December 2020

Mapping the Metadata challenges in Libraries: A systematic review

Imran Ghaffar Sulehri
Pakistan Institute of Fashion and Design, igsulehri@hotmail.com

Nosheen Fatima Warraich Dr
University of the Punjab Lahore Pakistan, nosheen.im@pu.edu.pk

Mapping the Metadata challenges in Libraries: A systematic review

Imran Ghaffar Sulehri
PhD Scholar, Information Management Department, PU
igsulehri@hotmail.com

Dr. Nosheen Fatima Warraich
Associate Professor,
Information Management Department, PU

Abstract

Background. In the information and knowledge world, libraries always played their important role and found as early adopters of new techniques and technologies for dissemination of information.

Purpose. If we understand the metadata as a researcher's perspective, it is exploratory in its nature which provides guidance to the further data which is explanatory. There are many metadata challenges which affect the execution and accessibility of relevant data. These challenges must be recognized at one place so that LIS professionals having interest in metadata could be able to understand these challenges and hurdles concerning with libraries. So, this this study is being conducted to find out the challenges of metadata and bring these challenges synthetically from scattered literature for the readers.

Design/methodology/approach. To compete this study, a systematic literature review approach has been followed. Thirteen paper are selected to find out the challenges faced by the libraries concerning with the metadata.

Findings. In this systematic review 85 challenges were found from the scholarly published literature which are categorized into 19 categories according to their nature and likeliness. Further, general challenges and project based challenges are presented separately.

Practical implications. Through this study scattered challenges of metadata faced by the libraries are grouped together to strengthen the lacking information. This paper will add knowledge in the existing literature in form of comprehensiveness.

Key Words: Meta Data, Libraries, Digital Library, Review

Background

During the past few years the term metadata got prominent place in the field of library and information science (LIS). Metadata is defined as as data about data which leads towards the informative documents (Dashrath, 2014). W3 defined metadata as "Data about data and is used to both describe and find resources". Many researches have conducted studies on various aspects of metadata, its understanding, application and projects. Sugimoto (2005) stated that digital libraries have important infrastructure for knowledge sharing. Metadata research area is getting notable appreciation in LIS and researches are being conducted on various aspects od metadata. Gradmann (2009) prescribed that interoperability is the basic feature and the libraries must follow uniform standards to gain interoperability.

Published literature on metadata provides understanding about the nature and application but still the topic needs to be discovered more. There are many studies (00000) which prescribe various dimensions like metadata interoperability, metadata schemas, metadata creation and management etc. are being explored through research so that a comprehensive scholarly material may be presented for the readers and researchers in this area. Challenges in implementation of metadata play significant redundant role and many researchers defined various issues of metadata which are of serious hurdles for the successful completion of the implementation of metadata. Calarco, Conrad, Kessler

and Vandenburg (2014) discussed some issues related to metadata which are harmful for discovery. Literature also guides that metadata is being used in every field of life and funded projects are initiated as well offered for its application. We also find some challenges which were faced during the metadata projects (Challenges of Using Metadata in a Library Setting: the Collection and Management of Electronic Links (CAMEL) Project at Oregon State University and Challenges in Digital Libraries - Key Issues Learned from Metadata-Centric Projects at Tsukuba) and some research articles also discussed the issues related to metadata.

It has been observed that there are many issues linked to the metadata application in libraries. Therefore, it is pertinent to synthesize the issues, challenges or hurdles related to metadata application specifically with respect to libraries, so that the library leaders and professionals who interested to implement metadata infrastructure in their libraries may become aware aboutexpected challenges. It will definitely help them to consider these challenges prior to start the project of metadata implementation. Furthermore, it will be beneficial for them to plan out that how to inculcate these challenges for the successful application. Additionally through this study readers, students, concerned persons and professionals will be able to know about the issues related to metadata. This study will be conducted by following the systematic review approach keeping in view the research question of synthesizing the challenges, issues and hurdles concerning with the metadata.

This study attempts to explore the metadata challenges for libraries by flowing the principals of systematic literature review so that extracted challenges from scholarly

literature through this study may help LIS community for more understanding of metadata.

Research Question

Following was the research questions which will be countered in this study. What are the challenges concerning with metadata application in libraries?

Methodology

To complete this study, the method of systematic review of previously published research articles is adopted. The adoption of systematic review is better to pin point specific issues related to any phenomenon from previous literature. It also logically guide to the researchers and provide a benchmark that what is to include or exclude. Hence, following the essence of systematic review, we follow all the steps required to complete any study.

Search strategy

To search out the relevant literature, a comprehensive search strategy was defined and opted. Comprehensive searches over the times in Google Scholar and Library, Information Science and Technology Abstracts (LISTA) were accomplished according to the topic and its research questions. Following keywords and search strategy were devised to locate the relevant literature from the both databases. Metadata challenges library/ libraries, metadata issues library/ libraries, metadata AND library, metadata library, metadata problem library, metadata problems library, metadata setback library, metadata difficulty/ difficulties library, metadata hurdle/ hurdles/ hindrance library, metadata opportunities library.

Representation of Search Results through PRISMA Flow Diagram

Following is the representation of search results and final selection of the articles. In first round we found 1025 research articles from the both databases. In the second round we sorted out the irrelevant articles and excluded989 articles.. Then we excluded such searches which were only citations and at the end we excluded books, bibliographies, websites and articles published in other language than English. Hence, 13 studies prevailed according to settled criteria and their systematic review was conducted.

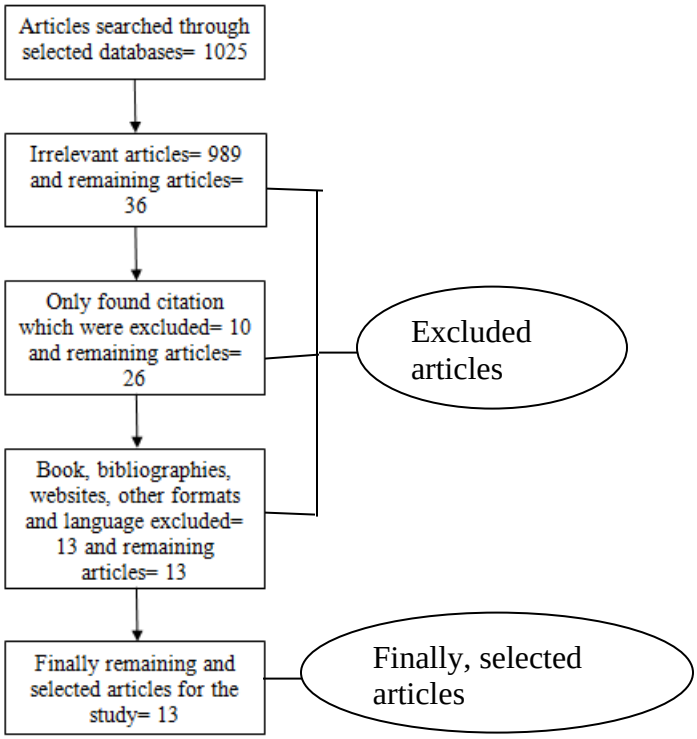

Figure. 1. PRISMA flow diagram of selection

Selection Criteria.

Following selection criteria were determine keeping mind the research question of the study.

Inclusion criteria of articles for this study as follows. All research articles published in journals or presented in conferences and available on Google Scholar as well as LISTA from 2000 to 2019 will be included for this study. Further, the articles which

describe metadata challenges of libraries in general or in a specific project of library/ libraries will also be the part of this review. Regarding language, the articles available in English language will be the part of this study.

Exclusion criteria for this study is as follows. Keeping in view the research question of the study following exclusion criteria was devised. Those articles which are not published in any research journal nor presented in any conference will not be the part of this study. Further, such kind of articles which published before 2000 will not be added for this study while articles other than English language will also be excluded. Moreover, books and general metadata challenges based articles which are not specifically concerning with libraries will not be the part of this study. The table 1 represents the systematically selected studies with challenges prescribed within them.

Table 1

Selected studies which defined metadata challenges with respect to libraries

Sr	***Study Title***	***Year of Publication***	***Metadata Challenges Faced by Libraries***
1	*An Analysis of the Named Entity Recognition Problem in Digital Library Metadata*	*2012*	*Lack of lexical evidence in text of metadata; Structured data for use in named entity recognition (NER) vary in each case*
2	*Challenges of Using Metadata in a Library Setting: the Collection and Management of Electronic Links (CAMEL) Project at Oregon State University*	*Published in 2000 and online published in 2013*	*Human and organizational issues; Unawareness about the use and strength of metadata systems; Using un-standardized metadata; Metadata skillful staff; Technical issues include local experts support for new project; compatibility with present access and mechanism and data; Lack of standards: Software tools; Maintenance*
3	*Challenges in Digital Libraries - Key Issues Learned from Metadata-Centric Projects at Tsukuba*	*2005*	*Interoperability and reuse of metadata schemas; Development and maintenance of metadata schemas; Typographical error; Inappropriate use of upper/ lower case letters; Assigning subject terms; Type and Granularity of Resources; Controlled vocabularies; Metadata schema sharing*
4	*Context and Meaning: The Challenges of*	*2004*	*Using different metadata terms for data elements affect results and*

	Metadata for a Digital Image Library within the University		*represent different functional uses; Specifying subject coverage; Standards; Meaning of values contained in elements; Variability among entered values; For image based data little amount of metadata is an issue for accessibility*
5	*Digitisation and Metadata Challenges: Experiences of the World Digital Library (Uganda)*	*2013*	*Metadata creation and management; Many metadata standards and schemas; Economics of metadata creation; Specialized knowledge requirement; Inconsistency; Continual evolution of standards; Lack of guidelines in describing information resources; Plagiarism; Inadequate and qualified metadata specialists; Inadequate IT infrastructure; Inadequate education and experienced digitization consultants*
6	*How Can We Achieve GLAM? Understanding and Overcoming the Challenges to Integrating Metadata across Museums, Archives, and Libraries: Part 2*	*2016*	*Uniformity in descriptive practices; Promise of linked open data; Controlled vocabularies; Creating* harmonious *conceptual reference model for description of metadata*
7	*Large-scale Metadata Harvesting—Tools, Techniques and Challenges: A Case Study of*	*2017*	*Harvesting of metadata; Untitled metadata; Junked Unicode characters; Incomplete harvesting; Connection time out; Multiple record harvesting;*

	National Digital Library (NDL)		*OAI index error; Curation of large scale harvested metadata*
8	*MASHing Metadata: Legacy Issues in OAI Harvesting From Three Digital Libraries*	*2013*	*Non-OAI accessible metadata; Variation in subject fields; Undocumented metadata aggregation; Normalization in union repository; Legacy issue of harvesting metadata; RDF triple describing item-collection relationship; Domain specific classification keywords*
9	*Metadata and Data Quality Problems in the Digital Library*	*2005*	*Creating metadata automatically holds some errors; Typographical error; Crosswalking metadata from one scheme/ format to another can be caused of source error; Metadata harvesting; Data transmission error; Incompatible data elements/ formats; Electronic metadata corruption during conversion to another scheme; Harvesting of metadata from multiple sources cause metadata varying structure, quality, content standard and schemes which make it inconsistence, unusable and unreliable.*
10	*Metadata Challenges in Library Discovery Systems*	*2014*	*Insufficient metadata; Inconsistency of metadata among disparate sources; Incorrect metadata; Unified index from different level of records which cause irrelevant results on the top; Linking of metadata with full text; Metadata connections with vendors, publishes and aggregators; Metadata*

			creation of hybrid documents; Normalization of data
11	*Metadata issues in Digital Libraries: key concepts and perspectives*	*2011*	*Use of different vocabularies for metadata; Metadata management; Interoperability; Models and schemas of metadata have setback to be organized organically in a single space of linked data*
12	*Moving Library Metadata toward Linked Data: Opportunities Provided by the eXtensible Catalog*	*2010*	*Conversion of MARC metadata into linked data; Mapping of data (Reuse the legacy MARC data); Use of single MARC record to describe more than one format or version; Difficulties in connecting some MARC fields (880&9XX) to linked data; Reuse of legacy metadata in other environment*
13	*Research Data and Repository Metadata: Policy and Technical Issues at the University of Sydney Library*	*2009*	*Loss of metadata granularity and inability to recreate the original records; Metadata would not be meaningful without contextual information provided by their native tags; Customize metadata schemas; OAI crosswalk; Hierarchical metadata schemas are not supported by DSpace; Less awareness about metadata preservation and technical aspect*

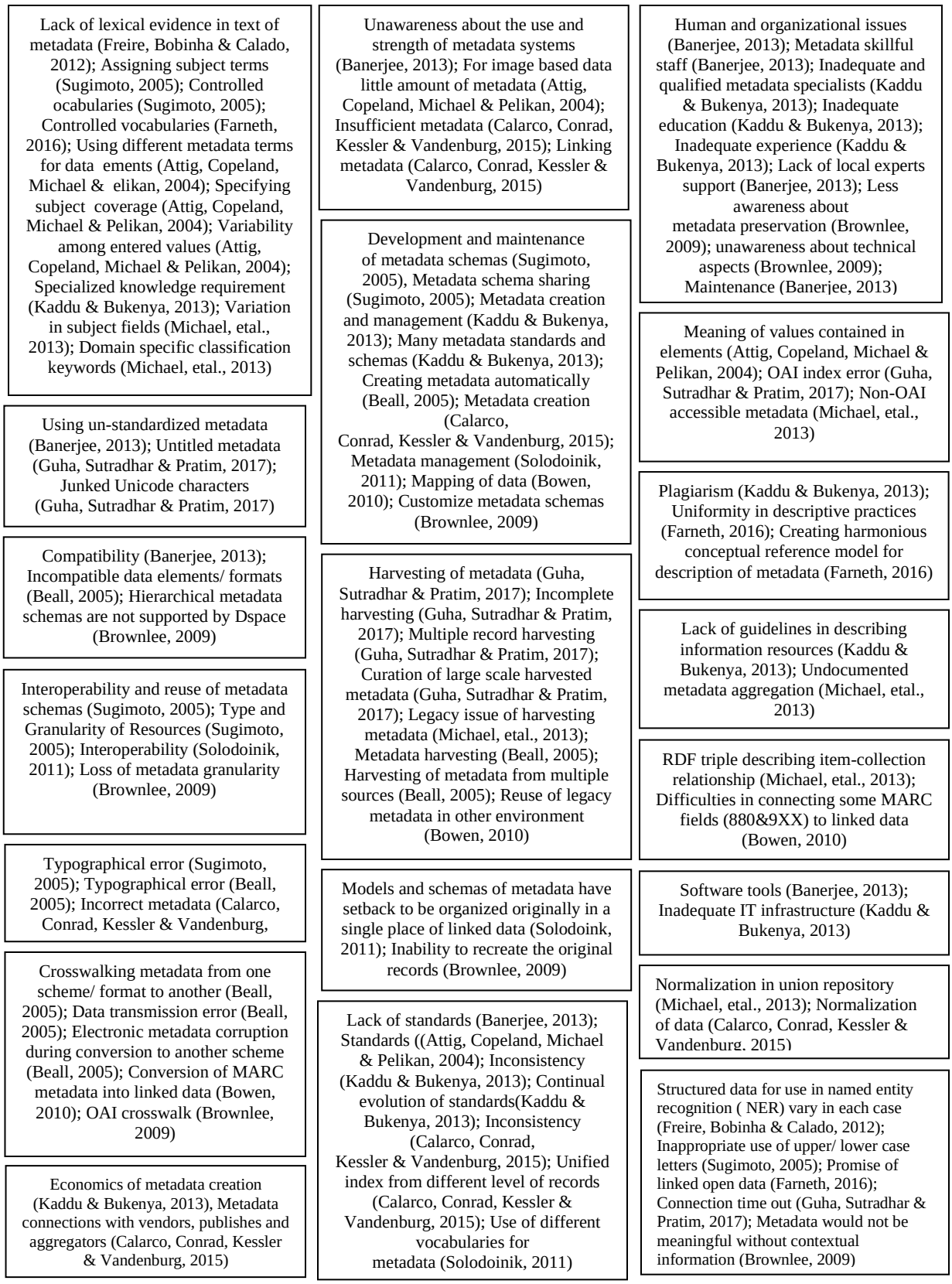

Figure 2. Metadata challenges extracted from the above mentioned selected studies

Results

There was diversity in the selected studies as some of the studies defined specific projects and some narrated general issues faced by digital or traditional libraries regarding metadata. Therefore, a large number of challenges as mentioned in (Figure 2) found through the literature. It is better for the readers and concerned professionals for understanding the issues related to metadata so that they may be well aware with the various types of challenges which can be raised during the metadata implementation in a specific context whether in digital or in traditional settings. Keeping in view the extracted challenges from the literature interested professionals in metadata implementation would be able to plan out in advance for encountering the expected issues.

The last dialog box consisted of miscellaneous issues. There were 82 issues pointed out from the selected studies and were categorized into 19 challenges. Major described issues in reviewed articles were; use of vocabularies for assigning metadata; lack of awareness, developing/ maintenance of metadata, lack of expertise, diversity in metadata standards, metadata harvesting, interoperability, compatibility, typographical errors and data transferring. Hopefully, this study will guide to the LIS professionals regarding problems, issues and challenges concerning with the metadata implementation in libraries and they will be able to keep in mind these challenges if willing to implement metadata in their libraries.

The researchers categorized these challenges into two thematic groups for the representation of their structural corpus.

Table 2

Project based challenges are thematically grouped to synthesize

Project Based Challenges Extracted from Selected Studies	Challenges Grouped Thematically
Human and organizational issues; Metadata skillful staff; Local experts support for new project; Unawareness about the use and strength of metadata systems; Typographical error	Human and organizational challenges
Using un-standardized metadata; Compatibility of data; Lack of standards; Lack of standards Software tools; Reuse of metadata schemas; Inappropriate use of upper/ lower case letters; Type and Granularity of Resources; Assigning subject terms	Lack of Standardization
Interoperability and reuse of metadata schemas; Compatibility with present access and mechanism; Controlled vocabularies;	Interoperability and Compatibility
Technical issues include; Maintenance; Development and maintenance of metadata schemas; Metadata schema sharing	Technical Challenges

Project Based Challenges of Metadata

The table 2 represents the project based challenges which were prescribed in the selected studies and are grouped thematically. There were many challenges whose essence was same. So, all those issues which were related to human or organizational element are grouped and their synthesized group named as “Human and organizational challenges”. Then there were some issues which meaningfully described the specific sort of standards so these challenges are grouped as “Lack of standardization”. In some studies compatibility type issued were presented so these challenges are grouped into the

group “Interoperability and compatibility”. While there were some issues which reflected the technical challenges are thematically grouped as “Technical challenges”.

Table 3

General challenges are thematically grouped to synthesize

General Challenges Extracted from Selected Studies	**Challenges Grouped Thematically**
Using different metadata terms for data elements affect results and represent different functional uses; Specifying subject coverage; Lack of lexical evidence in text of metadata; Meaning of values contained in elements; Variability among entered values; Lack of guidelines in describing information resources; Controlled vocabularies; Variation in subject fields; Domain specific classification keywords; content standard and schemes; Unified index from different level of records; Use of different vocabularies for metadata; lack of native tags for contextual information;	Lack of Lexical Terminologies and Controlled Vocabularies for Metadata
Standards; Many metadata standards and schemas; Continual evolution of standards; Models and schemas of metadata have setback to be organized organically in a single space of linked data; Structured data for use in named entity recognition (NER) vary in each case	Use of Various Standards
Harvesting of metadata; Incomplete harvesting; Multiple record	Harvesting and

harvesting; Curation of large scale harvested metadata; Normalization in union repository; Legacy issue of harvesting metadata; Metadata harvesting; Harvesting of metadata from multiple sources cause metadata varying structure; Normalization of data;	Curation of Metadata
Metadata creation and management; Little amount of metadata; Economics of metadata creation; Variation in descriptive practices; Untitled metadata; Undocumented metadata aggregation; Automatic creation of metadata; Insufficient metadata; Incorrect metadata; Metadata creation of hybrid documents; Metadata management	Metadata Creation and Management
Non-OAI accessible metadata; Connection time out; OAI index error; Mapping of data	Accessibility and Discovery
Inconsistency; Creating harmonious *conceptual reference model for description of metadata; Incompatible data elements/ formats; Inconsistency of metadata among disparate sources; Interoperability; unusable and unreliable*	Incompatible, Inconsistency and Interoperability
Specialized knowledge requirement; Plagiarism; Inadequate and qualified metadata specialists; Inadequate IT infrastructure; Inadequate education and experienced digitization consultants; Typographical error; Less awareness about metadata preservation and technical aspect; Customize metadata schemas	Human and Organizational Factors *(Lack of trained, aware specialist personal and technical issues)*

RDF triple describing item-collection relationship; Promise of linked open data; Linking of metadata with full text; Difficulties in connecting some MARC fields to linked data; Metadata connections with vendors, publishes and aggregators	Relationship among Metadata and Resources *(Connection with External/ Other Resources)*
Crosswalking metadata from one scheme/ format to another; Data transmission error; Electronic metadata corruption during conversion to another scheme; Conversion of MARC metadata into linked data; Reuse of legacy metadata in other environment; OAI crosswalk	Crosswalking/ Conversion of Metadata/ Data
For image based data little amount of metadata is an issue for accessibility; Use of single MARC record to describe more than one format or version	Little Amount of Metadata
Junked Unicode characters; quality; Loss of metadata granularity and inability to recreate the original records; Hierarchical metadata schemas are not supported by DSpace	Miscellaneous Issues

General Challenges of Metadata

The table 3 represents general challenges which were extracted from selected studies and are grouped thematically. The major issues in various studies were related to assigning metadata to material due to certain reasons so all of such issues are synthesized into the thematic group “Lack of lexical terminologies and controlled vocabularies for metadata”. Some different standards affect metadata so these types of issues are placed in

the group “Use of various standards”. Many studies prescribed issues related to data cleaning and such issues are grouped into “Harvesting and curation of metadata”. There were some issues related to creation and management of metadata are synthesized as “Metadata creation and Management’. Searching and their results based issues are group as “Accessibility and discovery”. issues related to human and organizations are grouped as “Human and organizational factors *(Lack of trained, aware specialist personal and technical issues).* Connectivity related issues are placed into the group “Relationship among Metadata and Resources *(Connection with External/ Other Resources).* Data conversion related issues were combined “Crosswalking/ conversion of metadata/ data” group. Metadata accessibility issues are placed in the group of “Little amount of metadata’. Issues which were not fallen in a specific category were grouped in “Miscellaneous issues”.

Discussion, Conclusion and Recommendations

As the intent of this systematic literature review study was to extract issues related to metadata so that these challenges may be presented at one single paper by obtaining from previously published literature. In this study 82 issues were extracted from the thirteen selected studies which are presented in figure 2. Further, these issues are divided into two major categories (Project based and general challenges) on the basis of thematic essence.

There are four thematic categories developed of with the metadata challenges related project based studies while eleven thematic categories are composed of with the metadata issues from the general issue based studies. If we have a critical look over the both thematic categories, we can observe that some thematic categories (Human and

organizational challenges; Standardization and Interoperability &Compatibility) are overlapping which means these categories are most prominent challenges of consideration concerning with the metadata. Furthermore, keeping in view the challenges extracted through this study one can better plan out and go ahead in the field of implementing metadata related initiatives in their workplaces, institutions, organizations etc. This study will add in the existing studies in form of synthesized thematic major challenges which need to be consider before starting metadata projects

For results extraction there were 13 studies which were relevant to the research question of the study. Keeping in view the challenges presented in the selected studies it is suggested before implementing a system which uses metadata, relevant people and organization must be on board so that human based and organizational issues may be reduced. Data or information providers should use standardized metadata. Metadata can be created for whole site, single page and even for single file but metadata should be formed in accordance with the need. The sharing of metadata schema information for customization of existing schema and to build new schema is important. Specificity of subject coverage vary discipline to discipline so keep it in mind. Variability among entered values means the terms used for data description are also important with respect to reduce the ambiguity of language, subjectivity of person describing the content and nature of disciplines. Metadata experts should thoroughly read about the the terminologies and may compose their items' descriptions toavoid plagiarism. During harvesting use MARC Edit, because some unicode Latin words come as junked characters so proper planning is required to avoid such issue. Large amount of records in a software may interrupt the metadata harvesting. Connection timed out may be well

managed when data server is not active on internet. Cross walking metadata arise errors when data is converted from less rigid metadata scheme (like Dublin Core) to data values which are tightly controlled (like MARC).

References

Attig, J., Copeland, A., & Pelikan, M. (2004). Context and meaning: The challenges of metadata for a digital image library within the university. *College & Research Libraries*, *65*(3), 251-261.

Banerjee, K. (2000). Challenges of using metadata in a library setting: the Collection And Management of Electronic Links (CAMEL) project at Oregon State University. *Library Collections, Acquisitions, and Technical Services*, *24*(2), 217-227.

Beall, J. (2005). Metadata and data quality problems in the digital library. *Journal of Digital Information*, *6*(3).

Bowen, J. B. (2010, September). Moving library metadata toward linked data: Opportunities provided by the eXtensible catalog. In *International Conference on Dublin Core and Metadata Applications* (pp. 44-59).

Calarco, P., Conrad, L., Kessler, R., & Vandenburg, M. (2015). Metadata challenges in library discovery systems.

Dashrath, V.B. (2014). Role of metadata in digital resource management. International journal of digital library services, 4(3).

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.

Freire, N., Borbinha, J., & Calado, P. (2012, June). An analysis of the named entity recognition problem in digital library metadata. In *Proceedings of the 12th ACM/IEEE-CS joint conference on Digital Libraries* (pp. 171-174). ACM.

Gradmann, S. (2009). Interoperability Challenges in Digital Libraries. First DL. In *org Workshop on Digital Library Interoperability, Best Practices and Modelling Foundations (ECDL 2009). http://www. dlorg. eu/uploads/Workshop% 20Corfu/Interoperability% 20Challenges% 20in% 20Digital% 20Libraries_Gradmann. pdf*.

Kaddu, S., & Kigongo-Bukenya, I. M. N. (2013, October). Digitisation and metadata challenges: Experiences of the World Digital Library (Uganda). In *2013 Digital Heritage International Congress (DigitalHeritage)* (Vol. 2, pp. 473-478). IEEE.

Khoo, M., Tudhope, D., Binding, C., Jones, H., Orrego, I., & Ahn, J. W. (2013). MASHing metadata: Legacy issues in OAI harvesting from three digital libraries.

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.

Solodovnik, I. (2011). Metadata issues in Digital Libraries: key concepts and perspectives. *JLIS. it*, *2*(2), 4-27.

Sugimoto, S. (2005, August). Challenges in Digital Libraries-Key Issues Learned from Metadata-Centric Projects at Tsukuba. In *Proceedings of IADLC (電子図書館国際会議, 名古屋大学)*.

Metadata definition from:

https://duckduckgo.com/?q=site%3Aw3.org+metadata&search-submit=&ia=web