preparation and data analysis.

In the analysis of metadata records, I compared the metadata characteristics, such as the number of metadata elements used per record, the number of instances of each element per record, the number of subject headings, and the controlled vocabularies used in different metadata fields, etc. The emphasis of this exploratory analysis was on the three major criteria of metadata quality: accuracy, completeness and consistency of metadata records. Other quality criteria were not examined in this study. The analysis included possible barriers for metadata interoperability in large-scale portals that aggregate metadata from databases in libraries and

museums in Kuwait, Oman, and Qatar. For example, I evaluated in the institutions that used the same metadata schemes and controlled vocabularies applied them in the same way. Equivalent metadata elements used by participating institutions were determined.

As the interviews were conducted over email, I believe there is no need for coding/categorization scheme for structured data analysis because of the short number of participants and the very clear/straight questions used in the interview. For those two reasons, I believe it is more efficient to analyze data following the interview question by question without the need to use a special coding schema. This methodology had been tested and applied in metadata research in the past (e.g., Park and Tosaka, 2015) and it was assumed that it would work successfully with this study.

3.5 Research Limitations

There are three main limitations for this study. First, interviews were only conducted with participants in cultural heritage institutions that have bibliographic databases; Second, the study covered only three of the six countries from the Arabian Gulf region: Kuwait, Oman and Qatar. Additionally, because the sample of metadata records, as a supplementary data collection, was relatively small (412 records total from 15 databases) it is not statistically representative of all metadata records in these collections.

As for limitations inherent in the data collection method, it is well-known that the memory is one of the limitations for conducting interviews which rely on human's recollection of events or even faulty memory (e.g., as reviewed in Alshenqeeti, 2014). Another known limitation of interview is that respondents may have incomplete knowledge to address researchers' questions (e.g., Alshenqeeti, 2014). By conducting the interviews via email, the proposed study will address these limitations as respondents will have more flexibility and

freedom to respond and will have time to refresh their memory by consulting relevant documents. Another interview data collection limitation is that respondents might provide answers that they think are expected in order to make themselves or their organization appear better. This and other limitations are commonly addressed by combining two or more data collection methods – a practice known as method triangulation (e.g., Carter, Bryant-Lukosius, DiCenso, Blythe, & Neville, 2014). This study will rely on method triangulation by combining interviews with content analysis of metadata records. Method triangulation allows for deeper analysis of the collected data and helps ensure accuracy in the investigation.

The proposed study is exploratory in nature. It is expected to start filling the gaps in understanding of the knowledge organization in libraries, archives, and museums in Arab Gulf counties. Future studies will address research questions that are beyond the scope of this study.

CHAPTER 4

FINDINGS

I was able to conduct 15 interviews with 15 participants from 15 different institutions via email. I reached 9 participants (60% of the sample) from Kuwait, 3 (20% of the sample) from Qatar and 3 (20% of the sample) from Oman, as shown in Figure 4.1.

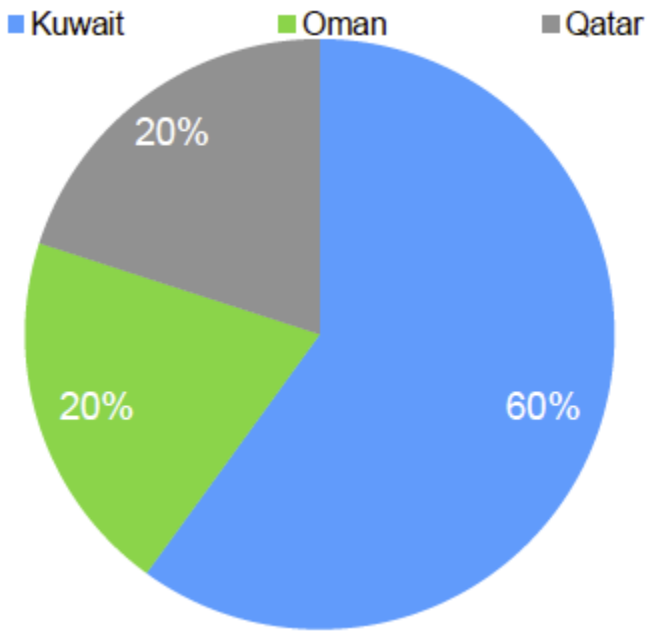


Figure 4.1: Participants by countries

Thirty email invitations were sent in total; there was a 50% response rate. A total of 412 metadata records were collected from databases of all participating institutions for content analysis. I picked "Gulf War" as a search topic to collect the metadata records. However, when the search result provided limited metadata records I used a broader search word— "War"—to collect the needed sample. The search was done by using the sample search with both Arabic and English languages. I focused on the "Title" field to ensure no duplication in the sample. The metadata records were collected and analyzed using an Excel spreadsheet as described in the methodology chapter. However, the administration fields and fields that have not been used were

excluded from content analysis in the whole samples. It is also important to mention that most of the participating institutions found to be using the same metadata fields in all records for both printed and electronic items. In other words, the metadata records provide data about the item along with the options to get either the electronic copy of it or place a request to reserve the physical copy. Some institutions listed metadata records alongside options to access either the electronic version or place a reservation for the printed copy. As shown in Figure 4.2, the 15 participant's institutions were divided between 11 libraries (73% of the sample), 3 archives (20% of the sample) and 1 museum (7% of the sample). The reason for getting a very low response rate from museums occurred because, in response to the coronavirus pandemic, most museums are either closed or have evacuated their technical employees to their original countries.

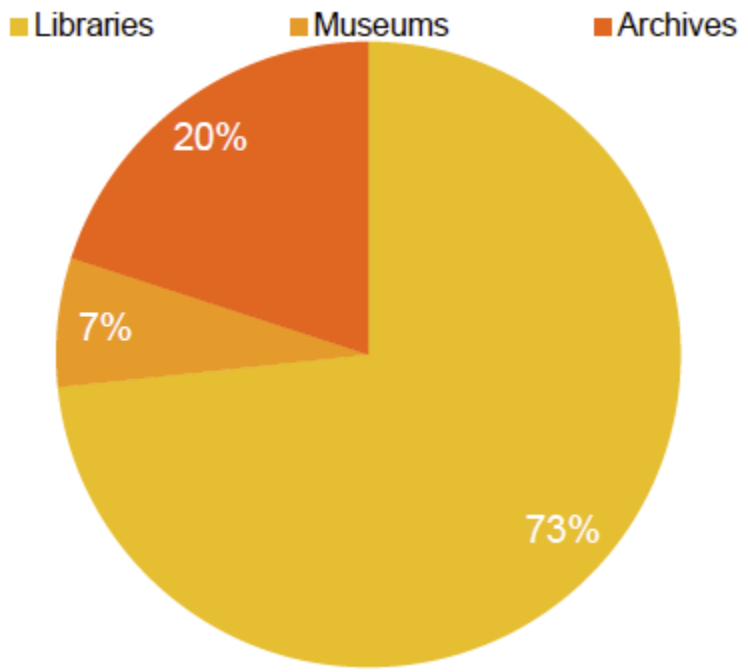


Figure 4.2: Participants by institutions type

All the personal information for participants and institutions are restricted to me; participants will be referred to by numbers 1 to 15, unrelated to their individual countries or institutions.

4.1 Participant 1

4.1.1 Interview Analysis

Interviews were conducted in the Arabic language and required translation into English by me to analyze the collected data from those interviews. Participant 1 holds a bachelor's degree in library and information science and began working at their institution in 2014; they took training courses before staring the job along with receiving mentoring from their supervisor. The institution started creating bibliographic databases in 2004 and it has more than 56,000 items with different types including e-books, e-journals, dissertations and more. The institution used to use the Virtua system for digital content management, however they switched to the Koha system one year before the interview. The metadata schema which is used in this institution is MARC 21 with RDA rules. The institution is using the Library of Congress classification system and the Library of Congress Subject Headings with controlled vocabulary because it works better for the institution's development. However, for some of the Arabic resources the institution uses the List of Standard Arabic Subject Headings[7]. Metadata records in this institution are currently available for copy and shared individually but may not be used for harvesting.

The cataloging department in this institution has three employees including the head of the department, who is responsible for checking metadata records and subject headings per their experience.[8] The institution follows mainly MARC polices and guidelines regarding the creation of metadata records, however, if something is not covered in those guidelines the institution tries

[7] QRMAK: Qa'imat Ru'us al-Mawdu'at al-Arabiyah al-Qiyasiyah lil-Maktabat wa-Marakiz al-Ma'lumat wa-Qawa'id al-Bayanat (List of Standard Arabic Subject Headings for Libraries, Information Centers and Databases); a special subject cataloging manual used for Arabic materials. More information can be found here: https://www.bibalex.org/libraries/presentation/static/156d0.aspx.

[8] The participant did not specify the educational background of their employees.

to follow other, bigger institutions' informal guidelines[9] in the same country.

The employees do not consider the users' need in the process of creating the metadata records, however, they do try to determine users' needs from the search log in the institution's catalog system. Feedback is obtained from operation departments such as the references desk, circulation department and some academic departments via emails or phone calls. Also, the institution gets in touch with users via phone calls, emails, workshops/orientations, and the annual survey. Usually, the person who is responsible for collection development is the one who contacts the users regarding their needs, however, sometimes catalogers also participate in the process.

The interaction options with the bibliographic database provided to the users are limited to conducting both simple and advanced searches, which includes Boolean search and limitation options such as type of documents, date, and more. Users have the option to download, browse and print metadata records. Searching and browsing options are available in both English and Arabic languages in the institution's collection. There are also tutorials that the institution provides which explains searching tips and introduces the users to the bibliographic database.

The participant appreciated the option of having a central portal that contains records from different cultural heritage organizations in Arabian Gulf Counties. However, some steps needed to be recognized before establishing that idea: contacting bigger libraries because they have more experience and collections; finding the administrators willing to establish and maintain this project; finding a better way to save the collections via clouds; setting up united guidelines for the libraries to make the harvesting process easier. On the other hand, the participant believes there are some barriers that this project might face, such as the uncooperative

[9] Not available as a document.

environment, different guidelines and polices, the fact that the Arabic content is still not completely available electronically, and the lack of technical support.

Overall, this institution adopted international systems and standards to organize their collections such as MARC 21, LC classification system and LCSH. The institution does not include the users' needs in making decisions, however, it has multiple interaction and channels to reach their users.

4.1.2 Content Analysis

I was not able to access the institution's bibliographic database due to technical error. However, the participant was willing to send a total of 20 metadata records in PDF format that represent textual items as shown in Figures 4.3 and 4.4. A total of 13 metadata fields were observed in the sample: Control Number (used for identifier), Library of Congress Call Number, Creator, Title, Edition, Publication, Physical Description, Type, Bibliography, Language, Subject (used for subject headings), Personal Name (used for contributor) and Summary (used for Description). Six metadata fields were found in 100% of records in this sample: Library of Congress Call Number, Title, Publication, Physical Description, Type and Subject. The lowest occurring field was Summary, found in only 15% of records in the sample. All records included more than one instance of two metadata fields: Subject and Personal Name. The average number of subject headings observed in the metadata records sampled from this institution's bibliographic database was 3.

Analysis of central tendency measures for the length of the data values (in characters without spaces) revealed the highest median data value length is in Subject field (72) and the lowest (0) in Control number, Bibliography, Language and Summary fields. The highest average of data value length is in Summary (77) and the lowest is in Type (2).

(Record no. 1930) : مبادئ ومسارات في الدرس اللغوي الحديث

[view plain]

000 -LEADER

fixed length control field 01586nam a2200349 i 4500

001 - CONTROL NUMBER

control field vtls000034378

003 - CONTROL NUMBER IDENTIFIER

control field UoNL

005 - DATE AND TIME OF LATEST TRANSACTION

control field 20200415132711.0

008 - FIXED-LENGTH DATA ELEMENTS--GENERAL INFORMATION

fixed length control field 171114s2009 ua a b 000 0 ara d

020 ## - INTERNATIONAL STANDARD BOOK NUMBER

International Standard Book Number 9773143627

Qualifying information (غلاف ورقي)

039 #9 - LEVEL OF BIBLIOGRAPHIC CONTROL AND CODING DETAIL [OBSOLETE]

Level of effort used to assign nonsubject heading access points marwa

Level of effort used to assign subject headings 201711141054

Level of effort used to assign classification marwa

040 ## - CATALOGING SOURCE

Original cataloging agency UoNL

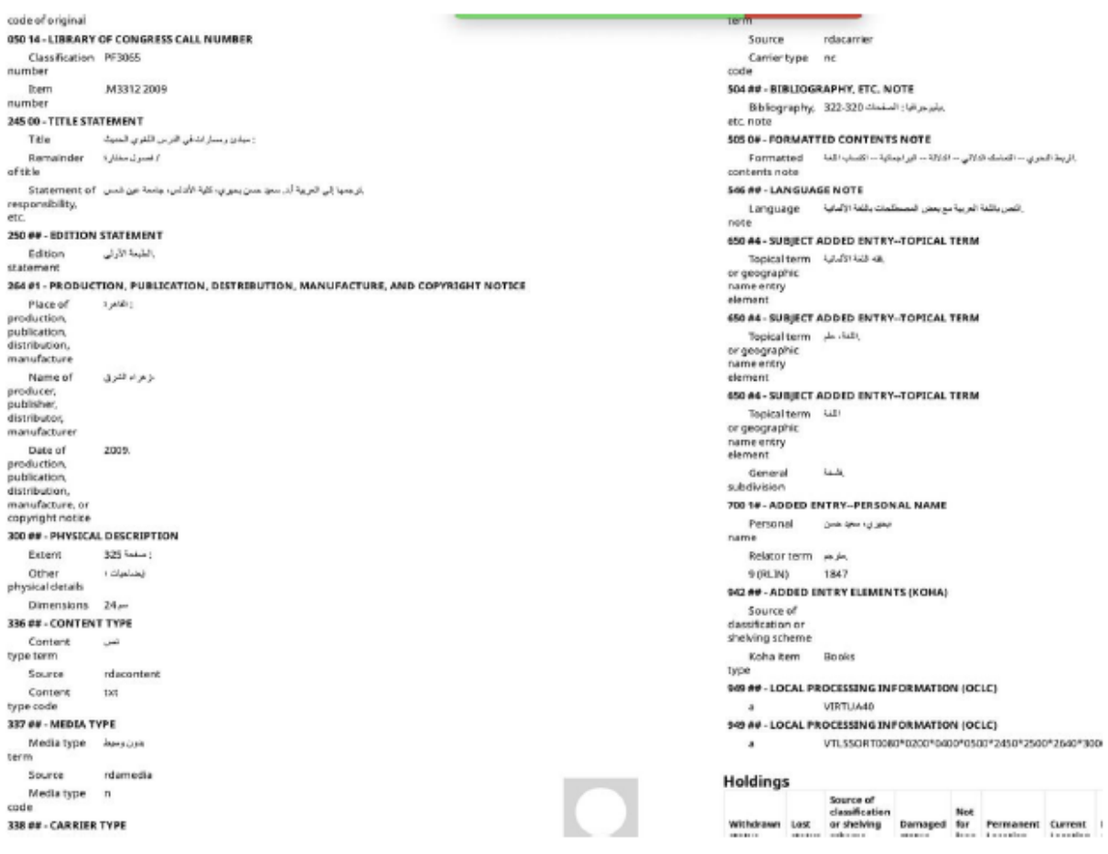
Language of cataloging ara

Description conventions rda

041 1# - LANGUAGE CODE

Language code of text/sound track or separate title ara

Figure 4.3: Example of a metadata record from Participant 1: textual item



code of original

050 14 - LIBRARY OF CONGRESS CALL NUMBER

Classification number P53065

Item number .M3312 2009

245 00 - TITLE STATEMENT

Title مبادئ ومسارات في الدرس اللغوي الحديث :

Remainder of title / فصول مختارة

Statement of responsibility, etc. ترجمها إلى العربية أ.د. سعيد حسن بحيري، كلية الآداب، جامعة عين شمس

250 ## - EDITION STATEMENT

Edition statement الطبعة الأولى.

264 #1 - PRODUCTION, PUBLICATION, DISTRIBUTION, MANUFACTURE, AND COPYRIGHT NOTICE

Place of production, publication, distribution, manufacture القاهرة :

Name of producer, publisher, distributor, manufacturer زهراء الشرق

Date of production, publication, distribution, manufacture, or copyright notice 2009.

300 ## - PHYSICAL DESCRIPTION

Extent 325 صفحة ;

Other physical details إيضاحيات ؛

Dimensions 24سم

336 ## - CONTENT TYPE

Content type term نص

Source rdacontent

Content type code txt

337 ## - MEDIA TYPE

Media type term دون وسيط

Source rdamedia

Media type code n

338 ## - CARRIER TYPE

term

Source rdacarrier

Carrier type code nc

504 ## - BIBLIOGRAPHY, ETC. NOTE

Bibliography, etc. note ببليوجرافيا : الصفحات 322-320.

505 0# - FORMATTED CONTENTS NOTE

Formatted contents note التربط الدلالي -- التماسك الدلالي -- الدلالة -- التراجمانية -- الكتساب اللغة.

546 ## - LANGUAGE NOTE

Language note النص باللغة العربية مع بعض المصطلحات باللغة الإنجليزية.

650 #4 - SUBJECT ADDED ENTRY--TOPICAL TERM

Topical term or geographic name entry element اللغة اللغة الألمانية.

650 #4 - SUBJECT ADDED ENTRY--TOPICAL TERM

Topical term or geographic name entry element اللغة، علم.

650 #4 - SUBJECT ADDED ENTRY--TOPICAL TERM

Topical term or geographic name entry element اللغة

General subdivision علمها.

700 1# - ADDED ENTRY--PERSONAL NAME

Personal name بحيري، سعيد حسن

Relator term مترجم.

9 (RLIN) 1847

942 ## - ADDED ENTRY ELEMENTS (KOHA)

Source of classification or shelving scheme

Koha item type Books

949 ## - LOCAL PROCESSING INFORMATION (OCLC)

a VIRTUA40

949 ## - LOCAL PROCESSING INFORMATION (OCLC)

a VTLSSORT0080*0200*0400*0500*2450*2500*2640*300

Holdings

Withdrawn	Lost	Source of classification or shelving	Damaged	Not for	Permanent	Current

Figure 4.4: Example of a metadata record from Participant 1: textual item

Variability measures for the length of the data values (in characters without spaces) were also assessed. The highest variance was observed for Summary field (at 39,745) and the lowest was Type field (at 1). The highest standard deviation was the highest in Summary (199) and the lowest in Type (1) and Library of Congress Call Number fields (2).

The MARC system allows catalogers to add additional entries as needed, which helped this institution to include all the contributors' names in every single item within the field "Personal Name Entry." The same applied for Subject Headings, which were listed in every record in Subject Entry following the LCSH format. According to the participant, LCSH applies to only the records that represent English items and have English data value as well. For records that represent Arabic items, the List of Standard Arabic Subject Headings for Libraries, Information Centers and Databases were applied.

Consideration of the three major metadata quality criteria—completeness, accuracy and consistency—revealed the following: Records in the sample were found to be mostly filled, but there was a lack of completeness in the appearance of fields Summary 15% and Language 35%; accuracy was high overall as most of the data values were included in the appropriate fields and records represented the items correctly, without obvious grammar or spelling errors; high consistency was observed in the sample as data in elements were consistent throughout the sample and several fields relied on controlled vocabularies; most of the data values in the sample followed the same formatting.

Overall, the analysis for this institution is that they work with high standards by following/applying international systems and standards such as LCSH, MARC and List of Standard Arabic Subject Headings for Libraries, Information Centers and Databases. However, the institution is missing the guarantee of following the rules of their metadata creation

guidelines, which is crucial in minimizing consistency errors that I noticed in the sample that, in turn, increases/maintains the level of accuracy of their collection.

4.2 Participant 2

4.2.1 Interview Analysis

The interview was conducted in the Arabic language and translation was needed to analyze the collected data from that interview. Participant 2 has a master's in library and information sciences and began working at the institution in 2000. The participant was trained through lectures and attended workshops as a personal effort to develop themself. The institution has had an bibliographic database for a long time, with around 30,000 book titles and documents in both Arabic and English languages. Most books and documents were produced by the institution itself. The institution uses the Symphony system for content management. The institution uses MARC 21[10] metadata schema for their collection. For classification system, the institution uses the Dewey Decimal Classification system version 20. The participant did not provide an answer about the controlled vocabularies used.

The participant mentioned that the metadata records in their institution are harvestable but did not specify what protocol they are harvestable with. According to the participant, they cooperated already with the Arabic Union Catalog. The institution does not have a guideline for creating metadata, however, they are following MARC 21 rules for that. The institution learns the users' needs by receiving their questions in person or comments via phone. The computer department is responsible for contacting the users regarding their needs. They are also responsible for providing digital and printed copy of the items in the institution's collection to

[10] Respondent indicated that MARC 21 is the metadata scheme used for digital collection, however only metadata in another scheme was available in the website. (See findings of metadata analysis for this institution below)

the users. There are both simple and advanced search options that includes Boolean search and limitation options, such as type of documents, date and more, in the institution's website that allows the users to search for the items they need in both Arabic and English languages. However, the search is limited for books as well as for the historical documents, as users need to submit a request to the computer department to search and receive needed documents.

The participant expressed their excitement regarding the idea of creating a large portal that contains records from different cultural heritage organizations in the Arabian Gulf countries. A meeting including some of the libraries and museums in the area is needed to discuss that idea with the support from the governments. However, the lack of trained professionals and the level of commitment on creating this project might be the two biggest barriers in creating that portal.

Overall, this institution adopted most of their systems from existing international systems such as the MARC and Dewey Decimal Classification system. It is also clear that this institution is trying to minimize and control ways to serve their users by limiting the interaction options, asking the computer department for document reserves, and not establishing public guidelines for users. This raises concerns of the limit of sharing metadata records that this institution is willing to offer for the centralized regional large portal of bibliographic database.

4.2.2 Content Analysis:

The metadata records were available to be collected directly through the institution's website. A total of 28 records were collected containing records represented by e-books, maps, documents, and pictures. The collected samples were divided by 14 metadata records representing books, seven metadata records representing maps, and seven metadata records representing historical documents.

Although the participant specified MARC 21 as the metadata scheme used for

bibliographic database, metadata records were available in different metadata scheme. As shown in Figures 4.5 and 4.6, there were seven metadata fields observed in the sample: Date, Source, Type, Relation, Creator, Name (used for title information), Description, Language, Pages number and Topic. Three metadata fields were found in 100% of records in this sample: Date, Type and Description. The lowest level of application was observed for Relation and Source 25%. There were no records included in more than one instance of any metadata fields. Users can search for an item only in the title field. The average number of subject headings observed in the metadata records sampled from this institution's bibliographic database was 0.

Analysis of central tendency measures for the length of the data values (in characters without spaces) revealed the highest median data value length in the Description field (at 448) and the lowest in Relation and Source fields (at 0). The highest average data value length occurred in the Description field (437) and the lowest were in Source (0) and Relation fields (1).

Variability measures for the length of the data values (in characters without spaces) were also assessed: the highest variance was observed for Description field (42912) and the lowest was Date and Source fields (0). The highest standard deviation was the highest in Description field (207) and the lowest in Date and Source fields (0).

I noticed three sets of metadata fields that represent the collection based on the item's type. As shown in Figure 4.5, metadata records representing books contain seven metadata fields: Date, Type, Creator, Name (used for title information), Description, Language and Pages Number. As shown in Figure 4.6, metadata records representing maps contain only five metadata fields: Date, Source, Type, Topic and Description. However, Source field was empty in all observed records. Lastly, as shown in Figure 4.7, metadata records representing historical documents contain five metadata fields: Date, Relation, Type, Topic and Description. However,

the field Relation was empty in all the observed metadata records, with the exception of two.

Figure 4.5: Example of a metadata record from Participant 2: textual item

Figure 4.6: Example of a metadata record from P.2: maps

Figure 4.7: Example of a metadata record from Participant 2: historical handwritten document item

Another point I noticed in the metadata records representing books in this collection was naming the Title field as “Name” instead of “Title field.” This could be considered confusing, especially for a user whose Arabic language is not perfect, because the participant might misunderstand what is meant by “Name.” Changing this to “Title field” would benefit the users as it would be very clear for them.

Another interesting finding was the inclusion of the names of the creators of the maps and historical documents in the Topic field instead of within a Creator field. This issue was observed in only the records representing maps and historical documents; the records that represent books have the creator’s name in the Creator field. This also could be confusing, as it might make the user misunderstand the topic of that item because it is just listing the creator’s name instead of mentioning the topic of that item.

A critical point that affects the quality of the metadata records in this collection is the absence of a filed subject heading, as well as including a clear and controlled vocabulary tool. By missing these two major factors it is important to point out the concerns of accessibility of searching and obtaining the accurate document within an adequate time. This raises the importance of meeting with users and asking them about their experiences with searching within the collection, because without a controlled vocabulary tool or subject headings it could prove harder to find an item within a reasonable time.

Lastly, there are very few options for users to interact with the items, as the institution allows only users to print the books and zoom/print the maps/historical documents. However, it is important to extend the interaction options by adding beneficial options such as email the item, save the item, mark the item as favorite, or download it.

Consideration of the three major metadata quality criteria—completeness, accuracy and consistency—revealed the following: records in the sample were found to be mostly filled, but a lack of completeness was observed in the appearance of the fields Relation (25%) and Source (25%); accuracy was medium overall as most of the data values were included in the appropriate fields and records represented the items correctly, without obvious grammar or spelling errors. Lack of accuracy was, however, observed in two fields: Topic field (50%) and naming the Title field as Name field (50%). High consistency was observed in the sample as data in elements were consistent throughout the sample. Most of the data values in the sample followed the same formatting.

In conclusion, based on the analysis of the sample of metadata records in this collection, it can be determined that that they have great descriptions about their items and users can easily understand what items represent by reading the Description field. However, the institution would

benefit from adding subject headings and a controlled vocabulary tool to their metadata element sets to make the search process easier. Also allowing more interaction options would make the metadata records more user-friendly and beneficial for obtaining accurate search results.

4.3 Participant 3

4.3.1 Interview Analysis

The interview was conducted in English language and no translation was needed to analyze the collected data from that interview. The third participant holds a master's in computer science and started working at their institution in 2008 along with self-training. The institution created its first bibliographic database in 2008 and it has two websites and three Desktop Windows software developed in-house for manuscripts data and library data. Most of the items in their collection are books and manuscripts. The institution is using an in-house developed system for digital content management. Also, they are using MARC[11] metadata schema with local modifications based on users' requirements. The institution found that classification systems do not fit their needs, so they developed their personal, local classification system inspired by the Dewey Decimal Classification system, which they call "Islamic Dewey."[12] This system is applied to subject headings[13] as well. The institution does not use any controlled vocabulary tool and their metadata record is not harvestable, however, it is easy to be exported dependent upon the need. There are no guidelines that control the metadata creation process in the institution.

[11] Respondent indicated that MARC is the metadata scheme used for digital collection, however only metadata in another scheme was available on the website. (See findings of metadata analysis for this institution below)

[12] This system was developed completely by the institution's president and it is used only within that particular institution. It not an official Arabic-Islamic version of Dewey Decimal Classification system but inspired by its idea. More information can be found: https://www.bibalex.org/als/en/page/dewey-overview

[13] Also developed locally by the institution's president based on the collections nature.

The institution interacts with users in person and via email regarding their needs. The person responsible for contacting and making decisions regarding users' needs is the head of the center, who also serves as the collection manager. There are several interaction options that users can utilize with the collection, such as email, sharing with social media, and using the institution's computers if they are visiting the institution's building. The institution offers simple search options in their website; the advanced search is only available in the institution's computers that are located in their building. The searching/browsing functions in the collection are available only in Arabic language; the institution does not see a need to add English language because the entirety of their collection is in Arabic language. The institution offers tutorials in YouTube regarding how to use their collection as well as in person monitoring for users who require more help. The participant believes the idea of creating a large portal that contains records from Arabian Gulf countries would be beneficial. Also, the participant mentioned that they provided their data to multiple information centers in different Arabian countries to show their willingness to cooperate with others. This would serve the goal of making information more accessible in the Arabian countries. The participant believes that the first step for creating a large-scale portal in the Arabian Gulf countries is to develop a proper metadata schema that fits the exact need and nature of the collection in those countries. It would also be necessary to ensure the acceptance from those institutions to exchange metadata records with each other. The participant believes that cost and the ability to create such metadata schema that covers all the data requirements are the two main challenges that this project might face.

Overall, this institution has a specific type of collection that in their opinion does not fit within existing systems in library and information science. As a result of that, they rejected using international classification systems, controlled vocabulary, and a content management system.

Instead, they created their own version that meets their needs, including tools such as the Islamic Dewey Decimal Classification system.

4.3.2 Content Analysis

The metadata records were available to be collected directly through the institution's website. A total of 28 records were collected, with all of them representing textual items. As shown in Figure 4.8, there were nine metadata fields observed in the sample: Identifier, Date (Contains both Gregorian and Islamic dates), Creator, Title, Classification (Used for Subject Headings), Publisher, Edition, Note and Pages number. All those metadata fields presented in the whole sample with 100% presence percentage except the Note field which was presented only in 65% of the sample. There were no records that included more than one instance of any metadata fields. Users can search for an item in the Title and Author name fields. The average number of subject headings observed in the metadata records sampled from this institution's bibliographic database was two. Metadata records in this institution found the same for both printed and electronic versions of the item. In other words, the records contain the same information regardless of the item's format.

Figure 4.8: Example of a metadata record from Participant 3: textual item

Analysis of central tendency measures for the length of the data values (in characters without spaces) revealed the highest median data value length in the Classification field (Subject headings) (at 80) and the lowest (3) in the Page numbers field. The highest average of data value length was in the Classification field (Subject headings) (at 73) while the lowest was in the Page numbers field (3).

Variability measures for the length of the data values (in characters without spaces) were also assessed. The highest variance was observed for the Classification field (323) and the lowest was the Identifier field (0). The highest standard deviation was the highest in the Classification field (18) and the lowest in the Identifier field (0).

I noticed that the institution uses the same fields for all their collection items. It is important to mention that there are many other important fields missing from those records such as Summary, Type, Format, and Contributor. This institution chose these metadata field sets based on their needs and interests. As the participant mentioned in the interview, the institution decided upon their own classification and subject headings systems because the international systems do not meet their needs. I noticed that all records have the same subject headings except one, which could give the collection a higher level of completeness and consistency; however, it would lower the accuracy level as there would not be specific subject headings for each item. It would also be helpful for the users to include all other missing fields to provide more information about the item.

I noticed several issues in the sample that decrease the consistency level in the collection. First, the Creator field had a different data value format in some of the records. For example, when there is more than one creator for an item, at times it was noted as “first author name and others,” while in other records it is listed as “group of authors.” Thus there is no specific format

for writing the Creator field when it comes to multiple authors. Second, Note filed was used for varying information that could be listed in its own field; this could be confusing for readers. For example, sometimes more subject headings are used, listing administration words that do not mean anything to the reader (such as “In table” or “Libraries group”). Because this is an incomplete sentence, the data value means nothing for the users, however, it might mean something for the employees in that institution. Third, the Page Numbers field has the same data value in most of the record except for one, which has additional wording: “10 additional pages are attached.”

Additional interesting findings include some errors and false data values in some fields, such as Date. One of the records had the Islamic date as 2000 and this cannot be correct because we are in 1442 now. Another error was found in the Edition field where the data value contained both the first and second edition for the same item. This cannot be correct because it should be either the first edition or second. Also naming the Subject Headings field as Classification field was confusing and could easily mislead users.

Overall consideration of the three major metadata quality criteria -- completeness, accuracy and consistency—revealed the following: Records in the sample were found to be mostly filled, but the lack of completeness was observed in the appearance of the Description, Type, Format, and Contributor 0% fields. Accuracy was medium overall, as most of the data values were included in the appropriate fields and records represented the items correctly without obvious grammar or spelling errors. However, lack of accuracy was observed in invalid data values, though only in the Date field of records (at 3%), Edition field records (3%) and naming the subject headings field as Classification field of records (100%). Medium consistency was observed in the sample as data in elements were consistent throughout the sample. Most of the

data values in the sample followed the same formatting except for Creator 12%, Note 9% and Page numbers of fields 3%.

Based on analysis of the sample metadata records in this collection, it is clear that this institution provides limited description of their items, primarily because many of other needed fields were missing. Users certainly would need help accessing the correct items, as subject headings are not functional professionally and are repeated in nearly the whole collection. Additionally, no controlled vocabulary tool is being used in this collection, which makes it challenging to find items from the first search. The institution would benefit from adding subject headings and a controlled vocabulary tool to their metadata schema to make the search process easier for users. Allowing more interaction options for users would also help, as it currently is extremely limited within the collection. Adding more interaction options would make the metadata records more user-friendly and allow the users get more benefits from using it.

4.4 4.4 Participant 4

4.4.1 Interview Analysis

The interview was conducted in English language and no translation was needed to analyze the collected data from that interview. Participant 4 holds a PhD degree in library and information science and started working in the institution in 2001 along with completing some workshop training. The institution has one bibliographic database that contains more than 700 digital books and many other information resources such as old magazines, newspapers, stamps, maps, audiovisual, dissertations and governmental documents. The institution uses the Medi INFO system for digital library content management and locally developed metadata schema from Dublin Core. Their plan is to develop their own metadata schema based on the Dublin Core by selecting fields that can be used as access points such as Author, Title, Publisher, Type,

Subject Heading and Barcode. The institution uses MARC 21 for cataloging books in their automation system only, however, they can retrieve the cataloging data (MARC form) and integrate it with the digital library system. The institution uses the Dewey Decimal classification system for their collection, as well as LSCH, controlled vocabulary for English recourses, The Big Subject Heading List, and controlled vocabulary by Dr. Mohammad Awad for Arabic resources.

The metadata records in this institution are not harvestable. The cataloging department is responsible for creating metadata records for the institution's items in the bibliographic database. The institution considers the users' needs in creating metadata records and they learn about those needs by examining user statistics of the library system and collection. However, the references service department is the only department responsible for direct contact with the users regarding their needs. The institution provides both simple and advanced search options that include Boolean search and limitation options such as type of documents, date and more in Arabic and English languages with multiple interaction options for the users such as searching, browsing, and reading full text.

The participant agreed to the idea of having a large-scale portal that contains metadata records from different institutions in the Arabian Gulf countries. The participant mentioned that the first step in this project should be selecting institutions that have bibliographic database, start a cooperative sharing system between those institutions to avoid duplicated records, find necessary financial support, finally create a large digital portal to be maintained by one of those institutions. However, the participant shared their personal experience with sharing metadata records with other institutions, mentioning that all institutions preferred to keep their records for their use only. Also, financial support is very weak because of the high cost for those

bibliographic databases. The lack of trained professionals needed to run such portals could be seen as one of the biggest barriers.

Overall, this institution uses most of the international systems to maintain their collection, such as the MediaINFO system, for digital library content management, locally developed metadata schema from Dublin Core, MARC21 for cataloging books in the library automation system, Dewey Decimal classification system, LSCH and LC controlled vocabulary for English recourses, and The Big Subject Heading List and controlled vocabulary by Dr. Mohammad Awad for Arabic resources. This institution is trying hard to meet the professional standards in maintaining their collection. The institution contacts the users from only the references service department and allows the users to interact with their collection and access it from distance.

4.4.2 Content Analysis

The sample from this institution was collected directly through their website as it is accessible to the public. A total of 28 metadata records were collected: ten of them representing books, nine representing maps and nine representing audiovisual items.

Twelve metadata fields were observed in the sample: Date, Country (Publication place), Type, Classification (Repeat the type), Format, Creator, Title, Page numbers, Description, Topic (Used for Subject Headings), Publisher and Ibarcode (Identifier). Some of these metadata fields appeared only in the records representing specific types of items, while some appeared in all records regardless of the item's type; this is shown in Figure 4.9. Four metadata fields were found in 100% of records in this sample: Format, Type, Classification, Title and Topic. However, the lowest level of application was observed for Description 32% and Ibarcode 17%. There were no records that included more than one instance of any metadata fields. Users can search for an item in the Country, Type, Format, Creator, Title, Description, Topic, Publisher

and Ibarcode (Identifier) fields. The average number of subject headings observed in the metadata records sampled from this institution's bibliographic database was two.

Figure 4.9: Example of a metadata record from Participant 4: textual item

Analysis of central tendency measures for the length of the data values (in characters without spaces) revealed the highest median data value length in Topic field (35) and the lowest (0) in Description and Ibarcode fields. The highest average of data value length is in Description field (74) and the lowest is in Page Numbers field (1).

Variability measures for the length of the data values (in characters without spaces) were also assessed. The highest variance was observed for Description field (14541) and the lowest was Date field (1). The standard deviation was the highest in Description field (121) and the lowest in Date and Page numbers of fields (1).

The item's type was listed twice with very similar data values (except for audiovisual items) in two separated fields: Type and Classification. For example, the data value for Type would be "Book - Arabic Book" and the data value for Classification field is "Book." This might cause duplications or overlap in the metadata records for the items and could affect the accuracy in the collection, as two fields would thus contain almost the same data value.

Accuracy errors were observed in some of the elements in this sample of metadata records. First, the field Country, which contains the publication place/country, has a city name in one of the records instead of mentioning the actual country name for that publication. Second I noticed two publishers' names listed in the Publisher field for one of the records without explaining which one is the current publisher or if they both are publishing the item at the same time. These types of errors need to be cleared so the reader can easily know the correct publisher for the items in which they are interested.

Another interesting finding was that metadata records followed the exact same format in the sample, except for the Creator field, which was mostly filed in the same format and written in Arabic language but contained one record that listed the author's name in English language.

Metadata fields sets were different based on the item's type. In the records representing books the following elements were included: Date, Country (Publication place), Type, Classification (Repeat the type), Format, Creator, Title, Page Numbers, Topic (Used for Subject Headings), Publisher and Ibarcode (Identifier), however, it was missing the Description field. In

the records representing maps the following elements were included: Date, Country (Publication place), Type, Classification[14] (Repeat the type), Format, Creator, Title, Page numbers, Topic (Used for Subject Headings), Publisher and Description, however it was missing the Ibarcode field (Identifier). As for the records representing audiovisual items the following elements were presented: Date, Country (Publication place), Type, Classification (Repeat the type), Format, Creator, Title, Topic (Used for Subject Headings) and Publisher, however it was missing the Page numbers, Ibarcode (Identifier) and Description fields. As for completeness, most fields appeared completely in those items except for one representing books, which was missing the Date field. Other than that, most of the fields appeared complete in the sample.

Consideration of the three major metadata quality criteria—completeness, accuracy and consistency—revealed the following: Records in the sample were found to be mostly filled, but incompleteness was observed in the appearance of field Description 0%; accuracy was medium overall as most of the data values were included in the appropriate fields and records represented the items correctly without obvious grammar or spelling errors; lack of accuracy was observed in only one record in the publisher field 3% of records, country field 3% records and naming the subject headings field as Topic field 100% of records; high consistency was observed in the sample, and most of the data values in the sample followed the same formatting, excluding one in the Creator field 3% of records that did not follow the same formatting.

Overall, based on the analysis of the sample of metadata records in this collection it shows that they are missing the Description in all items except the maps. It is very important to include that field as it provides important information to the users regarding the item. Another point this institution should consider is increasing the interaction options for users within the

[14]Repeating the item's type

collection. Currently the user can only browse and download the item, but it would be helpful to add more options such as the ability to email, print and share links via social media tools. This change would better serve more of the users' needs as some users would like to have those interaction options.

4.5 Participant 5

4.5.1 Interview Analysis

The interview was conducted in English language and no translation was needed to analyze the collected data from that interview. Participant 5 holds a Diploma of library and information science and started working at their institution in 2007. The participant took a lot of training remotely which increased their experience in the job's tasks. The institution started their bibliographic database and services in 2008; their collections contain printed books, eBooks, electronic articles, conference procedures and more. The institution uses the Sierra automated library system for digital content management and MARC 21 for metadata schema. The institution also uses the Library of Congress classification system, subject headings, and controlled vocabularies.

The participant mentioned clearly that their institution does not allow their metadata records to be exported to other institutions for privacy concerns. The head of technical services is responsible for creating metadata records within the institution with help from the Information Technology department. There are no metadata creation guidelines in the institution, however, the participant mentioned that they took into consideration users' needs while creating current metadata records. The institution allows direct contact between the person who is responsible for creating the metadata records and the users. The institution learns about users' needs by contacting them via email or from feedback received through the institution's system. Users can

conduct both simple and advanced searches using the Boolean search as well as limitation options such as type of documents, date and more in the institution's collection; these al can be done in English, Arabic, French, and German languages. They also can share, browse, and download the items they found in the institution's website. The institution offers training sessions for the users through Webex communication tool, MS teams and face to face training.

The participant supports the idea of creating a centralized digital library for the three Arabian Gulf countries, but also mentioned the importance of having one coordinator that contains all those shared records without causing copyright issues. The participant mentioned that this might be the only barrier that this project may face.

Overall, this institution adopted the international systems to maintain their bibliographic database such as the Library of Congress classification system, subject headings, and controlled vocabularies. They are using MARC 21 as metadata schema for their bibliographic database but do not have guidelines for creating metadata records, which is necessary to maintain the quality of the bibliographic database. It would be extremely beneficial to create or adopt clear guideline for creating their metadata records.

4.5.2 Content Analysis

The samples from this institution were collected directly thought their website as it is accessible to the public. A total of 28 metadata records were collected all representing textual sources.

As shown in Figure 4.10, there were twelve metadata fields observed in the sample: Title, Creator, Type, Publisher, Date, Source, Pages (Pages number), Subject (Includes subject headings), ISBN, ISSN, Volume, and Issue. Some of these metadata fields appeared in most of the metadata records and others appeared rarely. Six metadata fields were found in 100% of

records in this sample: Title, Creator, Type, Publisher, Date, Pages (Pages number). However, all other fields appeared less than 60% in the whole sample. The lowest level of application was observed for ISSN and Volume fields at 35%. No records included more than one instance of any metadata fields. Users can search for an item in the Title, Author, Publisher, Abstract, Call Number, CODEN, Dewey, DOI, Edition, Full Text, Genre, Geographic Location, OCLC number, Patent Number, Series, Time period, Subject, ISBN, ISSN, Volume, and Issue fields. The average number of subject headings observed in the metadata records sampled from this institution's bibliographic database was nine.

Figure 4.10: Example of a metadata record from Participant 5: eBook

Analysis of central tendency measures for the length of the data values (in characters without spaces) revealed the highest median data value length in the Subject field (117) and the lowest (0) in ISSN, Volume, and Issue fields. The highest average of data value length was in Subject field (120) and the lowest is in the Issue and Volume fields (0).

Variability measures for the length of the data values (in characters without spaces) were also assessed. The highest variance was observed for Subject field (3290) and the lowest

was Type field (0). The highest standard deviation was in Subject field (57) and the lowest were in the Type, Volume, and Issue fields (1).

I noticed that metadata records were missing Summary and Contributor fields, which would be very helpful to include, as those two fields add more valuable information to the records and would increase the quality of the metadata in the collection. However, the metadata fields that were listed provided sufficient information about the items. The large number of subject headings listed in the collection were very helpful; all the observed records contained at least five subject headings except one that had the Subject field but remained empty.

Another critical point I observed was that some articles were described as books in the Type field. I noticed that all records representing articles in the sample had "Book" as a data value in Type field instead of "Article" or "Journal Paper." This is a critical point that affects the accuracy in this collection.

I also noticed different formatting within the data value of the Date field. Some records included only the year of publication, while others included both the month and year of publication. Another point noticed was that some data values were written in English, while at other times written in Arabic, regardless the item's respective language. Those two mentioned points affect the consistency of the records in the collections as it proves there is no clear metadata creation rules being followed.

Consideration of the three major metadata quality criteria—completeness, accuracy and consistency—revealed the following: Records in the sample were found to be mostly filled, but incompleteness was observed in the fields of Description 0% and Contributor 0%; accuracy was high overall as most of the data values were included in the appropriate fields and records represented the items correctly and without obvious grammar or spelling errors, though lack of

accuracy was observed in the Type field in around 50% of records which had invalid data value; high consistency was observed in the sample as data in elements were consistent throughout the sample; most of the data values in the sample followed the same formatting except for in the Date field, where roughly 40% of records did not follow the same formatting and data values were written in varying languages.

Overall, the metadata records in this institution look sufficient as twelve metadata fields were presented to describe the items. It would be more efficient to include Description and Contributor fields, which would add more valuable information to the records. This institution allows five interactions for users: Save, Print, Share (via email), Download and Cite. These interaction options allow users to be more comfortable using the institution collection. It is very necessary for this institution to create their own metadata creation guidelines in order to limit the current significant issues pertaining to accuracy and consistency. They also have the option to adopt guidelines from other similar institutions, then modifying them to meet their needs.

4.6 4.6 Participant 6

4.6.1 Interview Analysis

The interview was conducted in Arabic language and translation was needed to analyze the collected data from that interview. Participant 6 holds a degree in library science and started working at their institution in 1990. The institution started creating ~~and owning~~ their bibliographic database in 2003 and include different types of items such as books, references, dissertations, manuscripts, encyclopedias, journals, databases, eBooks and audiovisual items. The institution uses Horizon 3.7 for digital content management and MARC 21 metadata fields. The participant mentioned that the Dewey Decimal Classification system is used for Arabic sources and LCSH is used for English sources. The participant did not clarify which controlled

vocabularies tool the institution uses, however, the participant confirmed that it is not one of the following: VIAF, LCNAF, LCSH, AAT, TGN, Maknaz. The institution uses The Big Subject Heading List by Dr. Mohammad Awad for subject headings, and metadata records within the institution are harvestable and shareable with other institutions. The Cataloging department is the primary creator of metadata records in the bibliographic database; however, the participant skipped the questions related to metadata creation guidelines and did not provide any information about it. I did not find any published guidelines for creating metadata records on the institution's website.

The institution learns about users' needs via social media tools, lectures, and workshops. The Public Relation committee is responsible for contacting users regarding their needs. The institution provides simple and advanced search options including Boolean search and limitation options such as type of documents, date and more in both Arabic and the English language to ensure a wide range of access to users of all types. Training is offered by the institution to increase the user's abilities to use and search within the institution system.

The participant believes the idea of creating a large-scale portal that contains records from different institutions in the Arabian Gulf countries would be a great idea. However, the participant noted a few steps that should be done before starting that project, such as studying the needs for, and the benefits of, pursuing that project. Also, the participant raises the financial and organizational concerns of the project, believing that cost, copyrights, and attaining trained human resources would be the three main barriers that this project might face.

This institution uses international systems to maintain their collection such as LCSH and MARC21. It is interesting to know this institution simultaneously uses The Big Subject Heading List by Dr. Mohammad Awad for subject headings, which is not very common in the Arabian

Gulf country. However, there is another institution participating in this study that also uses it, suggesting that it might become popular within the next few years. The participant skipped the questions related to the guidelines, and there were no guidelines found on the institution website.

4.6.2 Content Analysis

The sample from this institution was collected directly through their website as it is accessible to the public. A total of 28 metadata records were collected, all representing textual items as shown in Figure 4.11.

Figure 4.11: Example of a metadata record from Participant 6: textual item

There were nine metadata fields observed in the sample: ISBN, Creator, Title, Publication, Description (used for Physical description), Subject, Edition, Series and Contents (multiple use). Four metadata fields were found in 100% of records in this sample: Creator, Title, Description and Subject. However, the lowest level of application was observed for Series 10%, Contents 14%, and ISBN 25%. There were no records including more than one instance of any metadata fields. Users can search for an item in the Title, Author name and Date fields. The average number of subject headings observed in the metadata records sampled from this institution's bibliographic database was two. Metadata records in this institution were found to

be the same for both printed and electronic versions of the item. In other words, the records contain the same information regardless the item's format, allowing users to find a link that takes them to the electronic version of the item.

Analysis of central tendency measures for the length of the data values (in characters without spaces) revealed the highest median data value length in the Title field (54) and the lowest (0) in the ISBN, Contents and Series fields. The highest average of data value length was in the Subject field (65) and the lowest was in the Edition and Series fields (1).

Variability measures for the length of the data values (in characters without spaces) were also assessed. The highest variance was observed for Subject field (1973) and the lowest was Edition field (1). The highest standard deviation was the highest in Subject field (44) and the lowest in Edition field (1).

Some important fields were missing in the sample Summary field including those which provide information about the content of the item, and not just the Physical description of it. Another important missing field was Type, which describes the type of the item such as image or text. The Language field was also missing from the sample, which is important for collections that have items in more than one language. I also noticed that Contributor information appeared in the Creator field along with the Creator name instead of having its own field. This will affect the accuracy of the collection as contributors cannot be defined as creators. Some information such as date, publication place and publisher name appeared as subfields in the Publication field, which is very common if the institution is using the MARC metadata schema.

Another interesting point I noticed was that the Contents fields were used for multiple purposes. Sometimes this field had information about the item's language, however, in other record it was used to provide a brief description about the item. Also, in one of the records this

field contains information that belongs to the Relation field such as “this item is related to an old lecture.” This is a critical error that affects both the consistency and accuracy of the collection.

Consideration of the three major metadata quality criteria—completeness, accuracy and consistency—revealed the results reported below Records in the sample were found to be mostly complete as most of the fields had data values and were not blank. However, some of the fields were included only in a fraction of the records in the sample: Series (10%), Contents (14%), and ISBN (25%). Incompleteness was observed: fields such as Description, Language, Contributor and Type were missing in all records. Accuracy was high overall as most of the data values were included in the appropriate fields and records represented the items correctly, without obvious grammar or spelling errors. However, inaccuracy was observed in the Creator field when it included the Contributors names in 25% of records and when the data value of Contents field includes unrelated information in 9% of records. High consistency was observed as most of the data values in the sample followed the same formatting. This institution did not use any controlled vocabularies.

Content analysis of the sample of metadata records in this collection shows that they provide limited description about their items, as many required fields were missing. It is important to include the following fields: Summary, Language, Contributor and Type to increase the quality of the metadata in the collection. These fields provide very significant information about the items; including them would significantly benefit users. Another need is to establish metadata creation guidelines to control the creating/editing metadata records in this institution. I noticed some accuracy errors in the observed metadata records, as mentioned previously, and by creating these guidelines the institution would minimize those errors. Lastly, I strongly recommend that the institution create a friendlier system/website, so users feel more comfortable

using their collection. This is important because providing access in an easy and friendly way is a common user need.

4.7 Participant 7

4.7.1 Interview Analysis

The interview was conducted in Arabic language and translation was needed to analyze the collected data from that interview. Participant 7 started working in their institution in 1995 and holds a bachelor's in library and Information science. The participant was trained by expert employees in the department of educational resources management, now called the library administration. The institution started building their bibliographic database in 2002 and now have more than 250,000 items in different types including eBooks, journals and databases. The institution uses the Sierra library platform for digital content management and the MARC 21 metadata schema. The institution uses the Dewey Decimal Classification system for Arabic resources, and LC classification system for English resources. However, for items related to the Medical college they use The National Library of Medicine[15] (NLM) classification system. The institution also uses Alkhazindar Subject Heading List (A very old subjects headings list that was created by a person and stopped in the 1990s when he died) and the LSCH for English resources. The participant confirmed that metadata records in their institution are harvestable because it follows international standards, and all the metadata records were created and entered to the system by the cataloging department that contains experts' employees in cataloging. The catalogers in this institution follow MARC 21 guidelines as the institution does not have its own guidelines.

[15] More information can be found: https://classification.nlm.nih.gov/.

The participant stated clearly that they consider all the users' needs, including requesting new resources to offer information in easy ways. The institution learns about the users' needs from feedback that reached their librarians who are responsible for contacting users regarding their needs. The institution allows users to conduct both simple and advanced searches, which include the Boolean search and limitation options such as type of documents, date and more in Arabic and English languages within their collection. However, the institution does not offer any trainings or tutorials for the users regarding using their system. There are several interaction options available within the collection such as read, print, and download.

The participant believes that creating a large-scale portal containing records from different institutions in the Arabian Gulf countries is similar to the already established Arabian Union Catalog[16]. The participant believes the main challenge of having a large-scale portal containing records from different institutions in the Arabian Gulf countries would be that each institution uses different systems and standards, which would therefore create barriers in collecting records from those institutions and presenting them in one united metadata schema.

Overall, the institution uses international systems and standards, which should increase the quality of their metadata records. However, in this interview the participant skipped the question related to the controlled vocabulary tool that the institution is using. I will examine the controlled vocabulary tool used in the content analysis to identify it if there is an option for that. According to the participant, the institution has a strong bond with the users and they consider the users' needs in their services and protocols.

[16] That is not correct, the Arabian Union Catalog is a cooperated work that related to the Arabic content not the Arabic countries and it is included The British library and the Library of Congress.

4.7.2 Content Analysis

Before starting the content analysis, I noticed that this institution uses the exact same website design that the participant 6 is using, just with their own institution's logo. These two institutions are from the same country and both are governmental institutions but with totally different administrations and boards.


/ الأزمة والحرب فى الخليج : أمانة الكلمة والضبط البيليوجرافي

by سعد محمد الهجرسي, مصطفى حسام الدين م.مشارك

العربى للنشر و التوزيع، 1993

Subjects
- الضبط البيليوجرافي
- الكويت -- الغزو العراقي -- بيليوجرافيات

Description: ص ؛ 24 سم 352

Notes: يحتوى على كشافات وملاحق

Call No.: 016,9538091 2/3 سع.أز
016,9538091 2/4 سع.أز

Download MARC File: Click Here (Requires Password)

Figure 4.12: Example of a metadata record from Participant 7: textual item

The samples from this institution were collected directly through their website as it is accessible to the public. A total of 28 metadata records were collected and all presented textual items. As shown in Figure 4.12, there were seven metadata fields observed in the sample: Creator, Title, Publication, Description (used for Physical description), Subjects, Call Number (based on DDC) and Notes. Some of these metadata fields appeared in the whole sample. Five metadata fields were found in 100% of records in this sample: Title, Description, Publication, Call Number and Subjects were found in 100% of records in this sample. The lowest level of application was observed for the Notes field 10%. There were no records included in more than one instance of any metadata fields. Users can search for an item in the Title, Author name and Date fields. The average number of subject headings observed in the metadata records sampled from this institution's bibliographic database was one. Metadata records in this institution found

the same for both printed and electronic versions of the item. In other words, the records contain the same information regardless the item's format and allow users to find a link that takes them to the electronic version of the item.

Analysis of central tendency measures for the length of the data values (in characters without spaces) revealed the highest median data value length in the Title field (33) and the lowest (0) in the Notes field. The highest average data value length was in the Title field (40) and the lowest was in the Notes field (4).

Variability measures for the length of the data values (in characters without spaces) were also assessed. The highest variance was observed for Title field (509) and the lowest was Description field (4). The highest standard deviation was the highest in Title field (23) and the lowest in Description field (2).

I noticed several interesting points while analyzing this sample. It is clear that this collection is missing many important fields necessary to provide better description for the items such as Summary (contains information about the items content), Language, Type and Contributor. Including these fields would increase the completeness in this collection. It is important to point out that the Publication field includes the date of publication, place of publication and the name of publisher.

Another critical point I noticed is that the Subject fields contained only one subject heading and it was repeated in around 78% of the sample records. This means the metadata creators did not provide specific subjects headings for each item, which would provide better information and increase the accuracy of access and retrieval of items. However, this was expected because the participant mentioned that the institution is using the Alkhazindar Subject Heading List, which is an old subjects' headings list that has not been updated since the 1990s. It

is very important for this institution to replace this with a new subjects' headings list that provides more options for the metadata creators to increase the completeness and accuracy level of the collection.

Another point I observed was in using the Description field for describing only the physical part of the item, which affects the level of accuracy in the collection as the users would expect to read a description of the item's content in that field. Another point that affects the accuracy level in this collection is including the Contributors names in the Creator field. Users can easily get confused by the several names mentioned in the Creator field and may miss the main creator's name.

This institution uses the same system as the institution of participant 6, which explains why a similar website issue was observed. The system cancels the search findings every five minutes and takes the users back to the main searching page, which makes it inconvenient for me to use the system for hours during the content analysis process. The institution should consider fixing this error as they claimed they highly consider the users' needs and work toward achieving the best experience for the users.

Consideration of the three major metadata quality criteria—completeness, accuracy and consistency—revealed the following: Records in the sample were found to be mostly filled, but the overall completeness in the sample was low due to the short appearance of Notes field 10% and the Creator field did not appear in one of the records 3%; there are other missing fields such as Summary, Language, Contributor and Type 0%; accuracy was medium overall as most of the data values were included in the appropriate fields and records represented the items correctly, without obvious grammar or spelling errors.; inaccuracy was observed in the Creator field, where 27% of records included the Contributors names in this field, leading to redundancy by repeating

the same subject headings multiple times in Subject field (78%) of records; high consistency was observed in the sample as data in elements were consistent throughout the sample; most of the data values in the sample followed the same formatting.

After analyzing the records collected, it's clear that this institution needs to work more on developing and increasing their collection management. The institution should consider changing the website system as it is inefficient in retrieving the recourses and not user-friendly at all. Changing the subjects' headings list is necessary, as there are many other new/sufficient subjects headings lists being used by other institutions that lead to better results. Lastly, creating or adopting metadata creation guidelines would minimize the errors in the whole collection and, in turn, increase its quality.

4.8 Participant 8

4.8.1 Interview Analysis

The interview was conducted in Arabic language and translation was needed to analyze the collected data from the interview. Participant 8 holds a Master's in library and Information Technology along with some training from a group of information professionals; they started working with the institution in 1998. The institution began creating their bibliographic databases in 2010 and currently runs three bibliographic databases that contain 23,000 books. The institution also has different types of resources including articles, e-journals, books, eBooks, local reports and international standards reports. The institution uses the VIRTUA System for content management and MARC 21 metadata schema. As for the classification system, the participant noted that the university uses "Anglo-American rules."[17] The participant mentioned

[17] Invalid answer as Anglo-American rules are cataloging rules and thus cannot be used as a classification system.

that the institution uses LCSH and controlled vocabulary for most of their collections, however, they also use the Alkhazindar Subject Heading List (The same, dated list used by participant 7) for some Arabic items. The participant also claimed that all their metadata records are harvestable because they were created by catalogers who received a lot of training, including workshops[18]. This institution has metadata creation guidelines that determine which fields should be used from MARC 21, however, they are only available to the catalogers.

The institution contacts users via email and phone calls regarding their needs. There is a specific Information Professional responsible for contacting users to attain their feedback regarding the institution's collection and service. The institution allows users to search in both Arabic and English language, along with offering both simple and advanced searches including Boolean search and limitation options such as type of documents, date and more Additionally, the institution offers trainings and internal workshops for the users to increase their searching skills.

The participant agreed to the need of creating a large-scale portal containing records from different institutions in the Arabian Gulf countries, as it would benefit the users by providing a convenient central point of access. However, the participant believes the first step for conducting that project would be to raise the education and awareness of both employees and users regarding the importance of using/sharing those metadata records. The participant mentioned the main challenge that this project may face could be the lack of communication between the various intuitions and their different policies.

Overall, this institution uses international systems such as MARC 21, LSCH and other

[18]However, the harvestability of metadata records depends on many factors beyond training, including technological infrastructure, software being used by the institution, and more.

tools to manage their collections. It was interesting to learn that this intuition has its own metadata creation guidelines even though it is limited to the catalogers. The participant seemed to be knowledgeable, however, they did not understand the classification system which plays a main role in organizing collections. The intuition pays a lot of attention to the users' needs, which will be reflected in the metadata records display.

4.8.2 Content Analysis

The samples from this institution were collected directly through their website as it is accessible to the public. A total of 28 metadata records were collected, and all presented textual items. As shown in Figure 4.13, 11 metadata fields were observed in this sample: Locally Assigned Dewey Call Number[19], Author, Title, Publication, Physical Description, Analytical Parent (host item entry), Series, Summary (used for descriptions of the contents), Subjects[20], Added Entry Personal (used for contributor's names) and Notes (used for multiple general purposes such item type). Only one metadata field was found in 100% of records in this sample: Title field. However, other fields were presented in high percentage in the sample, such as Physical Description 97%, Note 92% and Author 85%. The lowest levels of application were observed for Series 9%, Summary 15%, and Analytical Parent 39%. There were some records that included more than one instance of two metadata fields: Subject and Personal Name. Users can search for an item in the Title, Author name, ISSN, ISBN, LCCN, UPC and Subject fields. The average number of subject headings observed in the metadata records sampled from this institution's bibliographic database was 5.

[19] 092 Locally Assigned Dewey Call Number. More information can be found: https://www.oclc.org/bibformats/en/0xx/092.html

[20] In some records the researcher found Local added subject heading: 699 Local Subject Added Entry--Uniform Title. More information can be found: https://www.oclc.org/bibformats/en/6xx/699.html

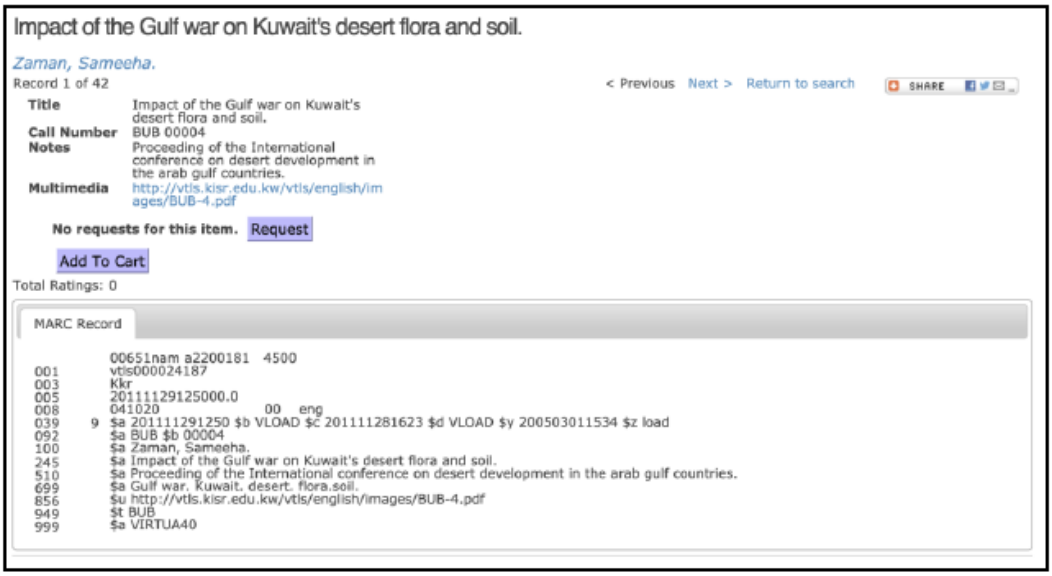


Figure 4.13: Example of a metadata record from Participant 8: digital publication of conference proceedings

Analysis of central tendency measures for the length of the data values (in characters without spaces) revealed the highest median data value length in Title field (101) and the lowest (0) in all the following fields: Series, Summary and Analytical Parent. The highest average of data value length was seen in Summary field (435) and the lowest was in Series field (2).

Variability measures for the length of the data values (in characters without spaces) were also assessed. The highest variance was observed for Summary field (1,135,034) and the lowest was Call Number field (26). The highest standard deviation was the highest in Summary field (1065) and the lowest in Call Number and Physical Description fields (5).

I found that the metadata fields presented in the sample did not provide completed information about the items; many other fields, such as Language and Type, were either missing or rarely presented in the sample, such as Summary. The data value for the presented fields were very limited and short, as well. However, some fields had several data values including the Publication field, which contained information regarding the publisher's name, place, and the publication date. Another example was Summary field, which was presented rarely in the

sample, but had a large data value of more than 3,000 characters in some records.

The observation of the sample also shows lack of consistency because different data value formats were observed across several fields. For example, the Author field had completed authors names in some records, but in other records only the first name and the first letter of the last name were listed. Additionally, some authors' names were written with dates while others were left without dates. There were not specific rules to write the authors names in the Author field because there were three different formats. Another example of the lack of consistency was the differences in subjects' headings as presented in the records. Some records had more than 15 subjects' headings while others had less than three or even none.

I noticed two main points affecting the accuracy of this collection. First, the Title field had the authors' names in almost the whole sample (97%) and then repeated those names in the Author field; this affected the accuracy level of the collection as the Title field contained another field's information thus making it incorrect. The second issue observed was that the data value in the Summary field was written in two different languages (translated) for 6% of records. This represents) 60% of the records that included the Summary field.

Consideration of the three major metadata quality criteria—completeness, accuracy and consistency—revealed the following: Records in the sample were found to be mostly filled; the overall completeness observed in the sample was medium, as eight out of the 11 metadata fields contained more than 55% presence in the whole sample; the collection was missing the crucial Language and Type fields; accuracy was low overall as data value of the Title field (100%) was incorrect in all records because they contained the authors' names; 6% of records included data values in two languages in the same instance of a metadata field (e.g., Summary); no obvious grammar or spelling errors were found; high consistency was observed in the sample as data in

elements were consistent throughout; most data values in the sample followed the same formatting, except for the Author field.

Overall, this institution has several metadata fields representing the information about items in a sufficient way. However, they should increase the appearance of the Summary field in order to provide more description of the item's content, which will offer users a clearer idea about the item. Following a specific metadata creation guideline would also help this institution to avoid the mistakes and invalid data value present in the current system, as well as enhance the consistency level of the collection as all data values in the records would follow the same format.

4.9 Participant 9

4.9.1 Interview Analysis

The interview was conducted in English language and no translation was needed to analyze the collected data from that interview. Participant 9 holds a bachelor's degree in Computer Science along with training by the institution's director and some colleagues. The participant started working at the institution in 2004—the same year the institution began their bibliographic database. The institution's bibliographic database contains 52 databases and offers more than 217,000 eBooks. They also offer other types of resources including references, eBooks, journals, repository for archives, online videos, lib-guide, oral history collection and printed collection. The participant mentioned they are using EBSCO admin for EDS discovery and OCLC hosted EZ-Proxy for digital content management. The institution uses Qualified Dublin Core standards as their metadata schema with the latest update version (DCMI Update 2020-01-20). They also use UNI-MARC for some of their collection as well. The institution uses the LC classification system and LCSH for their collection. The participant believes their metadata records are harvestable as they were created by the head of the cataloging department,

who took an online course on metadata. However, there are no metadata creation guidelines that the institution uses for creating metadata records.

The participant mentioned that the institution considers users' needs, but they do not contact them directly to learn about those needs. Instead, they note down access issues and suggestions by conducting searches on the system to determine the most available access points they can provide to users. The institution provides multiple interaction options to the users such as search, download, browse, cite, save search history, and share permalinks. Users can conduct simple or advanced searches which include Boolean search and limitation options such as type of documents, date and more in both Arabic and English languages. The institution provides information literacy presentations for students to learn searching tips along with tutorials and webinars for faculty members.

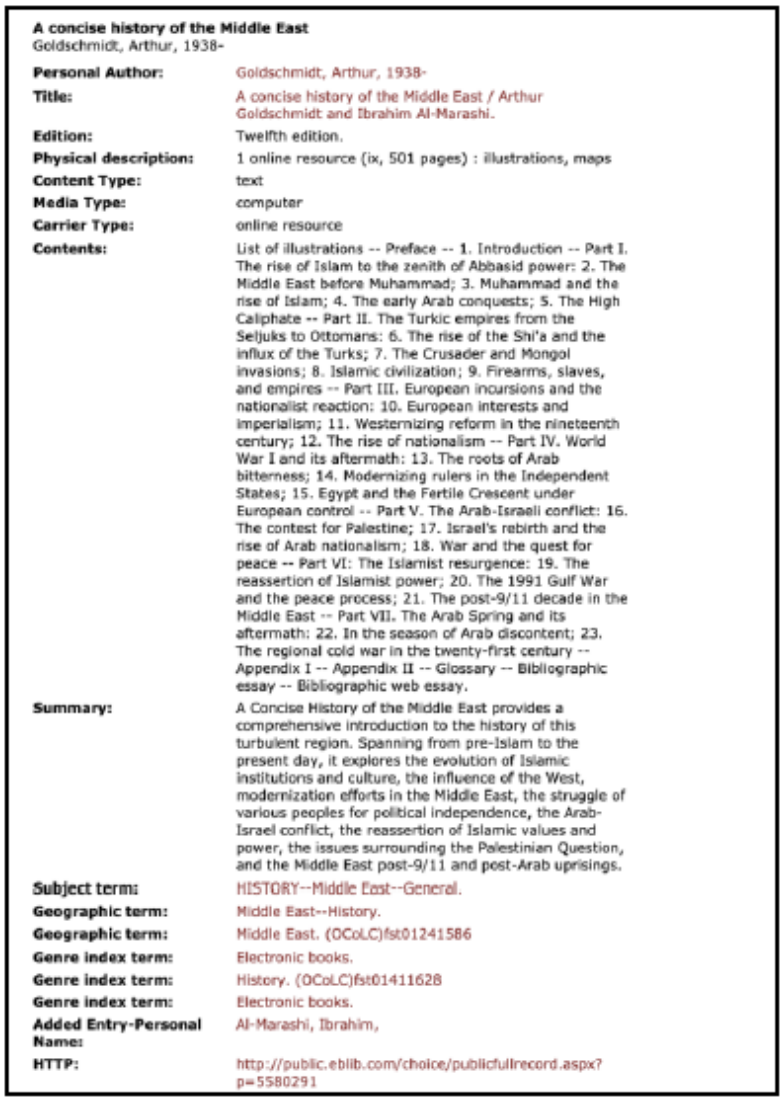
The participant believes the idea of having a centralized digital portal for institutions in the Arabian Gulf countries is interesting but noted several steps that would need to be taken before conducting that project. First would-be receiving approval from the potential participating libraries and information centers. Second, the need to specify and agree upon the information system and the metadata type that would be used. Third, the need to hold meetings that gather all participants in the project to discuss the terms. Fourth, identifying the required hardware and software needed for this project. Lastly, specifying the estimated budget and cost for this project. The participant also mentioned that this project will face four main barriers: lack of experts, the large amount of time needed to convert their metadata to a unified form, and lastly the possible high cost of this project might need while government support might not be available.

Overall, this participant was very knowledgeable about the questions and sounded very confident in their answers. The institution currently uses updated international tools and forms to

manage their collections, such as LCSH, DCMI and UNI-MARC. The institution does not have a direct connection with their users regarding their needs, which might create a gap between the users and the institution.

4.9.2 Content Analysis

The samples from this institution were collected directly through their website as it is accessible to the public. A total of 28 metadata records were collected; all of them represented textual items.



A concise history of the Middle East
Goldschmidt, Arthur, 1938-

Personal Author:	Goldschmidt, Arthur, 1938-
Title:	A concise history of the Middle East / Arthur Goldschmidt and Ibrahim Al-Marashi.
Edition:	Twelfth edition.
Physical description:	1 online resource (ix, 501 pages) : illustrations, maps
Content Type:	text
Media Type:	computer
Carrier Type:	online resource
Contents:	List of illustrations -- Preface -- 1. Introduction -- Part I. The rise of Islam to the zenith of Abbasid power: 2. The Middle East before Muhammad; 3. Muhammad and the rise of Islam; 4. The early Arab conquests; 5. The High Caliphate -- Part II. The Turkic empires from the Seljuks to Ottomans: 6. The rise of the Shi'a and the influx of the Turks; 7. The Crusader and Mongol invasions; 8. Islamic civilization; 9. Firearms, slaves, and empires -- Part III. European incursions and the nationalist reaction: 10. European interests and imperialism; 11. Westernizing reform in the nineteenth century; 12. The rise of nationalism -- Part IV. World War I and its aftermath: 13. The roots of Arab bitterness; 14. Modernizing rulers in the Independent States; 15. Egypt and the Fertile Crescent under European control -- Part V. The Arab-Israeli conflict: 16. The contest for Palestine; 17. Israel's rebirth and the rise of Arab nationalism; 18. War and the quest for peace -- Part VI: The Islamist resurgence: 19. The reassertion of Islamist power; 20. The 1991 Gulf War and the peace process; 21. The post-9/11 decade in the Middle East -- Part VII. The Arab Spring and its aftermath: 22. In the season of Arab discontent; 23. The regional cold war in the twenty-first century -- Appendix I -- Appendix II -- Glossary -- Bibliographic essay -- Bibliographic web essay.
Summary:	A Concise History of the Middle East provides a comprehensive introduction to the history of this turbulent region. Spanning from pre-Islam to the present day, it explores the evolution of Islamic institutions and culture, the influence of the West, modernization efforts in the Middle East, the struggle of various peoples for political independence, the Arab-Israel conflict, the reassertion of Islamic values and power, the issues surrounding the Palestinian Question, and the Middle East post-9/11 and post-Arab uprisings.
Subject term:	HISTORY--Middle East--General.
Geographic term:	Middle East--History.
Geographic term:	Middle East. (OCoLC)fst01241586
Genre index term:	Electronic books.
Genre index term:	History. (OCoLC)fst01411628
Genre index term:	Electronic books.
Added Entry-Personal Name:	Al-Marashi, Ibrahim,
HTTP:	http://public.eblib.com/choice/publicfullrecord.aspx?p=5580291

Figure 4.14: Example of a metadata record from Participant 9: E-book

As shown in Figure 4.14, there were 17 metadata fields observed in the sample: Personal Author, Title, Publication, Edition, Physical Description, Series, Content Type, Media Type, Carrier Type, Contents (used for listing the items chapters/index), Summary (used for descriptions of the contents), Subjects, Geographic Term, Genre Term, Added Entry Personal (used for contributors names), General Note (used for multiple general purposes such as more information about the item's edition or the publisher).Those were the metadata set elements that the institution used to describe their collection. Some of these metadata fields appeared in all metadata records in the whole sample. Two metadata fields—Title and Physical Description—were found in 100% of records in this sample, while Contents and Subjects and Personal Author were found in 97% and 85%, respectively. The lowest levels of application were observed for General Note 21%, Geographic Term 35% and Content Type, Media Type and Carrier Type 46%. Some records included more than one instance of two metadata fields: Subject, Content, Genre, and Personal Name. Users can search for an item in the Title, Author name, Series, ISBN, Electronic Resources, Periodical title and Subject fields. The average number of subject headings observed in the metadata records sampled from this institution's bibliographic database was five.

Analysis of central tendency measures for the length of the data values (in characters without spaces) revealed the highest median data value length in Contents field (504) and the lowest (0) in all following fields: Edition, Content Type, Media Type, Carrier Type, Geographic Term, Series and General Note. The highest average of data value length was in Contents field (949) and the lowest was in Content Type field (1).

Variability measures for the length of the data values (in characters without spaces) were also assessed. The highest variance was observed for Contents field (2497054) and the lowest was Content Type field (4). The highest standard deviation was the highest in Contents field

(1580) and the lowest in Content Type field (2).

I noticed this collection used a lot of metadata fields, however, not all of them were presented in all records. The highest number of elements presented in the one record were 13 metadata fields; the lowest was five metadata fields. There are some differences in the appearances of elements that affect the consistency in this collection. Another thing I noticed is that the fields names changed in some records. For example, Summary field was sometimes named Review field, even though it contained the same type of data value. Similarly, the General Note field was sometimes named Local Note field in some records. This is a critical point that negatively affected the consistency as the fields' names throughout the whole collection.

I observed redundancy in the Genre field, where the exact same data value was repeated in multiple records. This redundancy affected the accuracy of the metadata records of the collection. Another finding that affected the accuracy of this collection was the listing of dense text in the Contents field, assuming this would provide accurate information about the item. The data value for that field reached 8,448 characters without space. This seemed to be unnecessary text to be posted in that field and affected both accuracy and consistency as some of the records contained less than 300 characters in the data value of that field.

Consideration of the three major metadata quality criteria—completeness, accuracy and consistency—revealed the following: Records in the sample were found to be mostly filled. The overall completeness observed in the sample was high due the high number of metadata fields presented to describe most of the records sufficiently. However, two fields appeared rarely in the sample: General Note (21%) and Geographic Term (35%). Accuracy was high overall as most of the data values were included in the appropriate fields and records represented the items correctly, without obvious grammar or spelling errors. However, the data value was large and

unnecessary in some records in Contents field 9%. High consistency was observed in the sample as data in elements were consistent throughout the sample. Most of the data values in the sample followed the same formatting.

Overall, the metadata records in this institution follow international standards including LSCH and MARC. However, metadata creators in this institution need to agree on the terminology of some fields and unite them by choosing either Summary or Review field. The data value was nearly accurate in all fields, except for the Contents field, which needed to be more specific and limited. Having records that contain more than 8.000 characters in the data value of that field while others contain only 300 characters is not beneficial for the accuracy nor the consistency of the metadata records in the collection.

4.10 Participant 10

4.10.1 Interview Analysis

The interview was conducted in English language and no translation was needed to analyze the collected data from that interview. Participant 10 started working at their institution recently but has been working in the field since 1983. Participant 10 has a master's in library science and was trained by information services and cataloging departments. The participant mentioned that the institution currently runs four bibliographic databases. The institution's collections contain more than 30,000 items; most of them were originally printed copies that were digitized. The institution uses the Virginia Technology library system (VTLS) for digital content management and MARC 21 metadata schema. Also, the institution uses LC classification, controlled vocabularies, and Subject headings systems. The participant believes that their metadata records are harvestable (OAI-PMH) as they were created by very well-trained catalogers who took workshops from international organizations. The participant stated that they

have metadata creation guidelines, but they are not accessible to the users.

The institution has experts' librarians who have experience in providing information and research needs to users. Therefore, they know exactly how to organize the resources to be available to the researchers or the users based on their needs. Additionally, the library has liaisons (information specialists) who attend research meetings between the users and faculty members to provide information when it is needed. Those information specialists also meet with the catalogers regularly to update/inform them of the users' needs and feedback regarding the creation of the metadata records. Users can search, share, download and browse the items in both Arabic and English languages. They also can conduct simple and advanced searches that include Boolean search and limitation options such as type of documents, date and more in the collections with searching tips available on the website for more help.

The participant believes the idea of having a centralized digital library for Arabian Gulf countries is a great idea, but may be difficult to implement, as "Unclear roles and responsibilities of each library will lead to failure." The participant notes a few steps needed to create this project: ensure high commitment from the participating institutions, prepare financial and legal rules, establish an effective organizational structure with different committees and responsibilities.

This participant has had a lot of experiences working within this field since 1983. I found many other interesting points such as the institution's plan to learn about the users' needs. Having information professionals participate in research meetings and get close to the users is a unique and valuable method to learn, as it allows professionals to witness the users' needs directly. Those information professionals can then report back to the catalogers to modify their metadata creation based on users' needs. This process increases contact between users and

employees, which, by default, will increase the information services in that institution.

4.10.2 Content Analysis

The samples from this institution were collected directly through their website as it is accessible to the public. A total of 28 metadata records were collected; 14 records represented textual items and 14 records represented videos. As shown in Figure 4.15, there were 12 metadata fields observed in the sample: Personal Author, Title, Publication, Edition, Physical Description, Contents (used for listing the items chapters/index), Summary (sometimes called Subject and it was used for descriptions of the contents), Subjects term (for Subject headings), Geographic Term, Genre Term, Added Entry Personal (used for contributors names), General Note (used for multiple general purposes such as more information about the item originality).

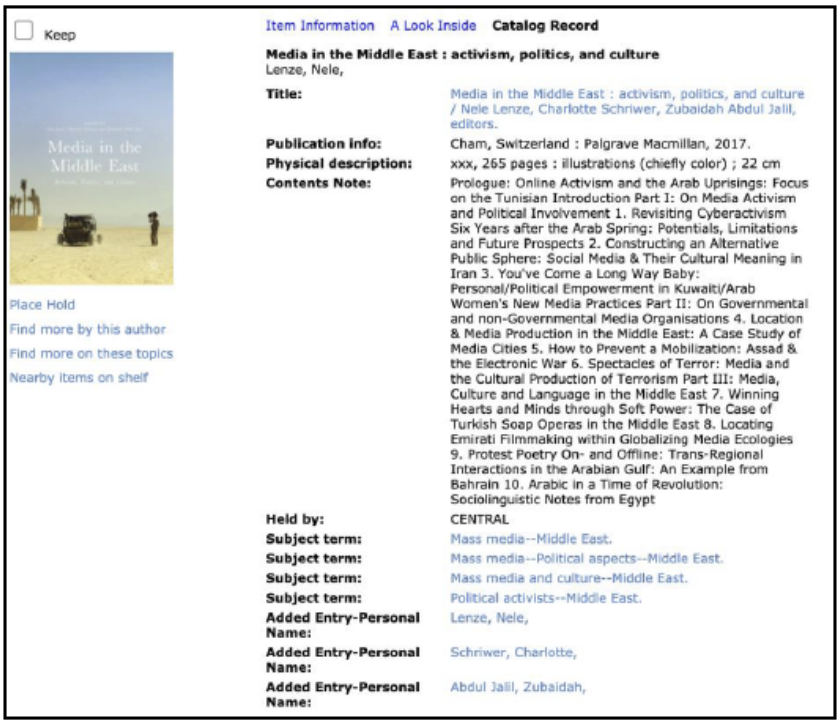


Figure 4.15: Example of a metadata record from Participant 10: textual item

I found the metadata fields set to be the same for both types of items, however, the Content field was not present in records representing videos nor in Arabic textual items. Two metadata fields

were found in 100% of records in this sample: Title, Publication and Physical Description appeared 100% while Subject appeared 85%. The lowest level of application was observed for Edition 21% and Genre Term 14%. Some records included more than one instance of two metadata fields: General Name, Subject and Personal Name. Users can search for an item in the Title, Author name, Series, Periodical title, and Subject fields. The average number of subject headings observed in the metadata records sampled from this institution's bibliographic database was 4. Metadata records in this institution were found to be the same for both printed and electronic versions of items. In other words, the records contain the same information regardless the item's format. Also, the users have access to links that take them to electronic versions of items.

Analysis of central tendency measures for the length of the data values (in characters without spaces) revealed the highest median data value length in Summary field (269) and the lowest (0) in the following fields: Edition, Contents, Genre Term and General Note. The highest average of data value length was in the Summary field (317) and the lowest was in the. Edition field (1).

Variability measures for the length of the data values (in characters without spaces) were also assessed. The highest variance was observed for Contents field (143,823) and the lowest was Edition field (15). The highest standard deviation was found in the Contents field (379) and the lowest was in the Edition field (4).

I noticed the Contents field presented only the records representing English textual items. This affected the completeness and consistency of the records, as Contents is an important field including information about the item's content. Another finding was that the Summary field appeared many times as Subject, which was very confusing for me as there was already another

field with the same title that contains subject headings. This was observed in only about half of the collection, and thus affects both the accuracy and consistency of the collection as Subject field does not substitute Summary field.

Overall consideration of the three major metadata quality criteria—completeness, accuracy and consistency—revealed the following: Records in the sample were found to be mostly filled, but the overall completeness observed in the sample was low as eight of the 12 metadata fields had less than 55% presence percentage in the whole sample; accuracy was medium overall as most of the data values were included in the appropriate fields and records represented the items correctly, without obvious grammar or spelling errors; naming the Summary field as Subject in some records (9%) was a critical point affecting the accuracy of this sample; high consistency was observed in the sample as data in elements were consistent throughout the sample; some fields appeared in records that were written in English but not in the Arabic regardless the item's type; most of the data values in the sample followed the same formatting.

The metadata schema in this collection used international systems such as LCSH, LC controlled vocabulary and MARC21. However, it is clear that catalogers are not following a specific guideline when creating the metadata records. For example, some fields were presented only in the English records and not in the Arabic records, even though both of those records represented textual items. I noticed that the Summary field was named differently in some records (Subject) but contained the same type of data value.

4.11 Participant 11

4.11.1 Interview Analysis

The interview was conducted in Arabic language and translation was needed to analyze

the collected data from that interview. Participant 11 began working in this institution in 1996 with a master's in libraries degree along with self and local training. The institution started their bibliographic database in 2000 and currently contains 38 documents, 38,000 articles, 150,000 dissertations, 155,000 books. They also offer both electronic and printed books, journal papers and dissertations. The institution uses EBSCO Discovery Service for everything including the content management tool, subject headings, classification system, metadata schema and controlled vocabularies. The participant decided to answer all the questions that related to the mentioned tools with one specific answer: "The option that the (EBSCO Discovery) offers." This response shows that either the participant does not know the actual answers, or they assume that the EBSCO Discovery Service is using well-known tools that do not need to be mentioned. The participant claims their metadata records are harvestable, but when I contacted the participant to clarify their answers, the participant did not respond.

The institution uses the metadata creation guideline provided by EBSCO Discovery Service; it is available to the public. The participant mentioned that the institution contacts the users regarding their needs via the website, workshops, ads, email, direct contact/training and visiting the departments in the institution. This process usually is done by the references disk and social media employees. Users can suggest adding new sources, asking for help to access articles that are not provided by the institution, and sharing records in the social media tools. Users can conduct both simple and advanced searches that include Boolean search and limitation options such as type of documents, date and more for items; it is available in both Arabic and English languages. The institution provides what the respondent believes are "the ideal use tutorials" to guide the users for the perfect search experience.

The participant believes the idea of having a centralized digital library for Arabian Gulf

countries is an excellent idea as it saves money and efforts for both the users and institutions. The participant suggested the first step to start that project would be to present it at the annual meeting of the administrators of libraries in the Arabian Gulf countries, or the official institutions. The participant believes if there is a true wish to do this project, then there are no barriers or challenges we should be concerned about.

Overall, the interview provided very limited information regarding the collection management, as many of the important questions were answered with a vague answer: "The option that the (EBSCO Discovery) offers." However, sufficient information was provided on the users' needs section as the participant sounded very confident answering those questions.

4.11.2 Content Analysis

The samples from this institution were collected directly through their website as it is accessible to the public. A total of 28 metadata records were collected and all presented textual items. As shown in Figure 4.16, 11 metadata fields were observed in the sample: Title, Authors, Publication, Type, Description, Subjects (used for subject headings), Categories (used for classification purposes), Language, Related ISBNs, OCLC and Accession Number (used as Identifier). Mostly all these metadata fields appeared in the whole sample. Six metadata fields were found in 100% of records in this sample: Title, Authors, Publication, Type, Languages and Accession Number were presented, and Categories field 97% in the whole sample. The lowest level of application was observed for Related ISBNs field with 42%. There were no records that included more than one instance of any metadata fields. Users can search for an item in the Title, Author name and Key Word fields. The average number of subject headings observed in the metadata records sampled from this institution's bibliographic database was 1.

Analysis of central tendency measures for the length of the data values (in characters

without spaces) revealed the highest median data value length in Description field (329) and the lowest (0) in Related ISBNs field. The highest average of data value length was in the Description field (448) and the lowest was in Type (5).

Figure 4.16: Example of a metadata record from Participant 11: eBook

Variability measures for the length of the data values (in characters without spaces) were also assessed. The highest variance was observed for Description field (306,798) and the lowest were Type, Language and Accession Number fields (all 0). The highest standard deviation was in the Description field (554) and the lowest were seen in the Type, Language and Accession Number fields (0).

In this sample some fields were presented as subfields for other fields. For example, the field Publication contains Edition, Place of publication, Date and Name of the publisher. The completeness level in this sample was high, as the presented fields described the items completely except for missing the Format field. However, incompleteness was observed in the Subjects field as it had a limited subject headings number of three. Additionally, some records

did not include any subject headings (32% of the sample); this affects the quality of the collection as it limit access and retrieval of the items. Another interesting point I noticed was that the data value in the Subjects field (subject headings) was always in English language even though the whole record was written in Arabic language. This affects the completeness level of the metadata records as Arabic items should at least have one Arabic subject heading so users who only write/read in Arabic can benefit of that subject heading.

I noticed invalid data value in the Description field where it contains "####" repeatedly for 255 characters without spaces. This error was found in three metadata records in the whole sample (10%) without explanation. This affects the accuracy level of the metadata records in the collections as the data value does not mean anything. Also, I noticed different data value format in the Authors names as two records in the sample contain different languages such as (Chinese and Arabic) or (English and Arabic), however, some records were written only in Arabic. This affects the consistency level of the sample as it should follow the same format.

Consideration of the three major metadata quality criteria—completeness, accuracy and consistency—revealed the following: Records in the sample were found to be mostly filled; the element set used completely described the items; accuracy was high overall as most of the data values were included in the appropriate fields and records represented the items correctly, without obvious grammar or spelling errors; lack of accuracy was observed the Description field when it included invalid data values (10%); high consistency was observed in this metadata sample as data in elements were consistent throughout the sample; most of the data values in the sample followed the same formatting, except for the Authors field, which contained two languages in the data value in two records (6%) of the sample.

Overall, the metadata records in this collection presents the items sufficiently as it

contains a high level of completeness and consistency. However, the metadata managers in this collection must consider following metadata creation guidelines to increase the accuracy level more. I noticed there were seven interaction options for users: users can upload the items to their google drive, print, email, save, cite, export and permalink the metadata records in that collection. This shows that this institution cares about the users' needs by providing multiple interaction options.

4.12 Participant 12

4.12.1 Interview Analysis

The interview was conducted in English language and no translation was needed to analyze the collected data from that interview. The participant from this institution began his job there in 2019 and holds a Bachelor's and Master's in computer sciences, as well as past experiences and self-training from their job at a previous institution. The institution started building the bibliographic database in 2012 but was not available to users until 2017. The collection includes more than 500,000 resources that are mostly textual. However, the participant mentioned that they still have in-site audio collection such as DVDs and microfilms. The institution is using MARC21 as their metadata schema and the Library of Congress classification system. The institution also uses LC controlled vocabularies list and subject headings. The institution uses Sierra as their content management tool.

The participant mentioned that users are welcome to provide their feedback to the front desk employees regarding their needs. The front desk employees are the responsible department to contact the users regarding their needs. The institution offers simple and advanced searches which include Boolean search and limitation options such as type of documents, date and more in both Arabic and English languages for the users.

The participant believes the idea of creating a centralized digital library for three Arabian Gulf countries is a very good idea but acknowledges that it would require many meetings and agreements to be established. Also considering the money needed for that project, including the salaries of the experts needed, would be a challenge.

Overall analysis of this interview reveals that this institution is using mostly American systems in managing their collection such as MARC21 and all LC systems. The thoughts that participant 12 shared regarding the idea of creating a centralized digital library for three Arabian Gulf countries is very important as it describes the level of challenge this project would entail.

4.12.2 Content Analysis

The samples from this institution were collected directly through their website as it is accessible to the public. A total of 28 metadata records were collected and all presented textual items. As shown in Figure 4.17, 17 metadata fields were observed in the sample: Title, Authors, Call Number (based on LCC), Edition, Description (used as format), Series, Contents (used to list the item's chapters), Summary (used as description of the content), Note (multiple random uses), Bibliography, Publication, Subject (used for LCSH-like subject headings) Indexed Term (used for non-standard controlled vocabulary subject terms), Genre/Form (used for item's type), Added Author Entry, Other Form (used to list the other available format of the item) and ISBN. Many of these metadata fields appeared many times in the whole sample. Four metadata fields were found in 100% of records in this sample: Title, Description, Publication and Subject. The Call Number field was presented 97% and the Authors field was at 94%. The lowest levels of application were observed for the Series field (25%) and the Summary field (32%). There were no records including more than one instance of any metadata fields. Users can search for an item in the Title, Author name, Key Words and Subject fields. The average number of subject

headings observed in the metadata records sampled from this institution's bibliographic database was 4.

Figure 4.17: Example of a metadata record from Participant 12: eBook

Analysis of central tendency measures for the length of the data values (in characters without spaces) revealed the highest median data value length in Title field (80) and the lowest (0) in the following fields: Series, Contents, Summary, Indexed Term, Genre/Form and Added Author Entry. The highest average of data value length was in the Contents field (4447) and the lowest was in Genre/Form (6).

Variability measures for the length of the data values (in characters without spaces) were also assessed. The highest variance was observed for the Contents field (2,021,131) and the lowest was the Call Number field (17). The highest standard deviation was found in the Contents field (1,422) and the lowest was in the Call Number field (4).

I noticed high completeness over the sample as many metadata fields were presented to describe the items. Most of those metadata fields were presented in a high number in each

record; the lowest number of metadata fields presented was ten for only one record and all other records had more than ten metadata fields describing them. All those fields were filled and there was no empty field.

Most of the metadata fields were filled with the correct data value, however I located some redundancy, which affected the accuracy of the metadata records in this collection. The first redundancy was very clear in the Series field, which was repeated twice with the same data value for the same record. However, this was observed only in one record (3%) from the sample. Another major redundancy was listing the authors names in the Title and repeating them in the Authors field. This is not only redundancy but also presents the wrong data value (authors names) in the wrong field (Title); a critical point that affects the accuracy level in this collection.

I was able to observe limited inconsistency in the Contents field as one of the records contained a large data value (7,487 characters without space) while most of the other records had less than 700 characters. The metadata element set's order was also different in the records, specifically in the fields of Contents, Note and publication. In some records they appeared at the top of the record, while in others they appeared at the bottom. These two points affect the consistency level of the metadata records in this collection; however, it is important to mention that all the data value was following the same format.

Consideration of the three major metadata quality criteria—completeness, accuracy and consistency—revealed the following: Records in the sample were found to be mostly filled; the element set completely described the items; accuracy was high overall as most of the data values were included in the appropriate fields and records represented the items correctly, without obvious grammar or spelling errors; limited inaccuracy was observed in the Series field of one record (3%) and Title field, which contained the authors names in the whole sample (100%);

high consistency was observed in the sample as data in elements were consistent throughout the sample; most of the data values in the sample followed the same formatting, except for a very limited inconsistency in the Contents field (3%)in one record that contained larger data value than other records; the ordering of the metadata fields was different several times in the sample.

Overall, this collection has high quality metadata records as it has high number of metadata fields presented completely and consistently. However, the main issue they might need to fix would be in removing the authors names from the Title field to increase the accuracy of their collection. The institution offers six interaction options to the users, which meet one of their needs.

4.13 Participant 13

4.13.1 Interview Analysis

The interview was conducted in English language and no translation was needed to analyze the collected data from that interview. The participant in this interview suggested to answer only the technical questions as they related to their job; questions related to the users' needs and metadata creation process were referred to another department. However, after contacting that department several times, I did not get any response from that department.

Participant 13 began working at the institution in 2012 and has a degree in History of Art, Conservation and Preservation Management of Collection. The participant was trained by the collections management system (CMS). The institution started their bibliographic databases in 2012 and it currently has over 400 online objects on the website and 300 objects on Google Arts and Culture webpage. However, the institution is working now to increase their online objects in their website to 800, which should be done in both Arabic and English languages. The institution uses Ke-Emu as the content management system (CMS) and OpenText as the Digital Asset

management system (DAMS). OpenText is used as the interface to pull the data into the web. The institution uses MARC metadata schema to describe their objects and The British Museum Treasure[21] (also known as Portable Antiquities Scheme) as their controlled vocabularies tool. The participant mentioned that they are using the Dewey Decimal Classification system and the Library of Congress subject headings list. The participant claimed that the metadata records in the intimation are harvestable.

The participant was able to answer three questions that related to the users' interaction options with the collection. Users can conduct searches in both Arabic and English languages with simple and advanced search option that includes Boolean search and limitation options such as type of documents, date and more Users also can search, download, share and save any record they find in the institution collection.

The participant believes the idea of creating a centralized digital library for three Arabian Gulf countries is a great idea, however, budget and grounded copyright agreements would be two main challenges that this project may face.

Overall, after analyzing this interview most of the questions were answered completely and a clear understanding of the technics and tools that the institution is using was delivered. However, I will pay more attention in the content analysis section to find the answers for at least the metadata creation process questions that address whether or not the institution is using a metadata creation guideline.

4.13.2 Content Analysis

The samples from this institution were collected directly through their website as it is

[21] More information can be found here: https://finds.org.uk/datalabs/terminology

accessible to the public. A total of 28 metadata records were collected and all presented textual items. As shown in Figure 4.18, there were 17 metadata fields observed in the sample: ISBN, Title, Authors, LC Call Number, Dewey Decimal Classification Number, Edition, Publication, Physical Description, Series, General Note, Bibliography, Contents, Summary, Subjects Terms, Geographic Term, Genre Term and Added Author (used as a contributor).

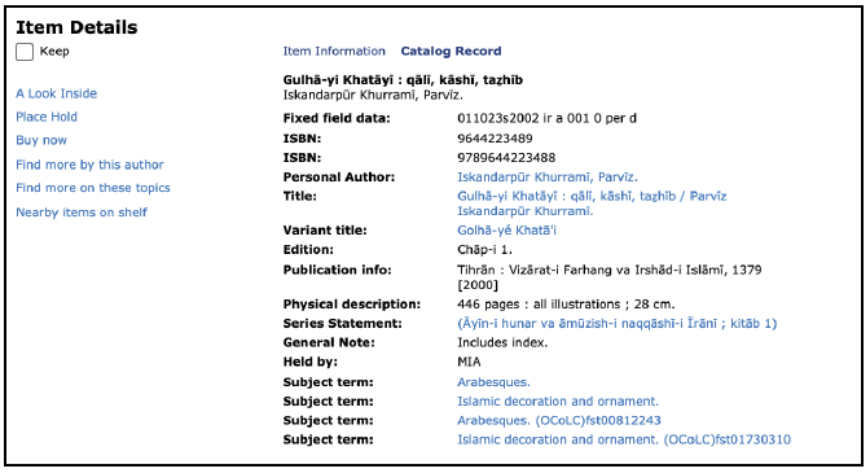


Figure 4.18: Example of a metadata record from Participant 13: textual item

Only six of those elements were presented more than 80% in the sample. Two metadata fields were found in 100% of records in this sample: Title and Physical Description. Also, Authors, LC Call Number and Subject Terms fields were presented (92%) in the sample, while the Publication field was presented 78%. All other fields were presented less than 67% in the whole sample. The lowest levels of application were observed for the Contents field (14%) and Genre Term field (17%). Some records were included in more than one instance of two metadata fields: Subject, Geographic term, General Note, Genre and Personal Name. Users can search for an item in the Title, Author name and Subject fields. The average number of subject headings observed in the metadata records sampled from this institution's bibliographic database was four. Metadata records in this institution were found to be the same for both printed and electronic versions of the item. In other words, the records contain the same information regardless the

item's format, and the users can find a link that takes them to the electronic version of the item.

Analysis of central tendency measures for the length of the data values (in characters without spaces) revealed the highest median data value length in the Subject field (71) and the lowest (0) in the following fields: Dewey Decimal Classification Number, Edition, Series, General Notes, Contents, Summary, Added Author, Geographic Term and Genre Term. The highest average of data value length was in the Subject field (114) and the lowest were in the Dewey Decimal Classification Number and Edition fields (4).

Variability measures for the length of the data values (in characters without spaces) were also assessed. The highest variance was observed for the Summary field (76,160) and the lowest was Dewey Decimal Classification Number field (19). The highest standard deviation was seen in the Subject field (276) and the lowest was in the Dewey Decimal Classification Number field (4).

Some missing fields needed to be included in the collection, such as the Language and Type fields. The fields that were present in the metadata records provided some beneficial information about the items, however, a lot of information was missing. For example, a low appearance rate was observed for the Summary (25%) and Contents (14%) fields in the collected sample. This means most of the records do not have fields that describe the content of the items, which is very important for the users. Another example is the random number for listing subjects' headings in the records; some records had 15 subjects' headings while other records had either none or just one. This affects both the completeness and the consistency level of the metadata records in this collection.

The author noticed several critical points in this sample that strongly affected the accuracy level of this collection. First, redundancy and invalid data values were observed 100%

in the Title field, which contained both the Authors and Added Author names next to the title name. The data value in Title field that related to other fields (Authors and Added Author names) also appeared the same in the actual fields. Second, I noticed invalid data values in the Dewey Decimal Classification Number in one of the records, which contained [Fic] as a data value in that field. However, the main and third critical point that sufficiently affected the accuracy level were the grammar mistakes and use of non-professional language in almost the whole sample. For example, one of the records contains "Kuwayt" in the data value of Publication field instead of the correct spelling, "Kuwait." An example of the use of nonprofessional language was seen when English letters were used to write Arabic words, such as [Altabah 1] which means edition 1. This was observed almost in the whole sample across different fields, which raised big concerns about how useful these metadata records are.

Consideration of the three major metadata quality criteria—completeness, accuracy and consistency—revealed the following: Records in the sample were found to be mostly filled, but the lack of completeness observed in the sample was missing two important fields—Language and Type; a lack of content description was observed in roughly 80% of the sample; accuracy was low overall as many of the fields included many grammar or spelling errors; invalid data value and redundancy was observed in 100% of the Title field; high consistency was observed in the sample as data in elements were consistent throughout the sample; most of the data values in the sample followed the same formatting, however a lack of consistency was observed in the Subject Terms field as some records had a high number of subjects' headings while other records did not have even one.

Overall, the sample observed contained enough metadata fields and subjects' headings, but it suffered from critical accuracy errors as the data value was written in a nonprofessional

way. This institution either does not have a metadata creation guideline or is not using it, as determined by the many errors found in the content analysis. The main positive side I noticed in this collection was that this institution offers seven interaction options to users.

4.14 Participant 14

4.14.1 Interview Analysis

The interview was conducted in Arabic language and translation was needed to analyze the collected data from that interview. Participant 14 began working at their institution in 2003 and holds a Bachelor's in library and information science. The participant received a lot of training from the head of the institution along with some training from local and international institutions in the library and information technology field. The institution started offering bibliographic database to users since its opening in 2006. Now they are running eight bibliographic databases all in Arabic language with more than 400 digital items. Most of their collection are textual items such as e-books, articles, and dissertations. However, a few DVDs and microfilms are available in the physical building of the institution. The institution uses the Horizon system for content management and MARC 21 metadata schema. The institution also uses the Dewey Decimal Classification system and The Big Subject Heading List for the bibliographic database. The participant skipped the question related to the controlled vocabularies tool used by the institution without mentioning a reason. The participant believes their metadata records are harvestable and easy to be integrated into other systems.

The catalogers in the institution took several trainings from local and international institutions as they are the primary creators of metadata records in the institution. They follow MARC 21 metadata creation guidelines to create their metadata records, which are not available to the public. The institution contacts the users via the library's website and some of the social

media tools that are run by the public relation department and services department. Users can conduct both simple and advanced searches which include Boolean search and limitation options such as type of documents, date and more in only Arabic language in the bibliographic database. The institution offers three interaction options to the users: search, browse and read. They also train new users to search and retrieve books from the shelves. Additionally, they also offer videos and paper tutorials that explain the use of the library's system along with help from the library service Helpdesk in every level of the building. The participant also skipped the question related to the institution's method in learning the users' needs.

The participant believes it is better for all institutions in the Arabian Gulf countries to focus in joining the Arabic Catalog Union instead of creating a local (Arabian Gulf area) portal that contains metadata records of collections in those countries, because the joint use of the Arabic Catalog Union would reach more users in the whole Arab world. The participant believes collecting databases from those institutions and presenting them in a united form would be the first step in this project. However, this participant states that institutions simply will not join this project and are not willing to share their metadata records with each other.

Overall, this institution has a small bibliographic database (only 400) even though they started building it in 2006. The institution follows international systems and standards such as the Dewey Decimal Classification system, MARC 21, and many others. However, this participant skipped two important questions regarding the controlled vocabulary tool being used as well as their method of learning the users' needs.

4.14.2 Content Analysis

The samples from this institution were collected directly through their website as it is accessible to the public. A total of 28 metadata records were collected and all represented textual

items. As shown in Figure 4.19, 11 metadata fields were observed in the sample: Title, Creator, Publication, Subjects, ISBN, Series, Description (used for physical description), Edition, Call Number (based on DDC with addition of page numbers), Notes (multiple uses) and Contributor.

Figure 4.19: Example of a metadata record from Participant 14: textual item

Many of these metadata fields appeared several times in the whole sample. Four metadata fields were found in 100% of records in this sample: Title, Subjects, Description, Call Number. Also, Creator and Publication fields were presented 92% in the sample. The lowest appearance was 14% for the Series field and 17% for Contributor field. There were no records included in more than one instance of any metadata fields. Users can search for an item in the Title, Author Name, Series and Subject fields. The average number of subject headings observed in the metadata records sampled from this institution's bibliographic database was 2. The same metadata records in this institution's collection represented both print and electronic versions of the item. These records included a link to the electronic version of the item.

Analysis of central tendency measures for the length of the data values (in characters without spaces) revealed the highest median data value length in the Subjects field (59) and the lowest (0) in the following fields: ISBN, Series, Notes and Contributor. The highest average of

data value length was in the Subjects field (59) and the lowest were in the Series and Edition fields (1).

Variability measures for the length of the data values (in characters without spaces) were also assessed. The highest variance was observed for the Notes field (3,762) and the lowest was the Call Number field (2). The highest standard deviation was seen in the Notes field (61) and the lowest was in the Call Number field (1).

I noticed that important fields were missing in the sample such as Summary (information about the content), Type and Language. It is true that 11 fields were presented in the sample, however, the data value was very short and presented very basic information about the items. Additionally, some fields had incomplete data value such as Description (which was used as a physical description of the item), which had only page number in some records. This affected the level of completeness in this collection as the information provided does not cover the items' information completely.

Another point I noticed was redundancy in the Title field as it contains the authors/contributors' names as well. This was observed in around 80% of the sample's records. This affected the accuracy in the collection by creating redundancy when it lists the authors/contributor names in a wrong field and repeats the same names in the Author and Contributor fields.

I noticed limited inconsistency in the data value format for both the Edition and Creator fields. In the Edition field, most of the records contain the first letter and the edition number (E1, or E2) however, one record (3%) has the complete words written (First Edition). In the Creator field, most records contained only the name of the creator, however, one record (3%) had the creator's name as well as the date. This was the only difference in data value format that I

noticed in the sample.

Consideration of the three major metadata quality criteria—completeness, accuracy and consistency—revealed the following: Records in the sample were found to be mostly filled, but incompleteness was observed in the appearance of the Series (14%) and Contributor (17%) fields; accuracy was high overall as most of the data values were included in the appropriate fields and records represented the items correctly, without obvious grammar or spelling errors; inaccuracy was observed in the Title field, which included the authors and contributors' names in roughly 80% of records and then repeated the same names in either the Authors or Contributors fields; high consistency was observed as data in elements were consistent throughout the sample; most of the data values in the sample followed the same formatting, except for one record that had different data value formatting in the Creator and Edition fields 3%.

Overall analysis of this sample confirms the fact that this institution should consider adding some metadata fields, such as Description (information about the content), Type and Language, to better represent their collection. Additionally, they should consider increasing the data value in the metadata fields as it currently provides very brief information about the items. This is important to increase the completeness level in this collection. Another thing to consider fixing would be errors in the Title field, done by removing the authors and contributors' names from it. All these issues could be easily addressed if this institution established and followed a specific metadata creation guideline to help them avoid such mistakes. Increasing the users' interaction options is another important thing that this institution needs to work on. I noticed only two interaction options available while using the institution system. The institution needs to meet or listen to the users' needs better and should increase interaction options, based on user feedback.

4.15 Participant 15

4.15.1 Interview Analysis

The interview was conducted in English language and no translation was needed to analyze the collected data from that interview. Participant 15 began their work in 2012 with a Bachelor's degree in library and information science and some training done by older colleagues. The participant also holds a Masters in the same field, which was received in 2015. The institution started their bibliographic databases in 2012 and currently has six bibliographic databases. The institution has only textual resources in their bibliographic database, in the form of eBooks and articles. The institution uses the Libero Library Management System as a tool for the digital content management and MARC 21 as metadata schema. The institution uses the Library of Congress Classification system, subject headings, and controlled vocabularies.

The participant skipped the question related to metadata creation guidelines and questions about persons responsible for metadata creation, as well as the question about collecting and utilizing user feedback. The participant stated that users can access the resources on and off campus, as well as conduct both simple and advanced searches which include Boolean search and limitation options such as type of documents, date and more. The institution offers only English collections that users can search for only in English, however, the participant mentioned that they started developing their Arabic collection and it should be done in the very near future. The institution provides off campus online training for the users that help them use/search the institution's collection easily.

The participant believed the idea of creating a large-scale portal containing records from different institutions in the Arabian Gulf countries would be a beneficial idea for all participating institutions. The participant stated that assigning an employee from each participating institution

to attend meetings to establish this project would be the first step. However, the participant believed that budget might be a main challenge to create the project along with identifying the biggest administer for the project.

Overall analysis of this interview shows that the participant was able to provide valuable answers, especially to those which related to organizing their bibliographic database. The institution is using international system, which is LC systems for subject heading, controlled vocabulary, and classification system. It was interesting to know that this institution is using Libero Library Management System as no other institution mentioned that previously.

4.15.2 Content Analysis

The samples from this institution were collected directly through their website as it is accessible to the public. A total of 28 metadata records were collected and all presented textual items. As shown in Figure 4.20, there were 12 metadata fields observed in the sample: Title, Author, Publisher, Subject, ISBN, Language, Specific Type of Material (used for type), Series, Description (used for physical description), Call Number[22], Contents and Other Name (used for contributors). Two metadata fields were found in 100% of records in this sample: Title and Subject field 100%. Also, Author and ISBN fields were presented 89% and Publisher filed 82% in the sample. The lowest level of application was observed for the Series field (10%) and Specific Type of Material field (17%). There were no records that included more than one instance of any metadata fields. Users can search for an item only using Key words. The average number of subject headings observed in the metadata records sampled from this institution's

[22] "other classification number" that "uses a scheme other than the *Library of Congress Classification*, the classification systems of Library and Archives Canada, the National Library of Medicine, the National Agricultural Library, the U.S. Government Printing Office, *Universal Decimal Classification*, or Dewey Decimal Classification schemes." More information can be found: https://www.loc.gov/marc/bibliographic/bd084.html.

bibliographic database was three. Metadata records in this institution were found the same for both printed and electronic version of the item. In other words, the records contained the same information regardless the item's format, and users can find a link that takes them to the electronic version of the item.

Catalogue Tag Display

MARC 21

1001 inventions that changed the world /

Tag	Description
020	$a0764161369
020	$a9780764161360
084	$ T15.A15 2009
245	$a1001 inventions that changed the world /$cgeneral editor, Jack Challoner ; preface by Trevor Baylis.
246	$aOne thousand and one inventions that changed the world
250	$a1st ed.
260	$aHauppauge, NY$bBarron's$c2009
300	$a960 p.$bill. (some col.)$c22 cm
505	$aIndex of inventions -- the ancient world -- from Rome to revolution -- the industrial age -- the age of empires -- birth of the modern age -- war and peace -- going global -- the internet age -- index by inventor
650	$aInventions$xHistory
650	$aTechnological innovations
700	$aChalloner, Jack

Figure 4.20: Example of a metadata record from Participant 15: textual item

Analysis of central tendency measures for the length of the data values (in characters without spaces) revealed the highest median data value length in the Subject field (48) and the lowest (0) in the following fields: Series, Specific Type of Material and Other Name. The highest average of data value length is in Content field (559) and the lowest is in Specific Type of Material and Language fields (1).

Variability measures for the length of the data values (in characters without spaces) were also assessed. The highest variance was observed for the Content field (1,953,867) and the lowest was the Language field (1). The highest standard deviation was seen in the Content field (1,398) and the lowest in the Language field (1).

The metadata fields in this sample were presented completely with a very short but informative data value. However, this collection was missing a field that provides a description

of the item's content, such as the Summary or Description field. I noticed that a small number of subjects' headings were provided for each item. Around 46% of the records in the collected sample had less than two subject headings. Also, one of the records had a very low number (only 4) of metadata fields presented and three other records have only five metadata fields. All these points affect the completeness level of the collection.

Another point I noticed that critically affected the consistency level was the different data value formats observed in several fields. For example, in the Language field I noticed three different data value formats describing the English language: Eng., E. and E (without period). Another different data value format was also observed in the Author field when some names appeared without the title (Mr. or Dr.) and others have it included. A final example was observed in the Specific Type of Material field, where most of the records have the data value "books", while one record had the data value "book-Arabic." All these differences in data value formats shows that there is no specific metadata guideline being followed, which affects the consistency level of the collection. Additionally, I noticed that different languages were used in data values (Arabic and English) for describing some items, particularly in the Subject field.

There were few redundancies observed in the sample in the Subject field when some records had the same subjects' headings repeated. However, this was rarely observed in the sample as half of the sample had less than two subjects' headings. Another point noticed was that two records (6%) have the edition information listed in the Title field, as well as including the author names in the Title for some records (14%). This is invalid data value as edition and author information should be listed separately in the Publisher and Author fields. Both redundancy and invalid data value affected the accuracy level of the collection and should be fixed.

Consideration of the three major metadata quality criteria—completeness, accuracy and

consistency—revealed the following: Records in the sample were found to be mostly filled. However, the overall completeness observed in the sample was medium as there were some short data values in the Subject field and the sample was missing proper description of the item's content, such as the Summary or Description fields. Accuracy was high overall as most of the data values were included in the appropriate fields and records represented the items correctly, without obvious grammar or spelling errors. However, lack of accuracy was observed in the Title field where the edition appeared in two records (6%) and author names (14%). A few examples of redundancy were observed in the Subject field, thus impacting the accuracy level, but all other fields had accurate data value. Low consistency was observed in the records because they were not using the same formatting in several fields, such as Language, Author and Specific Type of Material. Also, some records had two different language data values in some fields, such as Subject.

Overall analysis of this sample shows that the collection uses most of the needed fields except the field that describes the content of the item. This is a very necessary field as it provides an idea about the item for the users, which would help them in their search decisions. Major efforts should be focusing on fixing the data value formatting and uniting it into one clear format. This could be done easily by establishing, adopting, and continuing to follow metadata creation guidelines.

4.16 Findings Summary

The overall findings of this study raise several critical points regarding the organization of knowledge and bibliographic database management in the three countries of Kuwait, Oman, and Qatar. The collected data can be divided into 5 main parts, each of them presented in Tables 4.1 to 4.5. Some participants did not answer all the questions provided, so I used N/A to indicate

that the participant did not provide an answer to that part.

Table 4.1 shows the summary of findings in relation to the professional backgrounds of participants interviewed in this study. Seven of the participants have a graduate degree. 11 out of 15 participants have a degree in library and information science. The oldest participant started working in the field in 1983, while the newest one began in 2019. Eight of the participants had either personal or workshop trainings.

Table 4.1: Background information about participants

No.	Highest Academic Degree	Major	Yrs Exp	Types of Training
P.1	Bachelor	Library and information science.	6	Workshops By Colleagues
P.2	Masters	Library and information science.	20	Lectures Personal Training
P.3	Masters	Computer Science	12	Personal Training
P.4	PhD	Library and information science.	19	Workshops
P.5	Associate	library and information science	13	Online Training
P.6	N/A	Library science	30	N/A
P.7	Bachelor	library and information science	25	By Colleagues
P.8	Masters	library and information technology	22	By Professionals
P.9	Bachelor	Computer Science	16	By Colleagues
P.10	Masters	Library science	37	By Colleagues By Professionals
P.11	Masters	Library science	24	Workshops Personal Training
P.12	Masters	Computer Science	1	Personal training
P.13	N/A	History of Art	8	By Professionals
P.14	Bachelor	library and information science	17	By Colleagues
P.15	Masters	library and information science	8	By Professionals

Table 4.2 shows the summary of the findings regarding systems and tools being used for information organization by participants. There was no content management tool used in more

than three institutions. Twelve of 15 participating institutions use MARC as a metadata schema. Eight of 15 participants' institutions use the LC classification system and seven of the 15 use the Dewey Decimal Classification system. Seven of the participating institutions use the LC controlled vocabularies tool. Ten of the participating institutions use LCSH while three of the 15 use the Big Subject Heading List, especially for their Arabic records. One of the respondents (Participant 11) provided the same answer to all five questions.

Table 4.2: Systems and tools of information organization used by participants

No.	Content Management	Metadata Schemes	Classification Systems	Controlled Vocabularies	Subject Headings
P.1	Koha	MARC 21	LC	LC	LCSH = English LSASH = Arabic
P.2	Symphony	MARC 21	Dewey	N/A	Arabic Subject Headings
P.3	In-house developed system.	MARC	Locally developed - Islamic Dewey	None	Local created keywords
P.4	MediaINFO	Locally Developed from Dublin Core	Dewey	The Big Subject Heading List = Arabic LC = English	The Big Subject Heading List = Arabic LCSH = English
P.5	Sierra	MARC 21	LC	LC	LCSH
P.6	Horizon 3.7	MARC 21	Dewey = Arabic LC = English	None	The Big Subject Heading List
P.7	Sierra	MARC 21	Dewey = Arabic LC = English	N/A	Alkhazindar Subject Heading List = Arabic LCSH = English
P.8	VIRTUA	MARC 21	N/A	LC	Alkhazindar Subject Heading List = Arabic LCSH = English
P.9	EBSCO Admin for EDS Discovery OCLC Hosted EZ-Proxy.	Qualified DC	LC	LC	LCSH

(table continues)

No.	Content Management	Metadata Schemes	Classification Systems	Controlled Vocabularies	Subject Headings
P.10	VTLS	MARC 21	LC	LC	LCSH
P.11	EBSCO Discovery	EBSCO Discovery	EBSCO Discovery	EBSCO Discovery	EBSCO Discovery
P.12	Sierra	MARC 21	LC	LC	LCSH
P.13	Ke-Emu	MARC 21	Dewey	British Museum Treasure	LCSH
P.14	Horizon	MARC 21	Dewey	N/A	The Big Subject Heading List
P.15	Libero	MARC 21	LC	LC	LCSH

Table 4.3 presents the summary of interview findings in relation to evaluating user needs and metadata creation processes. Only seven of the 15 participating institutions have metadata creation guidelines, while eight of the 15 do not have or use one. Seven of the15 participating institutions offer more than one channel to contact their users while eight of 15 offer only one channel or none. Only four institutions allow the collection developer to have contact with the users regarding creating metadata records. 12 of 15 participating institutions allow more than three interaction options for the users with their collection. 15 of 15 participating institutions offer both simple and advanced search options to the users. 12 of 15 participating institutions allow users to search in both Arabic and English languages. 11 of 15 participating institutions offer either trainings or tutorials to their users to help them use their collection effectively.

Table 4.4 summarizes findings in relation to interview participants' views of the steps needed to create regional aggregations of metadata in the Arabian Gulf countries, as well as the barriers that this project might encounter. Thirteen of 15 participating collection managers support the idea. The necessary steps to create this project that were most mentioned by participants were establish meetings and creating united policies; adversely, cost and human sources were the most mentioned barriers.

Table 4.3: Users needs and metadata creation guidelines

No.	Using Metadata Creation Guidelines	How Users Contacted	Who Contacts Users	Options	Search Options	Search Languages	Offering Users Trainings or Tutorials
P.1	MARC	In person Emails Phone calls Surveys	Collection development department	4	Simple and Advanced	Arabic and English	Tutorials
P.2	MARC	In person Phone calls	Computer department	1	Simple and Advanced	Arabic and English	Direct support
P.3	None	In person Emails	Head of Center, Collections manager	5	Simple and Advanced	Arabic	Direct suppor & tutorials
P.4	None	Statistics	Reference service department.	3	Simple and Advanced	Arabic and English	N/A
P.5	None	Emails	Catalogarse	4	Simple and Advanced	Arabic English French German	Online Training Face to face training
P.6	N/A	Social media tools Workshops	Public relation committee	3	Simple and Advanced	Arabic and English	Workshops
P.7	MARC 21	In person	librarians	3	Simple and Advanced	Arabic and English	None
P.8	local developed	Emails Phone calls	- Assigned information professional	2	Simple and Advanced	Arabic and English	Internal workshops
P.9	None	None	None	6	Simple and Advanced	Arabic and English	Information Literacy Presentations & tutorials

(table continues)

No.	Using Metadata Creation Guidelines	How Users Contacted	Who Contacts Users	Options	Search Options	Search Languages	Offering Users Trainings or Tutorials
P.10	Internal guideline	In person	- Assigned information professionals	5	Simple and Advanced	Arabic and English	Online guidelines
P.11	EBSCO Discovery	In person Emails Workshops Website	References disk - Social media department	7	Simple and Advanced	Arabic and English	Online guidelines
P.12	N/A	In person	Front desk librarians	6	Simple and Advanced	Arabic and English	N/A
P.13	N/A	N/A	N/A	4	Simple and Advanced	Arabic and English	N/A
P.14	MARC 21	Social media tools Website	Public relation committee	3	Simple and Advanced	Arabic	Internal workshops & tutorials
P.15	N/A	N/A	N/A	N/A	Simple and Advanced	English	Online training

Table 4.4: The participants opinions regarding possible aggregation of metadata in a large-scale regional portal

No.	Opinion	Steps Needed	Challenges and Barriers
P.1	Great idea	Contacting the big libraries Find the administration. Collection preservation United guidelines	Uncooperative environment Different institutions' guidelines Arabic content is still not electronic. Lack of technical support
P.2	very needed	Meetings Government's support	Lack of professionals Willingness

(table continues)

No.	Opinion	Steps Needed	Challenges and Barriers
P.3	Great idea	Developing a proper container Acceptance	Containers/systems Cost
P.4	I agree	Selecting the institutions Cooperating Find the financial support. Create a large digital portal	Uncooperative environment Cost Lack of professionals
P.5	Great idea	- Find a coordinator	Copyright laws
P.6	Great idea	Study the needs and the benefits. Calculate the estimated cost. Calculate the staff number needed. Study the existed examples	Cost Copyright Lack of professionals
P.7	Like the Arabian Union Catalog	Unite the systems	Different systems
P.8	We need it	Increase the awareness. Training for both employees and scholars	The lack of communication between all those intuitions Different policies
P.9	Interesting	Taking the approval from the required libraries & information centers. Specify the information system & the metadata type used. Organize meetings. Specify the requirements (hardware &software) Specify the budget.	Not taking the project seriously Shortage in the Availability of specialized people on this field. The type of metadata used (if local then it will take time to convert to Dublin Core) Cost.
P.10	Great idea	Commitment Prepare financial and legal rules. Consortium should be planned very well with an organizational structure and different committees and responsibilities	Unclear roles and responsibilities of each library member

(table continues)

No.	Opinion	Steps Needed	Challenges and Barriers
P.11	Excellent idea	Meetings	None
P.12	Good idea	Meetings and found a base agreement. Study the possibility of it	Lack of professionals Cost
P.13	Great idea	N/A	Cost Copyright
P.14	Prefer libraries enjoy the Arabic Union Catalog instead	Collect databases. Find a united form	Libraries will not agree to join this project
P.15	Beneficial for all	Meetings Create clear policies	Cost Assigned the main administration

Table 4.5 summarizes findings in relation to content analysis revealing the most widely used metadata fields across the 15 institutions whose metadata was analyzed in this study. The average of metadata field being used in the whole sample is 12. The table shows that Author/Personal Author/Creator is the most used field – it was observed in metadata records from all 15 institutions. The same field was named slightly differently in metadata element sets used by different institutions.

The second most used field was the Title and Subject/Subjects/Subjects term/Topic which were used by 14 institutions. The third most used field was the Classification/Library of Congress Call Number/LC Call Number/Dewey Decimal Classification Number/Categories/Call Number (based on classification system) which was observed in 11 institutions. On the other hand, Table 4.5 shows the fields in the sample which were used less, in only one institution: Accession Number, Carrier Type, Country (Publication place), Ibarcode, Indexed Term, ISSN, Issue, Name, OCLC, Other Form, Pages, Specific Type of Material, Volume.

Table 4.5: Data collected from content analysis regarding the usage of metadata fields

No.	Name of the Field/Fields	# Using Field (n=15)
1	Author/Personal Author/Creator	15
2	Subject/Subjects/Subjects term/Topic	14
3	Title	14
4	Classification/Library of Congress Call Number/LC Call Number/Dewey Decimal Classification Number/Categories/Call Number (based on classification system)	11
5	Publication / Publication Info (including all 3 subfields: place, publisher, publication date)	10
6	Description	9
7	Edition	8
8	Note/Notes/General Note	8
9	Added Entry Personal/Added Author Entry/Personal Name/Contributor/Added Author/Other Name	7

(table continues)

No.	Name of the Field/Fields	# Using Field (n=15)
10	Series	7
11	Contents	6
12	ISBN	6
13	Summary	6
14	Type/Content Type	6
15	Physical Description	5
16	Date	4
17	Genre Term	4
18	Language	4
19	Publisher	4
20	Analytical Parent/Relation/Related ISBNs	3
21	Bibliography	3
22	Geographic Term	3
23	Control Number/Identifier	2
24	Media Type/Format	2
25	Pages number/Page numbers	2
26	Source	2
27	Accession Number	1
28	Carrier Type	1
29	Country (Publication place)	1
30	Ibarcode	1
31	Indexed Term	1
32	ISSN	1
33	Issue	1
34	Name [الإسم] (used for title information in 1 collection)	1
35	OCLC	1
36	Other Form	1
37	Pages	1
38	Specific Type of Material	1
39	Volume	1

CHAPTER 5

DISCUSSION AND CONCLUSION

5.1 Introduction

This chapter summarizes and discusses the findings reported in chapter 4. The collected data in chapter 4 addressed the answers to all research questions:

1. How is information organized in bibliographic databases at cultural heritage institutions in Arabian Gulf countries?
2. What tools and techniques in knowledge organization (data content standards, data value standards, metadata schemas, search options, database management tools) are used?
    a. What are the patterns of the metadata quality (completeness, accuracy, consistency) of metadata records?
    b. What levels and types of training, best practice documentation and policies are provided to metadata creators and/or managers?
    c. What role does user feedback play in decision-making?
    d. What are the similarities and differences in information organization within bibliographic databases among cultural heritage institutions in Arabian Gulf countries?
3. What are the views of metadata managers at cultural heritage institutions in Arabian Gulf countries on the creation of large-scale aggregations of digital content?

I believe the findings of this dissertation draw a very clear picture of the organization of knowledge in the Arabian Gulf countries.

5.2 Overview of the Problem

The practical problem this dissertation aims to explore is the fact that Arabian Gulf countries currently do not provide a tool that allows users to search for an item through all libraries and museum systems. There is a need for regional aggregations of digital content from libraries, museums, archives in the area (like Europeana or Digital Public Library of America) to

build that tool. To start such an aggregation, certain steps must be met. These steps include:

1. Understanding of user needs.
2. Common understanding/agreement should be developed by partnering institutions.
3. Policies and guidelines should be developed.
4. Metadata used by these institutions should be compatible (i.e., interoperable)

The research problem this dissertation aims to address is the lack of published studies to date that have examined:

1. The status of information organization, knowledge representation, and metadata practices in bibliographic databases in Arabian Gulf Countries
2. Awareness and attitudes towards creation of large-scale aggregations of digital content
3. Level of preparedness/training of information professionals
4. Potential barriers to metadata interoperability in regional aggregations.

5.3 Findings in the Context of the Relevant Literature

The findings of this dissertation show high connections of the literature in this field. Most of the studies in the literature, including Eid's 2019 paper, stated that many libraries and museums in Arab countries get the benefits of using international systems. The findings of this dissertation confirm that statement, as 14 of the 15 participants' institutions use international systems such as the Library of Congress classification system and the LCSH subject headings list. However, some institutions use Arabic developed subject headings lists, like The Big Subject Heading List, created by Dr. Mohammad Awad[23]. Another subject heading list used by other institution is the Alkhazindar Subject Heading List, which is a very dated list that has not been updated since in the 1990s when its developer died. The third Arabic subject heading list

[23] An Arabic subject heading list created by Dr Mohammad Awad that contains subject headings only in Arabic language. It is available electronically in amazon.com using this term in the search: القائمة الكبرى لرؤوس الموضوعات العربية.

being used by one of the institutions is Qa'imat Ru'us al-Mawdu'at al-Arabiyah al-Qiyasiyah lil-Maktabat wa-Marakiz al-Ma‘lumat wa-Qawa'id al-Bayanat (QRMAK), which is a special subject cataloging manual used for Arabic materials (Also called List of Standard Arabic Subject Headings for Libraries, Information Centers and Databases).

Another similarity between this dissertation’s findings and the existing literature is that the national libraries of some countries, such as Australia and Germany, use the Simple Dublin Core metadata standard (DCMES 1.1) for describing their digital resources; this was explained in Guinchard’s 2002 study. Two out of three national libraries participated in this study: one of them was using the Dublin Core metadata standard, while the other was using MARC 21. However, it is important to notice that the library using MARC 21 was established in 2012, after the Guinchard 2002 study had been published. This might be a new practice that national libraries in the Arabian gulf countries currently use; further examination in future studies is needed to test this assumption.

5.4 Unexpected Findings

The study released many interesting facts that can be discussed in several paragraphs. First, it is interesting to explore the variety of training that human resources of cataloging departments in the Arabian Gulf countries have gone through. A majority of the participants (eight out of 15) stated that they were trained by other, older colleagues/professionals. However, four out of 15 stated that they took personal training to do their job. This shows that most of the institutions in the Arabian Gulf countries require/offer onsite training for their employees to prepare them to do their job according to that institution’s standards and needs. Another interesting finding was that eight out of 15 participants have a graduate degree in Library and Information Science, while four out of 15 participants have an undergraduate degree in the field;

this makes a total of 80% of participants having an LIS educational background. Thus, it can be inferred that institutions in the Arabian Gulf countries aim to hire people with more knowledge and experience so as to maintain their bibliographic databases in higher quality.

There are a lot of similarities in the tools and technics used between the cultural heritage institutions in Arabian Gulf countries. There were 11 participating libraries in this dissertation study, and eight of them use MARC 21 metadata scheme. Also, seven of the 11 libraries use the Library of Congress classification system to organize their collection. A total of eight libraries use the Library of Congress controlled vocabularies and subject heading list. However, there are no main content management systems used by more than two libraries, which means each institution uses the system that basically most ideally matches their local needs.

All three archives that participated in the study use the MARC metadata scheme. Two of those archives use the Dewey Decimal Classification system to organize their collection, while the third has its own locally developed system inspired by the Dewey Decimal Classification system. All the participants from archives did not provide answers regarding the controlled vocabularies tools used by their institutions. I believe those archives are using free-text keywords instead of controlled vocabularies because they have a very limited budget and are not supported by the government. All those participated archives are using Arabic subject headings to describe their collection, and no two use the same content management system.

Unfortunately, only one museum participated in this dissertation study due to COVID-19 pandemic. This museum uses the Ke-Emu Content Management System and MARC metadata schema. The museum uses the Library of Congress subject headings, Dewey Decimal Classification system, and British Museum Treasure controlled vocabularies to describe the museum's items. Future studies will need to expand the participation by museums to allow

researchers to better identify the most common tools used in organizing information.

I found that some institutions were using older versions of information and knowledge organization standards . One participant indicated that their institution used the Dewey Decimal Classification version 20, which was published in 1989. There have been three updates to that system since then; the most recent version is 23, published in 2012, which has been applied by many libraries. The institution using the older version is an archive, and as I explained previously, archives in the Arabian Gulf countries are suffering from the lack of financial support from the governments, which might be preventing them from converting to newer versions of metadata standards.

The findings also revealed the most used fields in the metadata records of the participating institutions. There were 39 different metadata fields observed in the whole sample of 412 metadata records. As shown in Table 4.5, which present results of the author's mapping of the observed metadata fields and levels of use, a total of 12 fields appeared in only one institution's metadata. The most frequently used field appeared in metadata records of 15 institutions and represented the creator of an information object: different metadata elements sets had different names for this field, including Author, Personal Author, and Creator field. The Title field was found in the metadata records of 14 institutions (it is important to point out here that the 15th institution used the field called Name to include the Title information). Also, the field containing classification numbers or call numbers based on a classification system was used in metadata from 11 institutions: Classification, Library of Congress Call Number, LC Call Number, Dewey Decimal Classification Number, Categories, Call Number. The Publication field appeared in metadata from 10 institutions, and the Description field was found in metadata collected from nine institutions. Some metadata fields were used only in one institution each:

Accession Number, Carrier Type, Country (Publication place), Ibarcode, Indexed Term, ISSN, Issue, Name, OCLC Control Number, Other Form, Pages, Specific Type of feedback collected Material, and Volume.

Interoperability at this point might be challenging, as many issues were observed in the sample. Some institutions have high-quality metadata records in terms of completeness, accuracy, and consistency, but some others are clearly suffering from low quality metadata records—most especially those which do not have local metadata guidelines and do not follow an existing standard one (for example, the Dublin Core User Guide https://www.dublincore.org/resources/userguide/, or Resource Description and Access Toolkit https://www.rdatoolkit.org/). Another challenge to the interoperability is the fact that some institutions use the exact same metadata records for both printed and electronic versions of resources. Metadata guidelines for major metadata standards specifically instruct that electronic version should be represented by separate metadata records than the printed versions, as they have different data values for some metadata fields such as Type and Format. There is a lot of work to do for the Arabian Gulf countries institutions to ensure interoperability of future aggregations. Adopting and applying unified metadata creation guidelines could be one of the first steps towards interoperability in the aggregation. Generating crosswalks that show equivalences between metadata fields used by different institutions in the region is another important step, and this study has started the work on identifying such equivalences.

The participants' views regarding large scale portals that would aggregate metadata records representing items in the Arabian Gulf countries' cultural heritage collections were interesting:13 of the 15 participants enthusiastically supported the idea as they believe it is necessary within the region. However, they also raised some concerns, specifically emphasizing

the associated costs. The Arabian Gulf countries are generally rich countries, and they are in good standing financially so cost should not be a concern here. I believe that participants meant that the government would not be willing to spend high cost on such projects. Cultural heritage institutions such as libraries and archives in the Arabian Gulf countries typically do not receive significant funding from governments, which would make it very challenging to convince governments (especially while they are suffering financially now because of COVID-19) to invest in projects like this.

Another concern raised by participants was the lack of experts/human resources prepared to design and implement projects like this. This is a serious problem that institutions in the Arabian Gulf countries are facing due to the lack of experience in that area. There are no established large-scale aggregations hosted in that region, so employees' experiences are very limited. As a solution, the Arabian Gulf countries might need to hire experts from other countries that have experience working and managing large scale portals, which would increase the cost of the project as well. By observing the employees list of some institutions in Oman and Qatar, it is clear that both countries rely on the international workforce for the technical aspects of creating/managing digital collections.

One of the participants indicated that the idea of creating a large-scale portal containing metadata records of items in the Arabian Gulf countries is very similar to the Arabic Union catalog (https://www.aruc.org/en/home). However, union catalogs, including the Arabic Union Catalog, just like other national and international union catalogs (including the well-known global union catalog WorldCat) are aggregations of only MARC 21 bibliographic metadata and therefore differ from the aggregations of digital content with accompanying metadata such as DPLA, Europeana, Digital Library of The Middle East, and more. In addition, the Arabic Union

Catalog is a cooperative project that focuses on the Arabic content, but not on the Arabic countries. As a result, it aggregates metadata records that represent any Arabic content, including the content created in in non-Arabic Muslim countries such as Malaysia; and the content that is held by cultural heritage institutions outside of the region (e.g., in The British Library and the US Library of Congress collections).

Although this is beyond the study scope, I (as a user) observed serious usability problems when searching the databases of metadata records on the websites of two institutions. These websites cancel or time-out the user search every five minutes, even when the user is still actively browsing the list of search results. After cancelling the search page, it takes the user back to the main search page with cleaned search bars (and the last used search criteria are not preserved). This happened while searching on two different devices as well as different browsers. As a result, the user would need to conduct the same search multiple times: this is what I had to do to get the sample completed. Further research is needed to show how the institutions in question address these usability issues and concerns as raised by their users. Such usability issues can be a serious barrier in information discovery, even if metadata is accurate, consistent, and complete.

The findings of this study confirmed my expectations and assumptions as a researcher for most of the part. As a Kuwaiti citizen and a former Master student in Library and Information Science in Kuwait, I had a general initial idea, based on anecdotal evidence, of the information representation and knowledge organization status of institutions in my country and neighboring countries. However, a number of the findings of this study were surprising.

I was expecting to see that institutions do not examine and take into account the users' needs when choosing information systems for content management, selecting existing

international or developing local metadata schemes and information organization standards, creating metadata records, etc. I believe there are two reasons for that. First the role of the information professional as an intermediary is strong in the region as most of the information users in the Arabian Gulf countries tend to ask the information professionals (e.g., librarians) whenever they want more information about the item as opposed to reading/searching for it by themselves. I personally experienced that when I was doing my practicum course at Kuwait University and worked as a cataloger for 40 hours in one of the academic libraries in Kuwait. During my practicum, several users came to my office and asked me to give them bibliographic information regarding some books instead of searching the catalog. The second reason why institutions do not include user's opinion regarding creating metadata records is the relative lack of interest from the users in the Arabian Gulf countries who utilize the libraries infrequently compared to information users in the United States. There are some interested and loyal library visitors but they are not the majority. One of my colleaagues who works as a librarian in Kuwait shared interesting observations that sometimes users sent their nanny ort their personal driver to the library with a note requesting a book, treating libraries much like restaurant pick up services. Familiarity with these local specifics of library use led me to assume that users' needs will not be well addressed by the cultural heritage institutions in the Arabian Gulf countries.

Another expected finding of this study was that a lot of institutions do not use metadata creation guidelines. This could be due to several reasons. First, it is quite possible that most of cultural heritage institutions' employees are not aware of the existence of those guidelines. From my personal experience as a former student in both Bachelor's and Master's degree programs in library and information science in Kuwait, the only metadata creation guideline that student were taught about was the Anglo-American Cataloging Rules. This illustrates the lack in instruction

on metadata creation guidelines and their importance in the universities in the region. Another plausible explanation is that most of metadata creation guidelines currently exist on;y 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. While some of the guidelines were translated to Arabic (e.g., MARC guidelines), the question remains if the library archive, and museum workers in the region are aware of the existence of those translated versions and/or are able to locate those files on the internet? Answers to these questions would provide a clearer picture as to why metadata creation guidelines were not used in some institutions.

One interesting and unexpected finding of this study is that no one content management system was used by more than three institutions. This contradicted my assumption, that many institutions might be using Horizon system which from my experience is very popular in Kuwait. I believe the main reason for this unexpected variety is the different levels and sources of funding of the institutions that participated in this study. Not all the institutions are receiving the same funding or have sizeable budgets, therefore some of them different (often higher quality and more expensive) systems than others. This might also be the reason for the fact that many of the institutions that participated in this study do not rely on the same controlled vocabularies and knowledge organization systems. For example, one of the institutions used an old version of Dewey Decimal Classification system —most likely because they have limited budget that will not allow them to update their existing metadata when the new versions were released. Other institutions preferred to use older, no longer update versions of subject headings list such as Alkhazindar list because they invested in these lists in the 1990s and cannot afford paying more to switch to an up-to-date controlled vocabularies. The financial reasons for these findings are

also reflected in the fact that majority of the interview participants pointed out the cost as a barrier for the large-scale aggregation of digital content.

5.5 Limitations

This study is exploratory in nature and has several limitations. The first kind of limitations is related to data collection. The content analysis, when used alone, provides rich data on the metadata practices as embodied in metadata records but does not provide important contextual information for its findings: for example, infromation related to decision-making and factors affecting it. This study was designed to overcome this limitation by combining the interviews of metadata managers with content analysis of metadata records. The interviews preceded the content analysis, and follow-up interviews were planned to help clarify some of the content analysis findings. However, due to COVID-19 impact on the work of libraries, archives and museums that unfolded during the course of this study, I was not able to get any follow-up interviews with participants many of whom initially agreed to address remaining questions in the follow-up interview.

One of the known limitations of the interview as a 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. This might add provide useful contextual information overall, however, it raises some concerns regarding the accuracy of given answers. For example, the participant stated that their current institution started working on digital collections i in 1983, but the institution was only founded in 2002.

Another limitation I noticed with the collected interview data was that some participants

avoided providing information regarding their policies or practices related to users, as well as in regard to their metadata creation managers. This could be due to concerns that providing these answers might put them at risk of losing their job. For example, one of the participants had some concerns before conducting the interview as they had a bad experience while participating in a research project prior, which could be why they skipped some questions. Another example was that one of the participants took caution in replying to questions related to their administration or managers. These questions might include information about who is responsible for creating metadata or if there are specific metadata creation guidelines being used. The participant skipped these questions without explanation, which might be related to their concern for getting involved in any job risks for divulging information about the administration. Overall though, most of the interview questions were answered by all participants, which allowed to collect rich data in this study.

Another limitation of this study is its broader scope, without a focus on digital collections. While this study interview guide contained specific questions about digital collection metadata practices, participants often responded as if they were asked about institutions' entire collections as a whole (including physical collections) and sometimes confused digital libraries with online catalogs. This lack of distinction is quite common for librarians in the Arabian Gulf countries, and it is important to educate and teach the participants the differences between the bibliographic databases and the digital libraries. Currently, there is a small number of digital collection projects in the region held by the libraries (apart from the national libraries). However, some digital library projects are developed by museums (e.g., Museum of Islamic Art in Qatar) and archives (e.g., Alqabas newspaper archive in Kuwait). Future study is needed that will look specifically at the digital collection creation, information representation and knowledge

organization practices in digital collections developed by the cultural heritage institutions in the Arabian Gulf countries.

Low response rates from museum metadata managers was one of the main limitations of this study. Due to the COVID-19 pandemic most museums in the Arabian Gulf countries were closed. Also, most of those museums hired experts and employees from Europe or the United States, and when COVID-19 became an issue those employees preferred to travel back to their countries until the museums reopened after the pandemics. Some countries like Kuwait encouraged non-Kuwaiti residents to travel back to their countries and return when the pandemic is over to minimize the load on hospitals. I was able to contact more than six museums, but only representatives from three museums responded. After introducing the focus of the study to potential respondents from those three museums, only one of them reported that they have the metadata creator onsite and were willing to participate in the study. Although the data collected from this single museum participant was interesting, it cannot be generalized nor considered representative of all museums in the Arabian Gulf countries. This indicates the need for conducting future studies (under better circumstances) that examine the organization of knowledge in Arabian Gulf countries' museums.

5.6 Contribution and Implications

I believe this dissertation has several contributions that have the potential of improving the organization of knowledge in the Arabian Gulf countries. There is a lack of empirical studies on information organization in Arabian Gulf counties in general, as no studies examine the practices in the creation or adoption of metadata in cultural heritage institutions. Also, no published studies to date evaluate metadata in bibliographic databases for the common metadata quality criteria (e.g., accuracy, completeness, and consistency as defined by Bruce and Hillman)

nor assess its potential interoperability in the aggregation that would provide a central point of access to collections of cultural heritage institutions. This study addressed this gap and made a research contribution to developing an understanding of these issues, which are important for evaluating the feasibility and planning of future aggregations of digital content.

A practical implication of this dissertation is the first crude version of the mapping between fields used in the metadata of 15 cultural heritage institutions, which will need to be expanded upon by future studies but can serve as a first step in ensuring the interoperability of metadata. Another practical implication that can result from this study is the revision of metadata courses in LIS education programs in Arabian Gulf countries to place emphasis on collecting information on the needs of users and applying this information in metadata decision-making. This will help institutions to locate the problems that their users face when retrieving metadata records.

Furthermore, the findings of this study indicate the need for offering more metadata courses at universities within Arabian Gulf countries, as many of the participants mentioned a lack of human resources in this area. It is important for those universities to start preparing their students and teaching them the necessary information regarding metadata quality, standards, interoperability, and metadata harvesting. Likewise, adding the digital library coursework to the LIS, museum, and archival program, is necessary. This preparation is crucial for creating large scale portals that contain metadata records of items in the Arabian Gulf countries.

Future research is needed to examine the metadata and knowledge organization practices in museums in the Arabian Gulf countries, in terms of metadata quality, techniques and tools used for information representation and knowledge organization. Another recommended future study would be examining the status of the organization of knowledge in other Arabian Gulf

countries beyond the scope of this study: Saudi Arabia, Bahrain, and United Emirates. The patterns observed in metadata and metadata-related practices in these other countries might affect the establishment of large-scale portals that include metadata records from all Arabian Gulf countries.

Future research is needed that includes more metadata quality criteria to provide another prospective in examining the metadata quality of record on institutions in the Arabian Gulf countries. For example, I believe adding accessibility and conformance to expectations would provide more explanation regarding the usage of content management tools and interacting with the users. I believe evaluating metadata against these more metadata quality criteria in such a study will provide more robust understanding of metadata quality in Arabian Gulf bibliographic databases.

5.7 Conclusion

This dissertation is the first to explore the status of the organization of knowledge in bibliographic databases of cultural heritage institutions in Arabian Gulf countries. It will be useful as a benchmark for related future studies. The data was collected from 15 institutions within three different Arabian Gulf countries: Kuwait, Qatar, and Oman. The findings show that most of them use international standards of information organization that have been adopted or adapted from Europe or the United States. However, the lack of local metadata creation guidelines for the non-standard metadata resulted in problems with metadata quality in some institutions. Several steps are needed to ensure interoperability when creating large-scale portals to aggregate metadata records that represent items in the Arabian Gulf countries' bibliographic databases and start metadata harvesting. Future research and practical projects are needed at this point to adequately prepare for such projects.

APPENDIX A

THE EMAIL INTERVIEW QUESTIONS

1. When did you start your work at your institution?(متى بدأتِ العمل في مؤسستك)

2. Who trained you during your work life in the Information center? (من قام بتدريبك طوال فترة عملك في مركز المعلومات)

3. What is your academic background or your academic degree? (ما هي خلفيتك الأكاديمية او المؤهل العلمي)

4. When did the institution start creating digital project (Archive)? How many digital projects does the institution currently manage? (متى بدأت المؤسسه بتأسيس الأرشيف الالكتروني - ما عدد المشروعات الالكترونيه التي تديرها حاليا المؤسسه)

5. How many items the institution has in the digital archive? (كم عدد المصادر التي توفرها المؤسسه في ارشيفها الالكتروني)

6. What type of resources does the institution provide to the users? (ما نوع المصادر التي توفرها المؤسسه للمستفيدين)

7. What system is the institution using for digital content management? (ما هو النظام المستخدم من قبل المؤسسه لإداره المصادر الالكترونيه)

8. Which metadata schemes the institution is using (standard e.g. Dublin Core, MARC, EAD, MODS, or locally developed)? (أي نوع ميتاديتا تستخدمه المؤسسه حاليا)

a. If using a locally developed metadata scheme, how did you decide which elements to include? (اذا كانت تستخدم ميتاديتا محلي، كيف تم الاتفاق على الحقول المستخدمه)

b. Did you consult existing metadata schemes to base your unique scheme on? (هل تم الاستعانه بميتاديتا عالميه في تكوين الميتاديتا الخاصه بكم)

9. What version of metadata scheme is the institution using? (ما هي نسخه نظام الميتاديتا المستخدمه حاليا)

a. If it Dublin Core, which version: Simple DCMES or qualified DC (DC terms)? (اذا كانت دوبلن كور فأي نسخه هي)

b. If MARC is it MARC 21 or another version?(اذا نظام المارك فهل هي مارك ٢١ او نسخه اخرى)

10. What classification systems is the institution using for the digital collections? (ما هو نظام التصنيف المستخدم في المؤسسه لإداره المجموعات الالكترونيه)

11. What controlled vocabularies is the institution currently using? For example: VIAF, LCNAF, LCSH, AAT, TGN, Maknaz? Why did you choose it? (ما هو نوع مصطلحات مقيدة المستخدم في المؤسسة حاليا , لماذا تم اختياره)

12. What subject headings system is the institution using for the digital projects? (ما هو نظام تصنيف رؤوس الموضوعات المستخدم في المؤسسه)

13. Are the metadata records in your digital projects harvestable? OAI-PMH or others? (هل سجلات الميتاديتا في مجموعتكم الالكترونيه قابله للنسخ و المشاركه مع انظمه اخرى)

14. **Note:** Metadata Harvesting refers to gathering metadata from multiple places or archives and storing it in a central database. (حرث الميتاديتا مصطلح المقصود منه المعلومات الببليوجرافيه للمصادر قابله للتجميع من عدة مؤسسات لتتخزن في قاعدة بيانات رئيسية)

15. Who are the primary creators of metadata records in your institution for digital projects? What kind of training is offered? (من هو المؤسس الرئيسي للسجلات الميتاديتا في المؤسسه للمصادر الحديثه في الأرشيف - و ماهو نوع التدريب الذي تم تقديمه له)

16. Does the institution have metadata creation guidelines for digital projects? Is it available to the users? Where is it available? (هل هناك لوائح و قوانين لإنشاء سجلات الميتاديتا في الأرشيف - و هل هي متاحه للمستفيدين - و اين متاحه)

17. Does the institution consider the users' needs when creating the metadata guidelines? (هل تأخذ المؤسسه بعين الاعتبار احتياجات المستفيدين عند تحديد نظم و لوائح الميتاديتا)

18. How does the institution learn the users' needs regarding organizing information? (كيف تعلم المؤسسه احتياجات المستفيدين لعمليه تنظيم المصادر)

19. How does the institution contact the users to learn about their needs regarding organizing information? (كيف تتواصل المؤسسه مع المستفيدين للتعرف على احتياجات المستفيدين لتنظيم المصادر)

20. Who is responsible to contact the users regarding their needs? Is it done directly by the person who is responsible of creating the metadata guidelines or other departments? (من هو المسؤول عن مخاطبه المستفيدين للتعرف على احتياجاتهم - هل هي تتم بشكل مباشر عن طريق الشخص المسؤول عن انشاء سجلات المتياديتا او اقسام اخرى)

21. How can the users interact with the institution's digital collections (Searching - Sharing - Downloading - Browsing - others)? (ما هي طرق التفاعل التي يستطيع المستفيدين فعلها بالمصادر الالكترونيه للمؤسسه)

22. Does the institution provide advanced search or just sample search? (هل توفر المؤسسه بحث متقدم او فقط بحث بسيط)

23. Have you considered making searching/browsing functions available in languages other than Arabic? If not, what factors impact this decision? Is this something you plan to implement in the future? (هل فكرتم بتوفير لغه اخرى للبحث و التصفح - اذا لا ما هي العناصر التي جعلتكم تتخذون هذا القرار - هل هذا شيء تخططون لعمله في المستقبل)

24. What guidelines or tutorials the institution provides the users with using the institution's systems?(ما هي التعليمات التي تقدمها المؤسسه للمستفيدين عند استخدامهم نظام المؤسسه)

25. What do you think about the idea of creating a centralized digital library for three Arabian Gulf countries such as Europeana and Digital Public Library of America (DPLA)? (ما هو رأيك في انشاء مركز معلومات رئيسي يوفر عمليه البحث عن المصادر المتاحه في اكثر من مؤسسه خليجيه)

26. What steps do you expect this process will require?(ما هي الخطوات بإعتقادك ستحتاج هذه العمليه)

27. What challenges and barriers do you think the institutions that cooperate on this project might face in the process of creating a centralized library? (ما هي الصعوبات و المعوقات التي تعتقد ستواجه مشروع كهذا)

28. Would you mind sharing with us other metadata staff contact information in your country that you think they will be interested to participate in our interviews? (هل من الممكن ترشيح موظفين اخرين في بلدك تعتقد ان لديهم اضافه قيمه لمقابلة البحث)

29. Would you be able to participate in a follow up interview after the researcher reviews some of the metadata records in your institution? (هل بوسعك المشاركة بمقابلة اخرى مصغرة بعد ان تتم فحص عينه من السجلات الالكترونية في مؤسستكم)

APPENDIX B

UNT INSTITUTIONAL REVIEW BOARD LETTER OF APPROVAL FOR INTERVIEW DATA COLLECTION

May 4, 2020

PI: Oksana Zavalina
Study Title: The status of the organization of knowledge in cultural heritage institutions in Arabian Gulf countries.

RE: Human Subjects Application # IRB-19-875

Dear Dr. Oksana Zavalina:

In accordance with 45 CFR Part 46 Section 46.104, your study titled "The status of the organization of knowledge in cultural heritage institutions in Arabian Gulf countries." has been determined to qualify for an exemption from further review by the UNT Institutional Review Board (IRB).

Attached to your Cayuse application in the Study Detail section, under the Attachments tab, are the consent documents with IRB approval.

No changes may be made to your study's procedures or forms without prior written approval from the UNT IRB. Please contact The Office of Research Integrity and Compliance at 940-565-4643 if you wish to make any such changes. Any changes to your procedures or forms after 3 years will require completion of a new IRB application.

We wish you success with your study.

COVID-19 is having an impact on normal operations and procedures at UNT. Please follow all UNT rules and procedures regarding your human research studies, including cessation of face-to-face interactions with participants, during this time.

Note: Please do not reply to this email. Please direct all questions to untirb@unt.edu

Sincerely,

Gabe Ignatow
Professor
Chair - Institutional Review Board

APPENDIX C

IRB-APPROVED INFORMED CONSENT FORM FOR INTERVIEW DATA COLLECTION: ENGLISH-LANGUAGE VERSION

TITLE OF RESEARCH STUDY: The status of the organization of knowledge in cultural heritage institutions in Arabian Gulf countries

RESEARCH TEAM:
Principal Investigator: Saleh Aljalahmah Department of Information Science, College of Information, University of North Texas (UNT), phone: (541)514-0503, email: salehaljalahmah@my.unt.edu
Faculty supervisor's name (Chair of dissertation committee): Dr. Oksana L. Zavalina, Department of Information Science, College of Information, University of North Texas (UNT), office phone: (940)565-3776, email: oksana.Zavalina@unt.edu

You are being asked to participate in a research study. Taking part in this study is voluntary. The investigators will explain the study to you and will answer any questions you might have. It is your choice whether or not you take part in this study. If you agree to participate and then choose to withdraw from the study, that is your right, and your decision will not be held against you.

The research study examines the status of the organization of knowledge in libraries and museums in Arabian Gulf countries and attitudes towards an idea of creating a large scale portal that aggregates digital collection from different institutions to help users access information more easily.

Your participation in this research study involves answering research team questions about information organization in your institution's digital collection management and feedback from the users related to it. The online interview is expected to be sent to you by email. More details will be provided in the next section.

You might want to participate in this study if you are willing to share your views on and experiences with organizing information in your institution and are interested in contributing to developing better understanding of the current status of organization of knowledge in libraries and museums in Arabian Gulf countries. However, you might not want to participate in this study if you do not have the time to answer interview questions.

You may choose to participate in this research study if you are involved in making and/or implementing decisions on managing organization of information in your institution, for example in relation to descriptive metadata, etc.

There are no reasonably foreseeable risks or discomforts to you if you choose to take part in this study. You will not receive compensation for participation.

DETAILED INFORMATION ABOUT THIS RESEARCH STUDY: The following is more detailed information about this study, in addition to the information listed above.

PURPOSE OF THE STUDY: The purpose of this study is to examine the status of the organization of knowledge in libraries and museums in Arabian Gulf countries, and attitudes towards an idea of creating a large digital library that brings together digital collections from different institutions (similar to the Digital Public Library of America and Europeana) to help

users access information more easily. As part of this study, the researcher will analyze similarities and differences in how information is organized and presented to the users in libraries and museums in Arabian Gulf countries. Results of this examination are expected to provide a basis for mappings and crosswalks to facilitate interoperability and create a centralized digital library for the Arabian Gulf countries, which will harvest metadata from participating institutions from participating institutions.

TIME COMMITMENT: Participation in this study is expected to last approximately one hour.

STUDY PROCEDURES: You will participate in the following activity: Email interview via your email address. The questions we would like you to discuss in your email message to us will be sent to you after you agree to participate in this study. You may skip questions that may make you uncomfortable. We will synthesize data collected from answers to understand issues related to organization of knowledge in libraries and museums in Arabian Gulf countries.

POSSIBLE BENEFITS: If you choose to participate, there will be no personal benefits that have direct impact on you. However, your participation in this study may benefit the community of libraries and museums employees and users through improving information organization in libraries and museums among the Arabian Gulf countries that may result from this and related studies.

POSSIBLE RISKS/DISCOMFORTS: Participation in this online interview involves risks to confidentiality similar to a person's everyday use of the internet and as will all online communications, there is always a risk of breach of confidentiality. However, if you do experience any discomfort, please inform the research team.

COMPENSATION: No compensation will be offered for participation in this study. There are no alternative activities offered for this study.

CONFIDENTIALITY: Efforts will be made by the research team to keep your personal information private, including research study records, and disclosure will be limited to people who have a need to review this information. All electronic data collected from this study will be stored in a secure location on the UNT campus and/or a secure UNT server for at least three (3) years past the end of this research in the password-protected computer in Principal Investigator's office. Research records will be labeled with a code and the master key linking names with codes will be maintained in a separate and secure location.

You will be asked to provide the names of other potential recruits, but you have the right to decline to provide this information. The researcher will maintain confidentiality when you suggest other persons for inclusion in the research.

The results of this study may be published and/or presented without naming you as a participant. The data collected about you for this study may be used for future research studies that are not described in this consent form. If that occurs, an Institutional review Board (IRB) would first evaluate the use of any information that is identifiable to you, and confidentiality protection would be maintained.

While absolute confidentiality cannot be guaranteed, the research team will make every effort to protect the confidentiality of your records, as described here and to the extent permitted by law. In addition to the research team, the following entities may have access to your records, but only on a need-to-know basis: the reviewing IRB, and sponsors of the study.

Confidentiality will be maintained to the degree possible given the technology and practices involved when online mailing. Your participation in this study involves risks to confidentiality similar to a person's everyday use of the internet.

CONTACT INFORMATION FOR QUESTIONS ABOUT THE STUDY: If you have any questions about the study you may contact Saleh Aljalahmah at salehaljalahmah@my.unt.edu or Dr. Oksana Zavalina at oksana.Zavalina@unt.edu . Any questions you have regarding your rights as a research subject, or complaints about the research may be directed to the University of North Texas Office of Research Integrity and Compliance at 940-565-4643, or by email at untirb@unt.edu.

CONSENT:
Your signature below indicates that you have read or have had read to you all of the above.

- You confirm that you have been informed about the possible benefits, risks, and/or discomforts of the study.
- You understand that you do not have to take part in this study and your refusal to participate or your decision to withdraw will involve no penalty or loss of rights or benefits.
- You understand your rights as a research participant and you voluntarily consent to participate in this study; you also understand that the study personnel may choose to stop your participation at any time.
- By signing, you are not waiving any of your legal rights.

Please sign below if you are at least 21 years of age and voluntarily agree to participate in this study.

SIGNATURE OF PARTICIPANT DATE

*If you agree to participate, please send a signed copy of this form as an email attachment to the research team and please keep a copy for your records.

APPENDIX D

IRB-APPROVED INFORMED CONSENT FORM FOR INTERVIEW DATA COLLECTION: ARABIC-LANGUAGE VERSION

فريق البحث :

الباحث الرئيسي : صالح الجلاهمه - قسم علوم المعلومات - كلية المعلومات - جامعة شمال تكساس (University of North Texas) - هاتف (١٥٤١٥١٤٠٥٠٣+) - البريد الإلكتروني

Salehaljalahmah@my.unt.edu

المشرف المتابع : د. اوكسانا زافالينا - قسم علوم المعلومات - كلية المعلومات - جامعة شمال تكساس (University of North Texas) - هاتف (١٩٤٠٥٦٥٣٧٧٦+) - البريد الإلكتروني

Oksana.Zavalina@unt.edu

ندعوك للمشاركة في بحث علمي بشكل تطوعي تام. الباحثين سيقومون بشرح الدراسة لك و سيقدمون اجوبة عن أي سؤال ترغب بطرحه. لك كامل الحرية في اختيار المشاركة من عدمه في هذا البحث. كما يحق لك الانسحاب من المشاركة في الدراسة اثناء المقابلة متى ما رغبت بذلك.

أنت مدعوا للمشاركة في بحث يتعلق بتحليل و دراسة وضع تنظيم المعرفة في المؤسسات العلمية في دول الخليج العربي بغرض اقتراح فكرة إنشاء فهرس ضخم يضم سجلات مصادر المعلومات من مختلف المؤسسات كما هو الحال في المكتبة الرقمية العامة للولايات المتحدة الأمريكية و مكتبة أوروبا (يوروبيانا) و ذلك لما فيه من فائدة قيمة للمستفيدين و تمكينهم من الوصول إلى المعلومات بصوره أسهل و اسرع.

مشاركتك في هذا البحث تتطلب منك الإجابة على مجموعة من الأسئلة التي تتعلق بتنظيم المعرفة في أنظمة مؤسستكم و طرق تفاعلكم مع المستفيدين. الأسئلة ستوجه عن طريق مقابلة إلكترونية عبر البريد الالكتروني الوقت المتوقع لها يتراوح بين ٣٠ دقيقة إلى ساعة. مزيد من المعلومات في الفقرة القادمة.

نسعى من مشاركتك في البحث بإن تشارك معنا نظرتك و خبرتك في تنظيم المعرفة في مؤسستك من اجل تحسين الطرق المتبعة في دول الخليج العربي لتنظيم المعرفة و الارتقاء بها. كما انه يحق لك عدم المشاركة إذا كنت تعتقد بأنك لا تملك الوقت الكافي لذلك.

نثمن مشاركتك في البحث خصوصا إذا كنت من اصحاب القرار أو المؤثرين على القرار المتعلق بتنظيم المعرفة في مؤسستك. لاسيما القرارات المتعلقة بكتابة الميتاديتا أو خيارات البحث و التصفح أو لوائح تنظيم المعرفة .. إلخ.

لن يكون هناك ضغوط أو مخاطر على حضرتك في حال قررت المشاركة في هذا البحث. كما ان مشاركتك ستكون بشكل تطوعي تام بدون أي مقابل.

معلومات تفصيلية عن البحث : المعلومات القادمة هي معلومات تفصيلية عن البحث تحمل تفاصيل ادق مما تم ذكره سابقا.

الغرض من البحث : الغرض من البحث هو دراسة و فهم وضع و مستوى تنظيم المعرفة في المؤسسات العلمية في دول الخليج العربي بغرض اقتراح إنشاء فهرس شامل يشمل سجلات المصادر لمختلف هذه المؤسسات و الذي سيوفر الوقت و الجهد للمستفيد في ايجاد المعلومة. جزء من هذا البحث سيقوم الباحث بايجاد نقاط التوافق و الاختلاف في تنظيم المعرفة بين هذه المؤسسات في دول الخليج العربي. نتيجة هذه العلمية سترسم خريطة عمل واضحة لإمكانية مشاركة سجلات مصادر المعلومات في هذه المؤسسات عن طريق عملية (حرث الميتاديتا) و التي تتم من خلال مشاركة الميتاديتا بين هذه المؤسسات في جهة واحدة رئيسية.

تعهد الوقت : نتعهد بأن المشاركة في هذا البحث لن تستمر اكثر من ساعة.

إجراءات المشاركة في البحث : المشاركة في البحث ستتم عن طريق البريد الإلكتروني الخاص بكم لإجراء مقابلة إلكترونية لمدة تتراوح بين ٣٠ دقيقة إلى ساعة للإجابة على أسئلة المقابلة. سيتم إرفاق أسئلة المقابلة في بريد إلكتروني بعد موافقتكم الأولية على المشاركة في البحث حتى يتسنى لكم قراءتها و الإجابة عليها. لك الحق في عدم الإجابة عن أي سؤال تعتقد بأنه لا يناسبك. سيقوم الباحث بتحليل المعلومات التي تم تجميعها في المقابلة بوقت لاحق حتى يحصل على اجوبة لأسئلة البحث العلمي.

نحيطكم علما بأن التسجيل سيتم حفظه بشكل سري تام في جهاز الباحث طوال فترة البحث.

مميزات متوقعة من المشاركة في البحث : إذا قررت المشاركة في البحث فلن تكون هناك مميزات مباشرة شخصيا لك لكن مشاركتك ستساهم بتطوير مجتمعك و المؤسسات العلمية في دول الخليج العربي. ستساهم برفع كفاءة تنظيم المعرفة في دول الخليج العربي عن طريق مشاركتك في هذا البحث أو أي باحث سيستخدم هذا البحث لاحقا.

مخاطر متوقعة من المشاركة في البحث : المشاركة في مقابلة إلكترونية لمدة تتراوح بين ٣٠ دقيقة إلى ساعة قد تسبب شعور عدم ارتياح للبعض مشابه لشعور استخدام الإنترنت لفترة طويلة. لكن إن شعرت بهذا الشعور اثناء المقابلة الرجاء إبلاغ الباحث مباشرة.

المقابل للمشاركة في هذه الدراسة : لا يوجد مقابل للمشاركة في هذا البحث لأن المشاركة تطوعية بشكل تام.

الخصوصية : فريق البحث سيحرص بشكل كامل و تام على خصوصية معلومات المشاركين في البحث ولن يتم عرض المعلومات لأي شخص خارج فريق البحث. كل المعلومات المتعلقة في المشاركين بهذا البحث ستحفظ بشكل سري في جامعة شمال تكساس لمدة ٣ سنوات من إنشاء الدراسة. المعلومات ستكون محفوظة بنظام حماية يتطلب كلمة مرور لن يعرفها سوى فريق البحث. المعلومات و السجلات ستكون مرقمة بنظام أحرف عشوائية و ربط الأحرف بالأسماء الحقيقة للمشاركين سيكون محفوظ لدى الباحث فقط بمكان آمن.

سيتم طلب منك اسماء لأشخاص أخرين تعتقد بإنهم سيقدمون إضافة للبحث، لكن لك كامل الحق في رفض تقديم أي اسم إذا كنت تمانع ذلك. في حال تم تقديم أي اسم، سيتحفظ فريق البحث عن ذكر اسمك للشخص إذا تم عمل مقابلة معه.

نتائج البحث و سيتم نشرها و تقديمها بدون ذكر اسم المشاركين في المقابلات لحفظ خصوصيتهم. المعلومات التي ستقدمها في البحث قد يتم استخدامها في ابحاث اخرى تستعين بهذا البحث. في حال تم ذلك سيتم على من يستخدم المعلومات الحصول على موافقة IRB(و هي جهة علمية محايدة تحافظ على سرية معلومات البشر في الدراسات العلمية)

فريق العمل سيبذل قصار جهده للمحافظة على سرية و خصوصية معلوماتك الشخصية و هويتك بشكل قانوني. فقط الباحثين و مؤسسة IRBهم المخولين لطلب و الرجوع لمعلوماتكم الشخصية.

حفظ خصوصيتك و سريتك سيتم عن طريق استخدام أفضل وأمن وسائل التكنولوجيا. مشاركتك في هذا البحث لا تحمل مخاطر غير اعتياديه على معلوماتك الشخصية.

معلومات التواصل لمزيد من المعلومات : اذا كان لديك أي استفسار يتعلق بالمشاركة في هذا البحث الرجاء التواصل مع الباحث صالح الجلاهمه عبر بريده الالكتروني Salehaljalahmah@my.unt.edu او الدكتوره المشرفة اوكسانا عبر بريدها الالكتروني Oksana.Zavalina@unt.edu . و إذا كانت لديك اسئلة تتعلق بحقوقك كمشارك في الدراسه او شكوى بإمكانك التواصل مع مكتب البحث في جامعة شمال تكساس عبر الهاتف ١٩٤٠٥٦٥٤٦٤٣+ او عبر البريد الالكتروني untirb@unt.edu.

الموافقة :

- توقيعك أسفل هذه الورقة يؤكد بأنك قمت بقراءة كل ما سبق ذكره.
- تؤكد بأن تم إبلاغك بالمخاطر المتوقعة من المشاركة في البحث.
- تعلم بأن المشاركة في هذا البحث هي تطوعية و ليس إلزامية، و امتناعك عن المشاركة لن يضرك بأي شكل من الأشكال.
- تعلم حقوقك كمشارك في البحث بشكل تطوعي و يحق لك أو للباحث إيقاف مشاركتك في البحث في إي وقت.
- بتوقيعك هذه الورقة أنت لا تتنازل عن حق قانوني يتعلق بك.

من فضلك قم بالتوقيع أسفل هذه الورقة إذا كنت فوق سن ٢١ و ترغب بالمشاركة في البحث بشكل تطوعي.

توقيع المشارك
التاريخ

* إذا كنت موافق على المشاركة في البحث برجاء تزويد فريق البحث بنسخه موقعة من هذه الورقة. و يحق لك الاحتفاظ بنسخة منها.

APPENDIX E

IRB-APPROVED INVITATION EMAIL FOR INTERVIEW DATA COLLECTION:

ENGLISH-LANGUAGE VERSION

Subject line: Invitation to participate in research interview

Dear [NAME],

We are writing to invite you to participate in the study of organization of knowledge in libraries and museums in Arabian Gulf countries because you currently have experience managing the institution's digital collection. This study is being conducted as part of the doctoral dissertation of Mr. Saleh Aljalahmah who is a PhD candidate at the University of North Texas Information Science Department. The purpose of this study is to identify the information organization methods and techniques currently used in libraries and museums in Arabian Gulf countries (Kuwait, Qatar, and Oman).

Participation in this research includes an interview about your experience and thoughts related to information organization in libraries and/or museums in Arabian Gulf countries. Your participation will support development of understanding of information organization in digital projects in Arabian Gulf Countries and is expected to help in building collaborative initiatives.

If you would like to participate in the research or have any questions, we can be reached at salehaljalahmah@my.unt.edu. The interview questions will be sent to you by email and it will take you approximately 30 minutes to 1 hour to answer these questions. Alternatively, if you know of someone else at your institution who makes decisions related to information organization, could you please refer us to this person?

Thank you for your time.

Participation is completely voluntary. You may choose not to respond to any questions that you do not feel comfortable responding to.

Research team
Saleh Aljalahmah salehaljalahmah@my.unt.edu - Doctoral Candidate
Dr. Oksana L. Zavalina Oksana.Zavalina@unt.edu - Dissertation committee chair
University of North Texas

APPENDIX F

IRB-APPROVED INVITATION EMAIL FOR INTERVIEW DATA COLLECTION: ARABIC-LANGUAGE VERSION

الموضوع: دعوة للمشاركة في مقابلة لبحث علمية
السيد/ السيدة :
طلب مقابلة علمية : مصنفين معلومات

نتقدم إليكم بطلب المشاركة في دراسة علمية تحت عنوان " وضع تنظيم المعرفة في المؤسسات العلمية في دول الخليج العربي " و ذلك لما تمتلكونه من خبرة و تجربة عميقة في إدارة نظام و مصادر مؤسستكم. المشاركة في هذه الدراسة ستدعم طالب الدكتوراه صالح الجلاهمه و الذي هو طالب في جامعة شمال تكساس (University of North Texas) الولايات المتحدة الأمريكية بقسم علوم المعلومات. الهدف من هذه الدراسة هو التعرف و تحديد وضع تنظيم المعرفة المتبع في دول الخليج العربي و الذي قد يساهم في بناء مقترح لإنشاء فهرس رئيسي يتيح للمستفيدين البحث من خلاله على مصادر من مختلف المؤسسات العلمية في مؤسسات دول الخليج العربي (الكويت - قطر - عمان).
مشاركتكم في هذا البحث العلمي ستساهم بتطوير تنظيم المعرفة في المؤسسات العلمية حول الخليج العربي و تساهم ببناء منظومة شاملة لمصادر هذه المؤسسات العلمية.
المشاركة في هذه الدراسة يتطلب من حضرتكم الموافقة على إجراء مقابلة علمية رسمية بإستخدام البريد الإلكتروني تتعلق بخبرتكم و تجربتكم في تنظيم المعرفة في مؤسستكم. الوقت المتوقع للمقابلة يتراوح بين ٣٠ دقيقة إلى ساعة كحد أقصى للإجابة على كافة الأسئلة.

إذا كنت ترغب في المشاركة أو لديك أسئلة أخرى بإمكانك التواصل مع الباحث بشكل مباشر عن طريق بريده الإلكتروني (salehaljalahmah@my.unt.edu). و إن كانت لديكم ترشيحات أخرى لأشخاص لديهم خبرة موازيه في المجال برجاء الإشارة إليهم.

شكرا لوقتك.

المشاركة في هذه الدراسة هي تطوعية بشكل تام و يحق للمشارك عدم الإجابة على كامل الأسئلة أو بعضها في حال شعوره بعدم الارتياح.

فريق الدراسة
salehaljalahmah@my.unt.eduصالح الجلاهمه - طالب دكتوراه
oksana.Zavalina@unt.eduد. اوكسانا زافلينا - الدكتور المشرف

REFERENCES

Alimohammadi, D. (2004). Measurement of the presence of keywords and description meta-tags on a selected number of Iranian Websites. *Online Information Review, 28*(3), 220-223.

Allen, R., Cole, T., & Schmitz, J. (2002). Creating a digital library outreach information system planning at the University of Illinois. *Journal of Agricultural & Food Information, 4*(2), 3-22.

Al-Nawas, A. (2017). *A view of the conditions of Arabic literature in the Nordic region*. Culture for All Service.

Alshenqeeti, H. (2014). Interviewing as a data collection method: A critical review. *English Linguistics Research, 3*(1), 39-45.

Andrade, M. C., & Baptista, A. A. R. P. (2015). The use of application profiles and metadata schemas by digital repositories: findings from a survey. In *International conference on Dublin Core and metadata applications* (pp. 146-157).

Agocs, A., & Le Goff, J. M. (2018). A web service based on RESTful API and JSON Schema/JSON Meta Schema to construct knowledge graphs. In *2018 international conference on computer, information and telecommunication systems (CITS)* (pp. 1-5). IEEE.

Avram, H. D. (1969). *The MARC Pilot Project: Final report on a project sponsored by the Council on Library Resources*. Library of Congress.

Baca, M. (2016). *Introduction to metadata*. Getty Publications.

Bade, D. W. (2002). *The creation and persistence of misinformation in shared library catalogs: Language and subject knowledge in a technological era.* Graduate School of Library and Information science. University of Illinois at Urbana-Champaign.

Bade, D. W. (2004). *The theory and practice of bibliographic failure, or misinformation in the information society*. Chuluunbat.

Bade, R. (2010). *Digital library project aims at the Middle East.* Roll Call Retrieved from http://o-search.proquest.com.biance.penlib.du.edu/docview/324271156?accountid=14608

Bailey-Hainer, B., & Urban, R. (2004). The Colorado digitization program: A collaboration success story. *Library Hi Tech, 22*(3), 254-262.

Baker, T., & Coyle, K. (2009). *Guidelines for Dublin Core application profiles*. Retrieved from https://www.dublincore.org/specifications/dublin-core/profile-guidelines/

Beall, J. (2006). *Metadata schemes points of comparison*. (Unpublished). Retrieved from http://eprints.rclis.org/7162/

Beall, J., & Mitchell, J. S. (2010). History of the representation of the DDC in the MARC classification format. *Cataloging & Classification Quarterly, 48*(1), 48-63.

Bennett, R., Lavoie, B. F., & O'Neill, E. T. (2003). The concept of a work in World Cat: An application of FRBR. *Library Collections, Acquisitions, and Technical Services, 27*(1), 45-59.

Bhatt, R. K. (2011). Relevance of Ranganathan's laws of library science in library marketing. In *Library philosophy and practice* (pp. 1-7).

Bianchini, C., & Guerrini, M. (2007). International cataloguing tradition and Italian rules: Common ground and specific features. *Cataloging & Classification Quarterly, 44*(1-2), 131-150.

Bizer, C., Heath, T., & Berners-Lee, T. (2011). Linked data: The story so far. In *Semantic services, interoperability and web applications: Emerging concepts* (pp. 205-227). IGI Global.

Bowen, J. (2008). Metadata to support next-generation library resource discovery: Lessons from the extensible catalog, phase 1. *Information Technology and Libraries, 27*(2), 6-19.

Borbinha, J. L. (2004). An approach to creating a national digital library. *International Journal on Digital Libraries, 4*(1), 19-22.

Bruce, T. R., & Hillmann, D. I. (2004). The continuum of metadata quality: Defining, expressing, exploiting. In D Hillmann and E.L. Westbrook (Eds.), *Metadata in practice.* Chicago: American Library Association.

Budanović, M., & Žumer, M. (2018). The catalogers' thought process: A comparison of formal and informal context. *Cataloging & Classification Quarterly, 56*(5-6), 507-529.

Burke, M., & Zavalina, O. L. (2019). Exploration of information organization in language archives. *Proceedings of the 82nd Association for Information Science and Technology Annual Meeting, 56*(1), 364-367.

Burke, M., & Zavalina, O. L. (2020). Identifying challenges for information organization in language archives: Preliminary findings. In *iConference 2020* (pp. 622-629).

Calhoun, K. (2014). *Exploring digital libraries*. Chicago: Neal-Schuman.

Caplan, P., & Haas, S. (2004). Metadata rematrixed: Merging museum and library boundaries. *Library Hi Tech, 22*(3), 263-269.

Carter, N., Bryant-Lukosius, D., DiCenso, A., Blythe, J., & Neville, A., J. (2014). The use of triangulation in qualitative research. *Oncology Nursing Forum, 41*, 545–547. doi: 10.1188/14.ONF.545.547

Chan, L. M., & Zeng, M. L. (2006). Metadata interoperability and standardization: A study of methodology, part I. *D-Lib Magazine, 12*(6), 1082-9873. Retrieved from: dlib.org/dlib/june06/chan/06chan.html.

Chen, S. Y., Li, Z. F., & Chen, G. L. (2014). Design and implementation of an open-source content management system. In *Applied mechanics and materials* (Vol. 651, pp. 1821-1826). Trans Tech Publications.

Christel, M. G. (2009). Automated metadata in multimedia information systems: Creation, refinement, use in surrogates, and evaluation. *Synthesis Lectures on Information Concepts, Retrieval, and Services, 1*(1), 1-74.

Chowdhury, G., & Chowdhury, S. (2002). *Introduction to digital libraries*. Facet Publishing.

Cocciolo, A. (2014). Challenges to born-digital institutional archiving: The case of a New York art museum. *Records Management Journal, 24*(3), 238-250.

Copeland, A. (2014). The use of personal value estimations to select images for preservation in public library digital community collections. *Future Internet, 6*(2), 359-377.

Coyle, K. (2012). Metadata elements. *Library Technology Reports, 48*(4), 15-26.

De Groat, G. (2015). A history of video game cataloging in US libraries. *Cataloging & Classification Quarterly, 53*(2), 135-156.

Dewey, M. (1876). *Classification and subject index for cataloging and arranging the books and pamphlets of a library*. Project Gutenberg eBook.

Diao, J. (2018). Conceptualizations of catalogers' judgment through content analysis: a preliminary investigation. *Cataloging & Classification Quarterly, 56*(4), 298-316.

Dushay, N., & Hillmann, D. I. (2003,). Analyzing metadata for effective use and re-use. In *International conference on Dublin Core and metadata applications* (pp. 161-170).

Duval, E., Hodgins, W., Sutton, S., & Weibel, S. L. (2002). Metadata principles and practicalities. *D-Lib Magazine, 8*(4), 1-10.

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. (2000). Metadata and the future of cataloging. *Library Computing, 19*(3/4), 180-194.

El-Sherbini, M. (2015). Multilingual subject retrieval: Bibliotheca Alexandrina's subject authority file and linked subject data. In *Data science, learning by latent structures, and knowledge discovery* (pp. 535-546). Springer, Berlin, Heidelberg.

Expanded Thesaurus (Maknaz). (n.d.). Retrieved from http://en.maknaz.org

Farneth, D. (2016). How can we achieve GLAM? Understanding and overcoming the challenges to integrating metadata across museums, archives, and libraries: Part 2. *Cataloging & Classification Quarterly, 54*(5-6), 292-304.

Gartner, R. (2016). *Metadata: Shaping knowledge from antiquity to the semantic web*. Springer.

Geckle, B. J., & Nelson, D. N. (2017). Classifying librarians: Cataloger, taxonomist, metadatician? *The Serials Librarian, 72*(1-4), 57-64.

Gilliland, A. J. (2016). Setting the stage. In M. Baca (Ed.), *Introduction to metadata* (pp. 1-19). Los Angeles: The Getty Research Institute.

Greenberg, J. (2002). Metadata generation: Processes, people and tools. *Bulletin of the American Society for Information Science and Technology, 29*(2), 16.

Gregory, L., & Williams, S. (2014). On being a hub: Some details behind providing metadata for the Digital Public Library of America. *D-Lib Magazine, 20*(7/8), 25-32.

Guinchard, C. (2002). Dublin Core use in libraries: A survey. *OCLC Systems & Services: International Digital Library Perspectives, 18*(1), 40-50.

Guenther, R. S. (2004). Using the metadata object description schema (mods) for resource description: Guidelines and applications. *Library Hi Tech, 22*(1), 89-98.

Gutiérrez, A., & Ávila, D. (2014). Critical organization of knowledge in mass media information systems. *Ko Knowledge Organization, 41*(3), 205-216.

Han, Y. (2004). Digital content management: the search for a content management system. *Library Hi Tech, 22*(4), 355-365.

Hartman, C., Belden, D., Reis, N. K., Gelaw Alemneh, D., Phillips, M., & Dunlop, D. (2005). Development of a portal to Texas history. *Library Hi Tech, 23*(2), 151-163.

Haynes, D. (2004). *Metadata for information management and retrieval* (Vol. 1). Facet Publishing.

Heery, R., and Patel, M. (2000). Application profiles: Mixing and matching metadata schemas. *Ariadne, 25*. Retrieved from http://www.ariadne.ac.uk/issue25/app-profiles/

Henttonen, P. (2009). A comparison of MoReq and SÄHKE metadata and functional requirements. *Records Management Journal, 19*(1), 26-36. doi:10.1108/09565690910937227

Hillman, D. (2005). *Using Dublin Core*. Retrieved from https://www.dublincore.org/specifications/dublin-core/usageguide/

Hillmann, D. I., Dunsire, G., & Phipps, J. (2013, September). Maps and gaps: strategies for vocabulary design and development. In *International conference on Dublin Core and metadata applications* (pp. 82-89).

Holloway, M. F. (1959). Patterns of library service in the Middle East. *Library Trends, 8* (October), 192-208.

Isaac, A., & Haslhofer, B. (2013). Europeana linked open data–data.europeana.eu. *Semantic Web, 4*(3), 291-297.

Jackson, A. S., Han, M. J., Groetsch, K., Mustafoff, M., & Cole, T. W. (2008). Dublin Core metadata harvested through OAI-PMH. *Journal of Library Metadata, 8*(1), 5-21.

Jones, E. (2017). The public library movement, the digital library movement, and the large-scale digitization initiative: Assumptions, intentions, and the role of the public. *Information & Culture, 52*(2), 229.

Joint, N. (2006). Digital library futures: Collection development or collection preservation? *Library Review, 55*(5), 285-290.

Kawano, H. (2012). Digital archive systems using CMS and gallery tools: Implementation of anthropological museum. *International Journal of Heritage in the Digital Era, 1*(1_suppl), 125-130.

Krause, E. M., Clary, E., Ogletree, A., & Greenberg, J. (2015). Evolution of an application profile: Advancing metadata best practices through the Dryad data repository. In *International conference on Dublin Core and metadata applications* (pp. 63-75).

LaRue, J. (2010). The visibility and invisibility of librarians. *Library Journal, 135*(19), 10.

Lee, J. H., Tennis, J. T., Clarke, R. I., & Carpenter, M. (2013). Developing a video game metadata schema for the Seattle Interactive Media Museum. *International Journal on Digital Libraries, 13*(2), 105-117.

Lesk, M. (2005). *Understanding digital libraries*. San Francisco, CA: Morgan Kaufmann.

Lewis, P. H., Martinez, K., Abas, F. S., Fauzi, M. F. A., Chan, S. C. Y., Addis, M. J., et al. (2004). An integrated content and metadata-based retrieval system for art. *IEEE Transactions on Image Processing, 13*(3), 302-313.

Library and Information Technology Association (U.S.)., & Gorman, M. (1990). *MARC format integration: Three perspectives*. Papers presented at the second national conference of the Library and Information Technology Association, October 2-6, 1988, Boston, Massachusetts. Chicago: American Library Association.

Lin, C. H., Hong, J. S., & Doerr, M. (2008). Issues in an inference platform for generating deductive knowledge: A case study in cultural heritage digital libraries using the CIDOC CRM. *International Journal on Digital Libraries, 8*(2), 115-132.

Lemos, A.A. B. de (1981). A portrait of librarianship in developing societies. University of Illinois *Graduate School of Library and Information Science Occasional Papers, 148* (June), 1-46.

London: Her Majesty's Stationery Office. (1985). Reprinted in Michael Carpenter and Elaine Svenonius (Eds.), *Foundations of cataloging: A sourcebook* (pp. 18–47). Littleton, CO: Libraries Unlimited.

Lopatin, L. (2010). Metadata practices in academic and non-academic libraries for digital projects: A survey. *Cataloging & Classification Quarterly, 48*(8), 716-742.

Lund, B., Agbaji, D., Tijani, S., & Omame, I. (2019). Evaluating knowledge organization in developed and developing countries: A comparative analysis of Dewey Decimal and Library of Congress classification scheme preference and use in the United States and Nigeria. *Technical Services Quarterly, 36*(3), 249-268.

Lynch, C. A. (1997). The Z39. 50 information retrieval standards. *D-lib Magazine, 3*(4).

Machovec, G. S. (1998). Metadata and the Dublin Core: Emerging standards for digitization projects. *Online Libraries and Microcomputers, 16*(12), 1-5.

Mahdi, S., & Hariri, N. (2012). A comparative study on the indexing and ranking of the content objects including the MARCXML and Dublin Core's metadata elements by general search engines. *The Electronic Library, 30*(4), 480-491. doi:10.1108/02640471211252193

Mansor, Y., & Ramdzan, E. (2014). RDA perceptions among Malaysian catalogers. *Library Review, 63*(3), 176-188.

Marcum, D. (2014). Archives, libraries, museums: Coming back together? *Information & Culture: A Journal of History, 49*(1), 74-89.

Maurya, R. N. (2011). Digital library and digitization. *International Journal of Information Dissemination and Technology, 1*(4), 228-231.

McKenna, L., Debruyne, C., & O'Sullivan, D. (2018, May). Understanding the position of information professionals with regards to linked data: A survey of libraries, archives and museums. In *Proceedings of the 18th ACM/IEEE joint conference on digital libraries* (pp. 7-16).

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.

Mescan, S. (2004). Why content management should be part of every organization's global strategy: To meet the challenges of creating, using, and sharing content in a global marketplace, organizations should consider a content management system. *Information Management Journal, 38*(4), 54-58.

Miller, S. J. (2011). *Metadata for digital collections: A how-to-do-it manual*. New York, N.Y.: Neal-Schuman Publishers.

Moen, W. E., Stewart, E. L., & McClure, C. R. (1997). *The role of content analysis in evaluating metadata for the US government information locator service (GILS): Results from an exploratory study*. Retrieved from https://digital.library.unt.edu/ark:/67531/metadc36312/

Moulaison Sandy, H., & Dykas, F. (2016). High-quality metadata and repository staffing: Perceptions of United States based open DOAR participants. *Cataloging & Classification Quarterly, 54*(2), 101-116.

Nemeth, E. (2009). Metadata of art and architecture research databases: The value of abstracts and index terms to academics, librarians, and museum professionals. *Journal of Library Metadata, 8*(4), 293-313.

O'Dell, A. J. (2013). *An examination of the metadata captured about artifacts of bookbinding by United States research libraries* (Publication No.1537361) [Master's thesis, Corcoran College of Art). ProQuest Dissertations Publishing.

Palmer, C. L., Zavalina, O. L., & Mustafoff, M. (2006). Trends in metadata practices: A longitudinal study of collection federation. In *Proceedings of the 7th ACM/IEEE-CS joint conference on digital libraries* (pp. 386-395). ACM.

Panizzi, Anthony. (1850). Mr. Panizzi to the Right Hon. the Earl of Ellesmere. British Museum, 29 January 1828. In *Appendix to the Report of the Commissioners Appointed to Inquire into the Constitution and Management of the British Museum* (pp. 378–395).

Park, J., & Childress, E. (2009). Dublin Core metadata semantics: An analysis of the perspectives of information professionals. *Journal of Information Science, 35*(6), 727-739.

Park, J., & Tosaka, Y. (2010). Metadata creation practices in digital repositories and collections: Schemata, selection criteria, and interoperability. *Information Technology and Libraries, 29*(3), 104-116.

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.

Park, J., Tosaka, Y., Maszaros, S., & Lu, C. (2010). From metadata creation to metadata quality control: Continuing education needs among cataloging and metadata professionals. *Journal of Education for Library and Information Science, 51*(3), 158-176.

Phillips, M., Zavalina, O.L., & Tarver, H. (2020). Exploring the utility of metadata record graphs and network analysis for metadata quality evaluation and augmentation. *International Journal of Metadata, Semantics, and Ontologies, 7*(3), 1-13.

Phillips, M.E., Zavalina, O.L., & Tarver, H. (2019). Using metadata record graphs to understand digital library metadata. *Proceedings of the international conference and workshop on Dublin Core and metadata applications* (pp. 49-58).

Pierre, M. & LaPlant, W.P. Jr. (1998). *Issues in crosswalking content metadata standards*. Bethesda, MD: NISO Press

Qatar Digital Library. (n.d.). Retrieved May 20, 2021, from http://www.qdl.qa/en

Ravandi, N., & Zeinolabedini, H. (2013) Comparison of used metadata elements in digital libraries in Iran with Dublin Core standard. *International Journal of Research in Engineering and Technology, 2*(3), 398-403

Ray, J. (2017). Digital curation in museums. *Library Hi Tech, 35*(1), 32-39.

Robinson, B. (2016). Addressing bias in the cataloging and classification of Arabic and Islamic materials: Approaches from domain analysis. In A.B. Click et al. (Eds.), *Library and information science in the Middle East and North Africa* (pp. 255-269). Munich: De Gruyter Saur.

Robertson, R. J. (2005). Metadata quality: Implications for library and information science professionals. *Library Review, 54*(5), 295-300.

Roel, E. (2005). The MOSC project: Using the OAI-PMH to bridge metadata cultural differences across museums, archives, and libraries. *Information Technology and Libraries, 24*(1), 22-24.

Roy, S. G., Sutradhar, B., & Das, P. P. (2017). Large-scale metadata harvesting—tools, techniques and challenges: A case study of National Digital Library (NDL). *World Digital Libraries-An International Journal, 10*(1), 1-10.

Ruthven, I., & Chowdhury, G. G. (Eds.). (2015). *Cultural heritage information: Access and management* (Vol. 1). Facet Publishing.

Shreeves, S. L., Knutson, E. M., Stvilia, B., Palmer, C. L., Twidale, M. B., & Cole, T. W. (2005). Is quality metadata shareable metadata? The implications of local metadata practices for federated collections. In *Proceedings of the Association of College and Research Libraries (ACRL) 12th National Conference*. Minneapolis, MN.

Slavic, A. (2008). Use of the Universal Decimal Classification: A world-wide survey. *Journal of Documentation, 64*(2), 211-228.

SDL. (n.d.). Retrieved May 20, 2021, from https://sdl.edu.sa/sdlportal/en/publishers.aspx

Stegaeva, M. V. (2016). Cooperative cataloging: History and the current state. *Scientific and Technical Information Processing, 43*(1), 28-35.

Sokvitne, L. (2000), *An evaluation of the effectiveness of current Dublin Core metadata for retrieval.* Retrieved from: http://www.vala.org.au/vala2000/2000pdf/Sokvitne.PDF

Svenonius, E. (2000). *The intellectual foundation of information organization*. MIT Press.

Sweet, L. E., & Moulaison, H. L. (2013). Electronic health records data and metadata: challenges for big data in the United States. *Big Data, 1*(4), 245.

Taniguchi, S. (2013). Event-aware FRBR and FRAD models: Are they useful? *Journal of Documentation, 69*(3), 452-472.

Taylor, A. (2004). *The organization of information* (2nd ed). Westport, CN: Libraries Unlimited.

Tennant, R. (2002). *XML in libraries*. Edison, NJ.: Neal-Schuman Publishers.

Thiele, J. (2016). *Information access in rural areas of the United States: The public library's role in the digital divide and the implications of differing state funding models*. (Doctoral Dissertation, The University of Wisconsin-Milwaukee,2016) ProQuest Dissertations Publishing.

Thornburg, G., & Oskins, W. M. (2007). Misinformation and bias in metadata processing: Matching in large databases. *Information Technology and Libraries, 26*(2), 15-26.

Wallace, D. A. (2001). Archiving metadata forum: Report from the recordkeeping metadata working meeting, June 2000. *Archival Science, 1*(3), 253-269.

Weaver, M. (2007). Contextual metadata: Faceted schemas in virtual library communities. *Library Hi Tech, 25*(4), 579-594.

Woodley, M. S. (2016). Crosswalks, metadata harvesting, federated searching, met searching: Using metadata to connect users and information. In Baca M. (Ed.), *Introduction to metadata* (pp.1-25) Getty Research Institute.

Zare-Farashbandi, F., Ramezan-Shirazi, M., Ashrafi-Rizi, H., & Nouri, R. (2014). Hypertext Mark-up Language and Dublin Core metadata element set usage in websites of Iranian State Universities' libraries. *Journal of Education and Health Promotion*, 3:6.

Zavalina, O. L., Kizhakkethil, P., Alemneh, D. G., Phillips, M. E., & Tarver, H. (2015). Building a framework of metadata change to support knowledge management. *Journal of Information & Knowledge Management, 14*(01), DOI: 10.1142/S0219649215500057.

Zeng, M. L., & Chan, L. M. (2006). Metadata interoperability and standardization-A study of methodology, Part II. *D-Lib Magazine, 12*(6). Available at: http://www.dlib.org/dlib/june06/chan/06chan.html

Zeng, M. L., & Qin, J. (2008). *Metadata*. New York, NY: Neal-Schuman Publishers.

Zhang, X. (2004). Tutorial on metadata. In *Tutorials, 7th International Conference of Asian Digital Libraries (ICADL)* (pp. 13-17). Shanghai China: Shanghai Jiaotong University Library.

Žumer, M. (2018). IFLA Library Reference Model (IFLA LRM)—Harmonisation of the FRBR family. *Knowledge Organization, 45*(4), 310-318.