Metadata matters: evaluating the quality of Electronic Theses and Dissertations (ETDs) descriptions in Malaysian institutional repositories

R. Osman [1,2], A.M.K. Yanti Idaya[1]* and A. Abrizah[1]
[1]Department of Library and Information Science,
Faculty of Arts and Social Sciences, Universiti Malaya,
Kuala Lumpur, MALAYSIA
[2]Dar al-Hikmah Library, International Islamic Universiti Malaysia,
Kuala Lumpur, MALAYSIA
e-mail: raudhah@iium.edu.my;
*yanti@um.edu.my (corresponding author); abrizah@um.edu.my
ORCID ID: R.Osman: 0009-0002-6968-8977
A.M.K.Yanti Idaya: 0000-0002-7642-469X
A.Abrizah: 0000-0002-8224-5268

ABSTRACT

This study investigates the quality of metadata records in institutional repositories for Electronic Theses and Dissertations (ETDs) in three Malaysian research universities. The study employs metrics derived from Bruce and Hillmann (2004) to evaluate the completeness, accuracy, and consistency of 1138 metadata records from the three institutional repositories, and utilizes quantitative content analysis to identify variants of Dublin Core metadata elements used to describe ETDs. The study uncovers a range of issues with metadata completeness, accuracy, and consistency, emphasizing the need for standardized systems to ensure quality metadata. By highlighting the importance of quality metadata for ETD collections, the study provides valuable insights for academic libraries in designing and describing ETD collections. Overall, this study sheds light on the current metadata practices and elements used in ETD collections, and underscores the significance of effective metadata management for the dissemination and discoverability of scholarly works.

Keywords: Quality assessment; Metadata quality; Institutional repositories; Electronic Theses and Dissertations; Dublin Core metadata schema.

INTRODUCTION

The proliferation of digital repositories aimed at supporting teaching and research is rapidly growing, highlighting the need for comprehensive research on metadata quality assessment. The provision of high-quality metadata has become a crucial factor in ensuring that users can access and utilise accurate resources. Libraries, archives, and museums engage in metadata representation of information objects as a fundamental activity to enable users' access to their stored knowledge (Zavalin, Zavalina, and Miksa 2021). The importance of metadata quality has grown substantially in recent years as it allows for efficient structuring of resources to ensure their accessibility and discoverability by library users. Metadata plays a critical role in digital repositories as it ensures that resources are accurately and effectively presented, enabling users to gain the maximum benefit from the metadata provided. The provision of high-quality metadata for ETDs presents a significant

challenge (Yasser 2011) as these resources are a vital source of information, especially for researchers. ETDs are often considered a valuable and unique source of information, providing access to research findings that may not be published elsewhere (Ramdas Lihitkar and Lihitkar 2014).

Correctly organising and delivering data values in digital repositories is crucial for ensuring the quality of metadata. Therefore, metadata assessment and quantification of bibliographic records are necessary to achieve high-quality metadata (Reiche and Hofig 2013). Quality metadata is an essential component for users to easily identify and locate the resources they need (Currier et al. 2004). The importance of metadata quality is reflected in its role in the primary functions of a digital repository, and metadata quality metrics should be employed to assess it. High-quality metadata ensures that resources are fully described and that users receive accurate, relevant, and sufficient information, allowing them to easily access the required resources.

A number of previous studies have evaluated digital repositories on various metadata quality metrics, including completeness, consistency, and accuracy (Moen, Stewart, and McClure 1997; Park 2009; Weagley, Gelches, and Park 2010; Bellini and Nesi 2013; Steele and Sump-Crethar 2016; Aissi and Sboui 2017; Pelaez 2017; Romero-Pelaez, Segarra-Faggioni, and Alarcon 2018; Mayernik 2019). The study aims to fill the gap in research on metadata quality assessment of ETDs in Malaysian institutional repositories, which has been overlooked in previous studies. The study uses a quantitative content analysis approach to comprehensively evaluate metadata records in the repositories, focusing on completeness, accuracy, and consistency metrics to ensure universal access to high-quality knowledge. The data for the study was generated from web institutional repositories.

LITERATURE REVIEW

In today's digital age, metadata quality assessment is becoming increasingly important in ensuring the effectiveness of digital collections. The quality of metadata, which provides structured information about resources, is essential for facilitating the management, discovery, and retrieval of resources in digital repositories (Al-Khalifa and Davis 2006). In fact, metadata is a crucial feature of repositories that facilitates search and discovery tasks and ensures the reuse of educational resources (Romero-Pelaez, Segarra-Faggioni, and Alarcon 2018). Moreover, metadata enables users to identify and find the resources they need, making it an indispensable tool for efficient use of digital collections (Currier et al. 2004). As such, the need for comprehensive metadata quality assessment in digital repositories cannot be overstated.

Park (2009) reviewed metadata quality practices from a functional standpoint, focusing on metadata quality, measurement, and evaluation criteria in order to improve metadata quality in the current state of the art. His study found that correctness, completeness, consistency and accuracy are the most often utilised criteria for evaluating metadata quality. Metadata guidelines and (semi) automatic metadata generating tools appeared to be a quality assurance approaches, according to the study. In a 2010 online survey of cataloguing and metadata professionals in the United States, accuracy and consistency were identified as the most important quality metrics for digital repositories by 76.9 percent and 74.3 percent of respondents, respectively. Despite this, nearly two-thirds of the respondents (65%) still considered complete metadata description to be an important quality measurement metric (Park and Tosaka 2010).

Several studies have explored metadata quality in various contexts. For instance, one study aimed to address the issue of poor quality of metadata records describing educational content in Learning Object Repositories (LORs) by proposing a metadata quality assessment certification process for LORs (Palavitsinis, Manouselis, and Sanchez-Alonso 2014). The application of the certification process resulted in a remarkable improvement in the quality of 11,000 metadata records, with completeness for all metadata elements improving significantly from 30 percent to 85 percent. Another study evaluated completeness, accuracy, consistency, and the use of controlled vocabulary in 150 records of Dublin Core metadata elements from 6 collections (Weagley, Gelches, and Park 2010). The findings indicated that not all types of metadata were collected consistently and described accurately, particularly in collections that were a compilation of archives or repositories. An evaluation study on metadata quality in Open Course Ware (OCW) measured completeness, consistency, coherence, accuracy, provenance, and accessibility metrics. The study used metadata extracted from a data set stored in a semantic search engine's repository and found that all words from the metadata appeared in the title, indicating higher quality. Three out of nine OCW resources were also found to be of higher quality. None of the resources had a value of 0 for any of the metrics (Romero-Pelaez, Segarra-Faggioni, and Alarcon 2018).

Reiche and Höfig (2013) proposed an automatic assessment of metadata quality using established metrics from the field of metadata quality assessment. To demonstrate the application of these metrics, three public government data repositories (GovData.de, data.gov.uk, and publicdata.eu) were evaluated using five quality metrics: completeness, weighted completeness, accuracy, richness of information, and accessibility. The study emphasizes the importance of implementing robust metadata quality control measures to ensure reliable and accessible data in public repositories. In a related study, Bellini and Nesi (2013) evaluated metadata quality tools in Cultural Heritage Institutional Repositories and found that missing, outdated, and typo-ridden metadata fields pose challenges to searching, discovering, and retrieving materials. The study assessed metadata quality using high-level metrics such as completeness, accuracy, and consistency. Both studies demonstrate the need for metadata quality control measures to improve data accessibility and reliability in institutional repositories.

The assessment of metadata has not only been explored on institutional repositories, but also on other digital platform such as Volunteered Geographic Information (VGI), scientific research, weather and climate, and music. In the context of VGI, indicators have been proposed to evaluate the external quality of geospatial metadata with respect to the users' context, including completeness, trust, and freshness metrics. Previous research has shown that assessing VGI data for accuracy and credibility can be challenging when dealing with a large volume of events. Additionally, a study evaluated metadata records from 42 federal agencies using completeness and accuracy metrics for 67 metadata elements with both mandatory and optional GILS records structure. Completeness was measured in terms of description level, and accuracy was evaluated based on format, formatting errors, spelling, and typographical errors. Visible errors in each record were counted for the accuracy criteria, and the sample indicated that 10-30 percent of records contained such errors. The completeness criteria evaluated the completeness of sampled records in terms of included elements. Furthermore, quality metrics such as completeness, freshness, and trust were employed to assess the geospatial metadata quality of VGI data (Moen, Stewart, and McClure 1997; Aissi and Sboui 2017).

Theses and dissertations are a unique and valuable collection for universities, and making them visible and discoverable worldwide is essential. Steele and Sump-Crethar (2016) conducted a critical investigation to determine the current state of metadata quality control in institutional repositories, which hold these important documents. The study found a wide variety of metadata used to describe theses and dissertations, with frequent use of elements such as Abstract, Title, Author Subject, and Type (thesis, dissertation). The current study has a specific focus on evaluating the metadata quality of Electronic Theses and Dissertations (ETDs) in selected institutional repositories in Malaysia, utilizing metrics such as completeness, accuracy, and consistency to evaluate the quality of metadata associated with these important academic works.

OBJECTIVE AND METHODS

The objective of this study is to assess the quality of metadata specifically for Electronic Theses and Dissertations (ETDs) collections. Specifically, the study aims to analyze the metadata elements used to describe ETDs in institutional repositories in Malaysia and evaluate the quality of metadata records using completeness, accuracy, and consistency metrics. The assessment of metadata quality is based on the general characteristics of metadata quality and the definition of levels of quality for metadata as proposed by Bruce and Hillmann (2004).

The study employs a quantitative content analysis approach to assess the metadata records of the collection. Content analysis is a research method that provides a systematic and objective means of drawing valid conclusions from verbal, visual, or written data to describe and quantify specific phenomena. It goes beyond a simple counting game and focuses on meaning, intentions, consequences, and context (Downe-Wamboldt 1992). The study by Moen, Stewart, and McClure (1997) also used content analysis to assess the metadata records of 42 Federal agencies' implementation of the Government Information Locator Service (GILS).

To identify and evaluate the quality of metadata records in institutional repositories, the study carefully selected three Malaysian research university repositories based on multiple criteria. Firstly, the repositories contained a significant collection of ETDs, which were relevant to the study's objectives. Secondly, the repositories were registered with two of the most reputable repository directories (Registry of Open Access Repositories and Directory of Open Access Repositories), indicating their reliability and credibility. Thirdly, the repositories were affiliated with established Malaysian research universities, providing a strong foundation for the study. Finally, the researchers were able to access these repositories, allowing for a comprehensive analysis of their metadata records.

This study applied the following three stages in accessing metadata of the collection.
(a) The first stage was data collection which involved selecting theses and dissertations collection in three repositories, extracting metadata records of the collection in the repositories, and merging the metadata records into Microsoft Excel format.
(b) The second stage was preparation and evaluation of data, which included reviewing and identifying metadata elements, mapping each metadata record extracted with Dublin Core elements and in-house elements, checking the metadata elements used standard format practice, ensuring the description of metadata reflected the metadata element used, mapping the description of metadata element into completeness,

consistency, and accuracy metrics, and analyzing each metadata record based on the quality metrics.

(c) Finally, the third stage was result analysis, which involved interpreting results and categorizing frequencies based on the quality metrics used in the study. The content analysis steps are summarized in Figure 1.
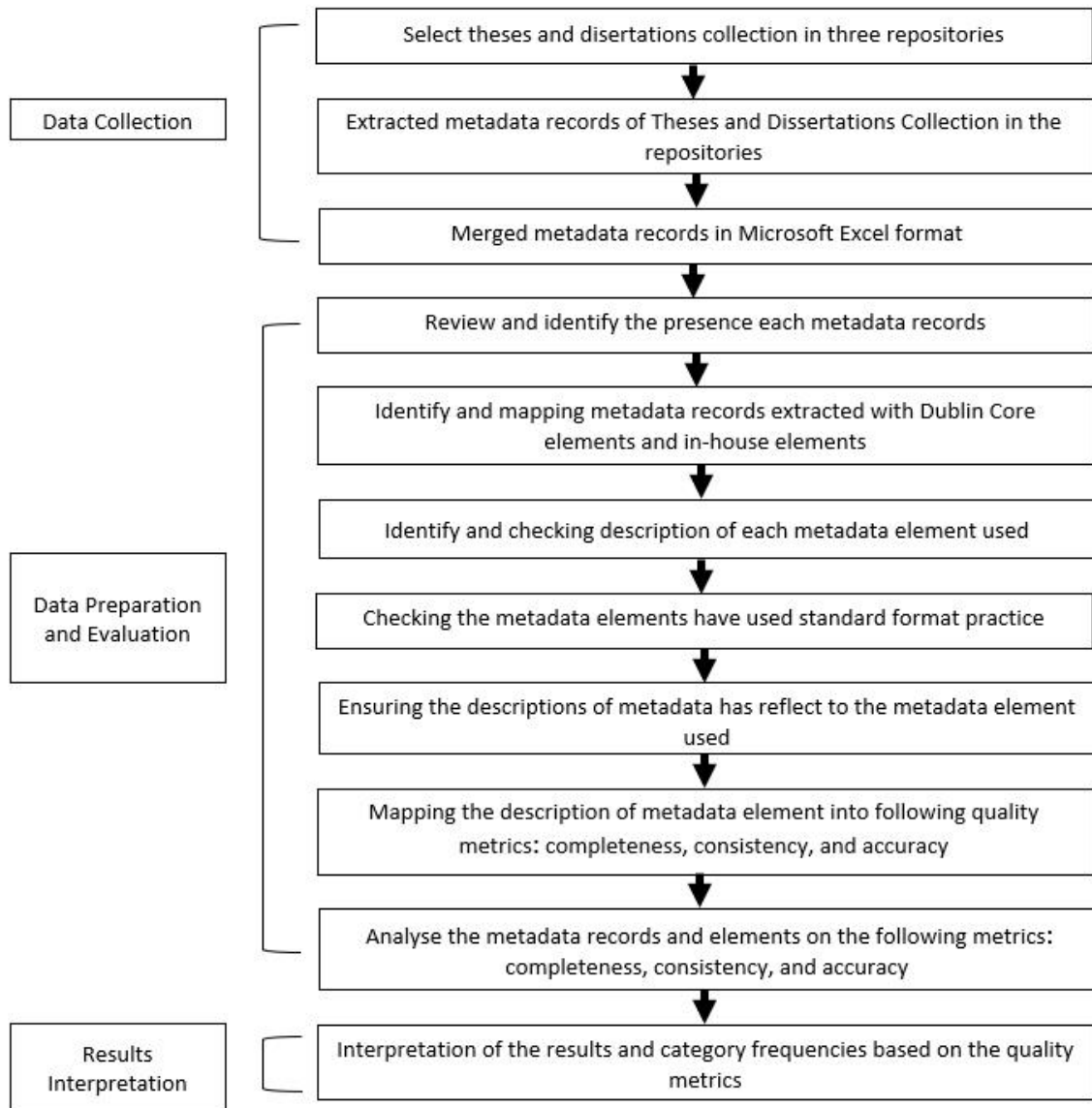

Figure 1: Content Analysis Steps for Evaluating Metadata Quality of Electronic Theses and Dissertations in Institutional Repositories

The study used metadata records from three Malaysian universities' institutional repositories (i.e. Eprints), labeled A, B, and C, to evaluate the quality of their ETDs collection. The metadata records were originally selected from the institutional repositories and primarily consisted of ETDs. The researcher accessed each institutional repository and selected the 'theses and dissertations' item type for metadata assessment. A compilation of all ETDs records and their associated metadata elements was generated from the web for further analysis.

In addition, the collection's metadata records were downloaded and retrieved in Dublin Core format from each institutional repository. To determine the sample size, Krejcie and Morgan's (1970) sampling size determination table was used. The Random Integer Generator tool from Random.org (2018) was then used to randomly select metadata records for the study. Additionally, each metadata record was manually checked for the inclusion of any in-house metadata used by each institution. The collected data was then compiled into an Excel format for metadata record evaluation, where each record was assessed for completeness, accuracy, and consistency. A summary of the extracted and sampled metadata records, totaling 1,138, is presented in Table 1.

Table 1: The Population and Sample Size for the Study

Repository	Size of the repository (N)	ETDs records sampled (n)
A	7,853	374
B	9.102	380
C	16,041	384

RESULTS

Presence of Metadata Elements

The ETDs collection in the repositories uses a Dublin Core metadata schema, with each institution using a slightly different set of metadata elements. Institution A uses nine metadata elements (Title, Creator, Subject, Date, Description, Type, Format, Relation, and Identifier), Institution B uses ten metadata elements (Title, Creator, Subject, Date, Description, Type, Format, Identifier, Language, and Relation), and Institution C also uses ten metadata elements (Title, Creator, Date, Subject, Description, Type, Format, Identifier, Language, and Relation). Table 2 provides a summary of the Dublin Core metadata elements used in each of the three repositories for their ETDs collection.

Table 2: Metadata Elements Used for ETDs Collection in the Three Repositories, with the Percentage of Occurrence

DC metadata elements	A (n=374)	% of DC elements used (n/6746)	B (n=380)	% of DC elements used (n/4197)	C (n=384)	% of DC elements used (n/4086)	Total (n/1138)	% of total usage of DC elements
Title	374	11.69	380	10.00	384	11.59	1138	11.03
Creator	374	11.69	380	10.00	384	11.59	1138	11.03
Subject	374	11.69	377	9.92	380	11.46	1131	10.97
Description	206	6.4	380	10.00	216	6.51	802	7.78
Date	374	11.69	380	10.00	384	11.59	1138	11.03
Type	374	11.69	380	10.00	384	11.59	1138	11.03
Format	374	11.69	380	10.00	219	6.61	973	9.43
Identifier	374	11.69	380	10.00	384	11.59	1138	11.03
Language	-	-	380	10.00	194	5,85	574	5.56
Relation	374	11.69	380	10.00	384	11.59	1138	11.03
Total	3198	100.00	3797	100.00	3313	100.00	10308	100.00

Table 3 provides a summary of the metadata elements used to describe the ETDs collection in the three institutional repositories. A total of ten metadata elements were identified,

including Title, Creator, Subject, Description, Date, Type, Format, Identifier, Language, and Relation. However, in comparison to the 15 Dublin Core elements, four elements were not utilized to describe the ETDs collection: contributor, source, coverage, and rights.

Table 3: Metadata Elements Used to Describe the ETDs Collection in the Institutional Repositories

Institutional repository		A	B	C
Number of metadata records sampled		374	380	384
Number of Dublin Core metadata elements used		9	10	11
Number of in-house metadata elements used		3	3	2
Dublin Core Metadata present	Metadata Element Definition	A	B	C
Title	A name given to the resource.	√	√	√
Creator	An entity primarily responsible for making the resource.	√	√	√
Date	A point or period of time associated with an event in the lifecycle of the resource.	√	√	√
Subject	The topic of the resource.	√	√	√
Description	An account of the resource.	√	√	√
Publisher	An entity responsible for making the resource available.	-	-	√
Contributor	An entity responsible for making contributions to the resource	-	-	-
Type	The nature or genre of the resource.	√	√	√
Format	The file format, physical medium, or dimensions of the resource.	√	√	√
Identifier	An unambiguous reference to the resource within a given context.	√	√	√
Language	A language of the resource.	-	√	√
Relation	A related resource.	√	√	√
Source	A related resource from which the described resource is derived.	-	-	-
Coverage	The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant.	-	-	-
Rights	Information about rights held in and over the resource.	-	-	-
In-house metadata element present				
Call number	Call number of the resource.	-		-
Chairman supervisor	Supervisor name.	-		-
Division	Faculty.	√		√
Additional Information	Additional Information.	√		√
Uncontrolled keyword	Keyword of the resource.	√		-

It was found that the Dublin Core Scheme allows for adaptability with each repeating element, but despite this flexibility, only a few metadata elements were used to characterize the ETDs collection in the investigated repositories. The following metadata components were identified:
   a. The Creator metadata element in institutional repository A is repeatable.
   b. The Language metadata element in institutional repository B is repeatable.
   c. The Type metadata element in institutional repositories A, B and C is repeatable.
   d. The Format metadata element institutional repository A is repeatable.

e. The Identifier metadata element in institutional repositories A, B and C is repeatable.
f. The Relation metadata element in institutional repositories A and C is repeatable.
g. The Subject metadata element in institutional repositories C and B is repeatable.

Metadata Quality

The metadata assessment of records in three Malaysian public university repositories involved evaluating completeness, accuracy, and consistency metrics. A total of 1,138 metadata records were extracted from the three repositories, comprising 374 metadata records from repository A, 380 metadata records from repository B, and 384 metadata records from repository C.

(a) Completeness

The completeness metric revealed that the metadata records assessed contained incompleteness errors. In institution A, several metadata records had incompleteness errors, with the Description element having the highest incompleteness error of n=166 (84.69%), followed by the Additional Information element with n=17 (8.67%). The Uncontrolled Keyword element had the lowest incompleteness error of n=13 (6.63%). The deposition guideline in institution A requires three metadata elements to be filled, namely Description, Additional Information, and Uncontrolled Keyword. Table 4 illustrates the incompletely filled Description element (in italic), among other sample metadata completeness errors presented in Tables 5 and 6.

Table 4: Missing Information for "Description" Metadata Element

Title	Strategi Pembelajaran Kosa Kata Bahasa Arab Pelajar Sekolah Menengah Agama di Malaysia / Harun Baharudin
Creator	Baharudin, Harun
Subject	L Education (General)
	LB Theory and practice of education
Description	***Description not available***
Date	2014
Type	Thesis
	NonPeerReviewed
Format	application/pdf
Item type	Thesis (PhD)
Uncontrolled keyword	Arabic language -- Vocabulary -- Study and teaching (Secondary) -- Malaysia; Vocabulary -- Study and teaching (Secondary)
Subject	L Education > L Education (General)
	L Education > LB Theory and practice of education
Divisions	Faculty of Education

Institution B had the highest number of missing information errors, with the Notes element having n=366 (91.08%) missing information, followed by the Subject element with n=23 (76.66%) and the Chairman Supervisor element with n=7 (23.33%). In institution C, incompleteness errors were identified in various metadata elements, namely Format, Language, Description, Division, and Subject. The Format and Language elements had the highest completeness error, with n=190 (20.34%), followed by the Description element with n=168 (17.98%). The Division and Subject elements had the lowest incompleteness errors, with n=7 (0.74%) and n=4 (0.42%), respectively.

Referring to Table 5, it can be seen that a completeness error occurred in the "Chairman Supervisor" element of the metadata record (in italic), which is one of the required fields

for describing the collection. The deposition guideline in institution B requires this element to be used to describe the name of the thesis Chairman or Supervisor with their academic title and qualifications as available in the thesis. The assessment of metadata records indicated that the Chairman Supervisor element in Table 5 was incompletely filled in the repository.

Table 5: Missing Information for "Chairman supervisor" Metadata Element

Title	Perceived Barriers to Trade Unionism in Malaysia
Creator	Ramasamy, Nagiah
Description	Although trade unions in Malaysia generally play an important role in protecting the rights of workers, the movement is facing a declining trend. This study…
Date	2010-01
Type	Theses
	NonPeerReviewed
Format	application/pdf
Language	en
Item Type	Thesis (PhD)
Subject	Labor unions - Malaysia
	Employee rights - Malaysia.
Call No.	GSM 2010 7
Chairman Supervisor	***Chairman supervisor not available***
Division	Graduate School of Management

In addition, the assessment of metadata records in institution C revealed an incompleteness error in the Format element (in italic), which is used to describe the file format of the resource. A sample of missing information in the Format field is presented in Table 6 to illustrate this incompleteness error.

Table 6: Missing Information for "Format" Metadata Element

Title	Shear behaviour of reinforced concrete beams strengthened with carbon fibre reniforced polymer fabrics
Creator	Ishak, Mohd. Yunus
Description	Goodness of fit (GOF) test is a statistical technique in selection of an appropriate probability distribution…
Subject	T Technology (General)
Date	2012
Type	Thesis
	NonPeerReviewed
Format	***Format not available***
Language	Language not available
	-
Item type	Thesis (PhD)
Subject	T Technology > T Technology (General)

(b) Accuracy

The repository at institution A showed that the title element had the highest accuracy error rate of 31.25% (n=5), followed by the description and division elements at 25% (n=4). The creator/author element had an accuracy error rate of 12.5% (n=2), while the additional information element had the lowest accuracy error rate of 6.25% (n=1).

In institution B, the highest inaccurate errors identified is Identifier element with the total frequency n=5 (50%), followed by Title element with the total frequency n=4 (40%) and Creator element with the total frequency n=1 (10%). Meanwhile in institution C, incorrect value was identified as the highest error occurred in Subject element with the total frequency n=55 (63.21%) and followed by Relation element with total frequency n=31 (35.35%). Type metadata element was the lowest error occurred with total frequency n=1 (1.14%).

The study identified accuracy errors in selected metadata records from the repositories, specifically in the Description, Subject, and Title elements. These errors are illustrated in Tables 7, 8, and 9 (in italic). Table 7 displays incorrect or inaccurate value in the Description element, which is a required element for describing the collection's abstracts. According to the deposition guideline in Institution A, the Description element should contain information about the abstract available in the theses, dissertations, or academic exercises. However, the study found that words were repeated in the Description element, indicating an accuracy error.

Table 7: Incorrect Value for Description Metadata Element

Title	Institutionalization of state policy: The role of state-owned enterprises, provinces and municipalities in China's Urban housing market / Zhang Miao
Creator	Zhang, Miao
Subject	H Social Sciences > HB Economic Theory
Description	***The allocation of urban housing remains a serious challenge in China, which has The allocation of urban housing remains a serious challenge in China, which has The allocation of urban housing remains a serious challenge in China…***
Date	2014
Type	Thesis NonPeerReviewed
Format	application/pdf
Item type	Thesis (PhD)
Uncontrolled keyword	The role of state-owned enterprises
Divisions	Faculty of Economics & Administration

Table 8 illustrates inaccurate information found in the Creator metadata element, which is a mandatory element in repositories used to describe the author's name. The repository's deposition guideline specifies that the creator's name should be entered as the family name, followed by the author's first name, and any titles such as *bin, binti, s/o, a/k* or honorifics such as *Haji, Dato', Tan Sri,* etc. should be omitted. However, this study identified inaccuracies in the Creator metadata element, specifically in the spelling of the surname as "Lalbakshs" instead of "Lalbakhsh" (as found in the full-text). This inaccuracy may lead to difficulties in searching and retrieving information related to the author.

The subject element is a required field in describing a thesis collection in an institutional repository, according to deposition guideline. However, the record in Table 9 shows that the subject element was incorrectly entered in the repository, with an unspecified value. This is not in accordance with what the guideline indicates and contradicts other records.

Table 8: Incorrect Value for Creator Metadata Element

Title	Engendering a new feminine identity and reconstructing consciousness-raising in Doris Lessing's selected novels
Creator	***Lalbakshs, Pedram***
Description	Looking from a socialist feminist viewpoint, the present study discusses selected Doris Lessing?s novels to identify and analyze her engendered New Feminine Identity…
Date	2012-07
Type	Thesis NonPeerReviewed
Format	application/pdf
Language	en
Item Type	Thesis (PhD)
Subject	Feminist criticism Lessing, Doris May, 1919 - Criticism and interpretation Feminism
Call No.	FBMK 2012 38
Chairman Supervisor	Associate Professor Wan Roselezam Wan Yahya, PhD
Division	Faculty of Modern Language and Communication

Table 9: Incorrect Value for Subject Metadata Element

Title	Cadangan pengambilalihan pengurusan penyelenggaraan harta bersama Rumah Pangsa Kos Rendah oleh pihak berkuasa tempatan : kajian kes di kawasan pihak berkuasa tempatan Majlis Perbandaran Johor Bahru Tengah
Creator	Yusuf, Shubanah
Description	-
Subject	***Unspecified***
Publisher	-
Date:	2006
Type:	Thesis NonPeerReviewed
Format	-
Language	En
Item type:	Thesis (Masters)
Subject:	Unspecified
Divisions:	Geoinformation Science And Engineering (Formerly known)

(c) Consistency

The study identified inconsistencies in various metadata elements across the institutional repositories. Institution A had the highest number of missing information errors, with the Title element having the highest frequency of errors (n=275, 34.85%), followed by the Additional information element (n=271, 34.34%), Creator/author element (n=177, 22.43%), Date element (n=62, 7.85%), Description element (n=2, 0.25%), and Identifier element (n=1, 0.12%).

In contrast, the study revealed that Institution B had a different pattern of inconsistency errors. Specifically, the highest frequency of errors was found in the Title element (n=260, 65.65%). This was followed by the Date element (n=102, 25.75%), and the Format element, (n=24 6.06%). In comparison, the Description element had a total frequency errors of n=5

(1.26%), the Subject element had a total frequency errors of n=4 (1.01%), and the Creator element had a total frequency of errors n=1 (0.25%).

In addition, the study revealed that institution C had the highest occurrence of consistency errors among the three institutions. Specifically, the Date element had the highest frequency of inconsistency errors with a total of n=56 (21.37%), followed by Additional information metadata element with a total frequency of n=23 (8.78%), Identifier element with a total frequency of n=12 (4.58%), Title element with a total frequency of n=11 (4.2%), and Description element with a total frequency of n=5 (1.91%). The study further highlights consistency errors in selected metadata elements, including Date, Description, and Title, which are presented in Tables 10, 11, and 12, respectively.

Table 10 illustrates an example of a date formatting error (in italic) that was identified in institution A, where dates were coded based on month and year. However, according to institutional practice, the date formatting for the thesis collection in their repository must be in the YYYY (year) format.

Table 10: Structural Consistency Error of Date Metadata Element

Title	Experimental investigation and constitutive modeling of randomly oriented electrospun nanofibrous membranes / Wong Dannee
Creator	Wong , Dannee
Description	The recent advancement of nanotechnology has enabled the fabrication of nanofibers through a number of processing techniques. Among these, electrospinning offers a unique ability to produce nanofibrous membranes for different materials and of different assemblies or textures that make them suitable…
Date	*2018-08*
Type	Thesis NonPeerReviewed
Format	application/pdf
Item type	Thesis (PhD)
Uncontrolled keyword	Constitutive modeling; Nanofiber; Membrane; Mechanical response; Electrospinning
Subject	T Technology > TJ Mechanical engineering and machinery
Divisions	Faculty of Engineering

A technical consistency error was identified in the capitalization of the "Title" metadata element (in italic), which is used to describe the title of the thesis. According to best practices, capital letters should be used for the first word of the title and proper nouns, such as the names of racial or ethnic groups. However, errors in capitalization were identified and are shown in Table 11.

A structural consistency error was discovered in the Description element which is used to provide the abstract of the thesis. The study revealed that 6 records (1.90%) in the repository had an unfilled Description element. This omission led to challenges in searching and retrieving the required resources by users. The abstract metadata element is expected to be in English for a resource that is in same the language. However, during the assessment, it was observed that for some resources written in English, the Description element was written in the Malay language. Even though the actual PDF version of the thesis contained an English abstract, errors still occurred in the data. Table 12 illustrates an example of this inconsistency.

Table 11: Structural Consistency Error of Title Metadata Element

Title	***Typology of traditional courtyard houses as muslim functional spaces in low rise residential units, Iraq***
Creator	Haraty, Hayder Jawad Shakir
Description	A courtyard is an important transitional functional space with unique aesthetic design element used for Islamic social and cultural activities in traditional Iraq's…
Date	2015-12
Type	Thesis
	NonPeerReviewed
Format	Text
Language	En
Item Type	Thesis (Masters)
Subject	Courtyard houses – Iraq
	Architecture - Islamic influences - Iraq
	-
Call No.	FRSB 2015 7
Chairman Supervisor	Mohammad Yazah Bin Mat Raschid, PhD
Division	Faculty of Design and Architecture

Table 12: Structural Consistency Error of Description Metadata Element

Title:	Development of general purpose graphics processing unit with shading capability
Creator:	H'ng, Gaik Hi
Subject:	T Technology (General)
Description	***Pembangunan GPU dalam FPGA telah banyak diusahakan sejak kebelakangan ini. GPGPU merupakan…***
Date:	2013-01
Publisher:	-
Type:	Thesis
	NonPeerReviewed
Format:	application/pdf
Language:	En
Item type:	Thesis (Masters)
Subject:	T Technology > T Technology (General)
Divisions:	Electrical Engineering

DISCUSSION

The study conducted an assessment of 1,138 metadata records across three Malaysian institutional repositories, which revealed the use of nine Dublin Core metadata elements to describe ETDs collections. These elements include Title, Creator, Date, Subject, Description, Type, Format, Identifier, and Relation. This finding is consistent with a previous study by Steele and Sump-Crethar (2016). On the other hand, Park and Richard (2011) found that Title, Abstract, Type, Author, and Subject are commonly used to describe theses and dissertations collections. Additionally, Zavalin, Zavalina, and Miksa (2021) noted that Author, Title, and Subject are frequently included in resource descriptions. The study also found support for the use of ten descriptive elements, including Title, Author, Advisor, ORCHID identifier, Date, Keywords, Disciplines, Citation, Comment, and Abstract, in the institutional repository of the University of Nebraska-Lincoln (UNL) (Mering 2019). It is recommended to adopt a standardized set of metadata elements to ensure high-quality

metadata for ETDs collections. It is important to note that theses and dissertations collections are unique to each institution.

The study identified that the Format, Language, and Description metadata elements had the highest incompleteness errors, indicating missing information. In contrast, the Chairman Supervisor, Division, and Subject elements had the lowest incidence of missing information. It is noteworthy that the Description element is a critical element used to record the abstracts of the theses and dissertations, and it is highly recommended to be used for all resources. The study's findings support previous studies by Bui and Park (2013), Yasser (2012), and Sharifah Nur Amirah (2017), which also found the highest incompleteness errors in the Format, Description, and Keyword metadata elements.

Another important finding highlighted in this study concerns accuracy quality metrics. The study found that the metadata elements Subject and Relation had the highest accuracy errors, with incorrect values being the main issue. Specifically, the Subject element contained misinformation and "Unspecified" data values. This is consistent with previous research on mismatches of metadata values in metadata elements (McClelland et al. 2002). In institutions A and C, the Subject selected should be based on the subject of the resources supplied in the system, while in institution B, it should be as it appears in the Library of Congress Subject Headings (LCSH) or Medical Subject Headings (MeSH). Inaccurate data values in the Author and Subject elements can cause potential search interference and seriously impede metadata quality in institutional repositories (Wilson 2007). This finding supports previous studies that have also reported inaccurate data values in the Subject, Author, and Language elements, which can hamper metadata quality in institutional repositories and the possibility of search results (Sharifah Nur Amirah 2017).

The last finding of this study relates to the consistency quality metric and identified inconsistencies in the Title, Creator, and Additional Information metadata elements. The Title element should be entered with the first letter of each word capitalized in certain situations, while the Creator element should contain information about the author of the resource. Additionally, in Institution A, the Additional Information element should include information supplied by the system. This finding is consistent with previous research by Sharifah Nur Amirah (2017), which found missing data values in the Creator, Title, and Date elements. Date elements were also found to have discrepancies between the three institutions, with Institution C reporting the highest number of inconsistency errors. These inconsistencies can be attributed to structural inconsistencies and variations in date encoding. It is important for the date to be entered correctly and consistently in accordance with metadata creation standards described in guidelines or manuals (Moen, Stewart, and McClure 1998; Bellini and Nesi 2013; Chuttur 2014).

Overall, the study emphasizes the importance of metadata quality for institutional repositories, especially for ETDs collections, and highlights the critical role of clear and simple metadata guidelines in improving the quality of metadata records. The study identified several quality metrics, including completeness, accuracy, consistency, and timeliness, which are essential for assessing metadata quality. This study supports a previous one that found over half of the Association of Research Libraries (ARL) provide metadata documentation, but only 57.14 percent of them integrate it with other non-metadata documentation, and 17.14 percent of institutions do not provide metadata guidance (Kenfield 2019). The study's findings also reveal the need for establishing a common data model that is interoperable across institutional repositories, particularly for ETDs collections, to improve the quality of metadata records and search results.

Ultimately, by ensuring high-quality metadata, institutions can enhance the discoverability and accessibility of their digital resources and provide better support for research and scholarship.

CONCLUSIONS

This study offers insights into the level of metadata quality of ETDs in Malaysian institutional repositories, focusing on three core metrics: completeness, accuracy, and consistency. The study's assessment of metadata quality contributes valuable findings to the literature by highlighting the errors that occurred in various metadata elements, specifically in the ETDs collection. The study's human evaluation approach, which involved statistically analyzing a significant sample of metadata records, enabled a thorough analysis of the metadata quality of the collection.

However, it is important to acknowledge that the study's conclusions are based on a limited sample size and collection type. Further research is necessary to confirm the study's findings, including replicating the study with additional digital repositories in different types of institutions and collections. Moreover, conducting comparative studies with other countries would provide valuable insights into the metadata quality assessment of Malaysian higher learning institutions. Future studies may also consider assessing the metadata quality of additional resources housed in institutional repositories using different quality metrics, which would allow for a more comprehensive evaluation of resource metadata quality.

Overall, this study's findings provide practical suggestions for creating a metadata framework for institutional repositories in Malaysia and potentially in other nations, highlighting the importance of metadata completeness, accuracy, and consistency in facilitating resource discovery and use.

ACKNOWLEDGMENT

AUTHOR DECLARATION

The authors have no conflict of interest to declare.

REFERENCES

Aissi, S. and Sboui, T. 2017. Towards evaluating geospatial metadata quality in the context of VGI. *Procedia Computer Science*, Vol. 109:686–91. Available at: https://doi.org/10.1016/j.procs.2017.05.377.

Al-Khalifa, H.S. and Davis, H.C. 2006. The evolution of metadata from standards to semantics in e-learning applications. *Proceedings of the Seventeenth Conference on Hypertext and Hypermedia, Odense, Denmark,* 69-72. New York, NY: Association for Computing Machinery.

Bellini, E. and Nesi, P. 2013. Metadata quality assessment tool for open access cultural heritage institutional repositories. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 7990 LNCS: 90–103. Available at: https://doi.org/10.1007/978-3-642-40050-6_9.

Bruce, T. R., and Hillmann, D. I. 2004. Metadata in practice. In *The Continuum of Metadata Quality: Defining, Expressing, Exploiting*, 238–236. Chicago: American Library Association.

Bui, Y. and Park, J.R. 2013. An assessment of metadata quality: a case study of the National Science Digital Library metadata repository. *Proceedings of Information Science Revisited Approaches to Innovation / Actes Du congrès Annuel De l'ACSI.*, 1–13. Available at: http://idea.library.drexel.edu/handle/1860/1600.

Chuttur, M.Y. 2014. Investigating the effect of definitions and best practice guidelines on errors in Dublin Core metadata records. *Journal of Information Science,* Vol. 40, no. 1: 28–37. Available at: https://doi.org/10.1177/0165551513507405.

Currier, S., Barton, J., O'Beirne, R. and Ryan, B. 2004. Quality assurance for digital learning object repositories: issues for the metadata creation process. *ALT-J Reesearch in Learning Technology,* Vol. 12, no. 1: 5–20. Available at: https://doi.org/10.1080/0968776042000211494.

Downe-Wamboldt, B. 1992. Content analysis: method, applications, and issues. *Health Care for Women International,* Vol. 13, no. 3: 313–21.

Kenfield, A.S. 2019. Metadata documentation practices at ARL institutional repositories. *Portal: Libraries and the Academy,* Vol. 19, no. 4: 667–99. Available at: https://doi.org/10.1353/pla.2019.0041.

Krejcie, R. V. and Morgan, D.W. 1970. Determining sample size for research activities. *Educational and Psychological Measurement,* Vol. 30: 607–10. Available at: https://doi.org/10.1177/001316447003000308.

Mayernik, M.S. 2019. Metadata accounts: achieving data and evidence in scientific research. *Social Studies of Science,* Vol. 49, no. 5: 732–57. Available at: https://doi.org/10.1177/0306312719863494.

McClelland, M., McArthur, D., Giersch, S. and Geisler, G. 2002. Challenges for service providers when importing metadata in digital libraries. *D-Lib Magazine,* Vol. 8, no. 4. Available at: https://doi.org/10.1045/april2002-mcclelland.

Mering, M. 2019. Transforming the quality of metadata in institutional repositories. *Serials Librarian,* Vol. 76, no. 1–4: 79–82. Available at: https://doi.org/10.1080/0361526X.2019.1540270.

Moen, W.E., Stewart, E.L. and McClure, C.R. 1998. Assessing metadata quality: findings and methodological considerations from an evaluation of the US Government Information Locator Service (GILS). *Proceedings IEEE International Forum on Research and Technology Advances in Digital Libraries -ADL'98-*, 246–55. Available at: https://doi.org/10.1109/ADL.1998.670425.

Moen, W.E., Stewart, E.L. and 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. *IEEE Computer Society Metadata Conference*, 1--14. New York, NY: IEEE.

Palavitsinis, N., Manouselis, N. and Sanchez-Alonso, S. 2014. Metadata quality in learning object repositories: a case study. *Electronic Library,* Vol. 32, no. 1: 62–82. Available at: https://doi.org/10.1108/el-12-2011-0175.

Park, E.G. and Richard, M. 2011. Metadata assessment in E-Theses and Dissertations of Canadian institutional repositories. *Electronic Library,* Vol. 29, no. 3: 394–407. Available at: https://doi.org/10.1108/02640471111141124.

Park, J.R. 2009. Metadata quality in digital repositories: a survey of the current state of the

art. *Cataloging & Classification Quarterly,* Vol. 47, no. 3–4: 213–28. Available at: https://doi.org/10.1080/01639370902737240.

Park, J.R. and Tosaka, Y. 2010. Metadata quality control in digital repositories and collections: criteria, semantics, and mechanisms. *Cataloging and Classification Quarterly,* Vol. 48, no. 8: 696–715. Available at: https://doi.org/10.1080/01639374.2010.508711.

Pelaez, R. 2017. Metadata quality assessment metrics into OCW repositories." In *ICETC 2017: Proceedings of the 2017 9th International Conference on Education Technology and Computers*, 253–57. Available at: https://doi.org/https://doi.org/ 10.1145/3175536.3175579.

Ramdas Lihitkar, S. and Lihitkar, R.S. 2014. Electronic Theses and Dissertations (ETDs) in India: a comparative study. *Library Hi Tech News,* Vol. 31, no.2: 9–14. Available at: https://doi.org/10.1108/LHTN-10-2013-0061.

Random.Org. 2018. Random Integer Generator. Available at: https://www.random.org/integers/.

Reiche, K.J. and Höfig, E. 2013. Implementation of metadata quality metrics and application on public government data. *IEEE 37th Annual Computer Software and Applications Conference Workshops*, Japan, pp. 236-241. Available at: doi: 10.1109/COMPSACW.2013.32.

Romero-Pelaez, A., Segarra-Faggioni, V. and Alarcon, P.P. 2018. Exploring the provenance and accuracy as metadata quality metrics in assessment resources of OCW repositories. *ACM International Conference Proceeding Series*, 292–96. Available at: https://doi.org/10.1145/3290511.3290540.

Sharifah Nur Amirah, S.A. 2017. Conceptions and practices of metadata quality for FRBR user tasks resources discovery in institutional repositories. [Unpublished Doctoral Thesis, International Islamic University Malaysia]. IIUM Repository (IRep). Available at: http://irep.iium.edu.my/74362/.

Steele, T. and Sump-Crethar, N. 2016. Metadata for Electronic Theses and Dissertations: a survey of institutional repositories. *Journal of Library Metadata,* Vol. 16, no. 1: 53–68. Available at: https://doi.org/10.1080/19386389.2016.1161462.

Weagley, J., Gelches, E. and Park, J.R. 2010. Interoperability and metadata quality in digital video repositories: a study of Dublin Core. *Journal of Library Metadata,* Vol. 10, no. 1: 37–57. Available at: https://doi.org/10.1080/19386380903546984.

Wilson, A. J. 2007. Toward releasing the metadata bottleneck: a baseline evaluation of contributor-supplied metadata. *Library Resources & Technical Services,* Vol. 51, no. 1: 16–28. Available at: https://journals.ala.org/index.php/lrts/article/view/5384/6604.

Yasser, C.M. 2011. An analysis of problems in metadata records. *Journal of Library Metadata,* Vol. 11, no. 2: 51–62. Available at: https://doi.org/10.1080/19386389.2011.570654.

Yasser, C.M. 2012. An experimental study of metadata training effectiveness on errors in metadata records. *Journal of Library Metadata,* Vol. 12, no. 4: 372–95. Available at: https://doi.org/10.1080/19386389.2012.735573.

Zavalin, V., Zavalina, O.L. and Miksa, S.D. 2021. Exploration of subject representation and support of linked data in recently created library metadata: examination of most widely held WorldCat bibliographic records. *Library Resources & Technical Services,* Vol. 65, no. 4: 154–167. Available at: https://doi.org/10.5860/lrts.65n4.154.