Metadata Quality Benchmarks of ETDs in International Institutional Repositories: An Automated Appraisal

Aditi Roy[1] and Saptarshi Ghosh[2]

[1]Research Scholar, Department of Library & Information Science, University of North Bengal, Raja Rammohunpur, West Bengal, India

[2]Professor, Department of Library & Information Science, University of North Bengal, Raja Rammohunpur, West Bengal, India

Abstract

Good metadata quality makes a record more discoverable, facilitating search and retrieval. In this study, three methods are used to determine the quality of metadata of Electronic Theses and Dissertations; these are - Marc Report Analysis of Metadata, Metadata Quality parameters suggested by data.europa.eu, and lastly, a java based pre-compiled program by Peter Király has been used in this study. This study provides a brief comparative account of Electronic Theses and Dissertation Metadata structure of Institutional Repositories and Libraries. The comparative analysis of each repository shows that the total number of the record count is much higher in the case of the libraries as it was downloaded using z39.50/ SRU client. In contrast, in the case of repositories, only four sets of data are harvested using OAI-PMH. The field 040$e=rda is absent throughout the records of the institutional repositories. However, the field is present in a few records of the library, i.e., the University of Colorado shows 930 occurrences out of 1000 records. The process of metadata quality analysis involves a combination of automated tools and human expertise, ensuring a comprehensive evaluation of metadata attributes and relationships.

Keywords: Institutional Repositories, Libraries, Marc Report, MarcEdit, Metadata Quality, OAI-PMH, z39.50

1. Introduction

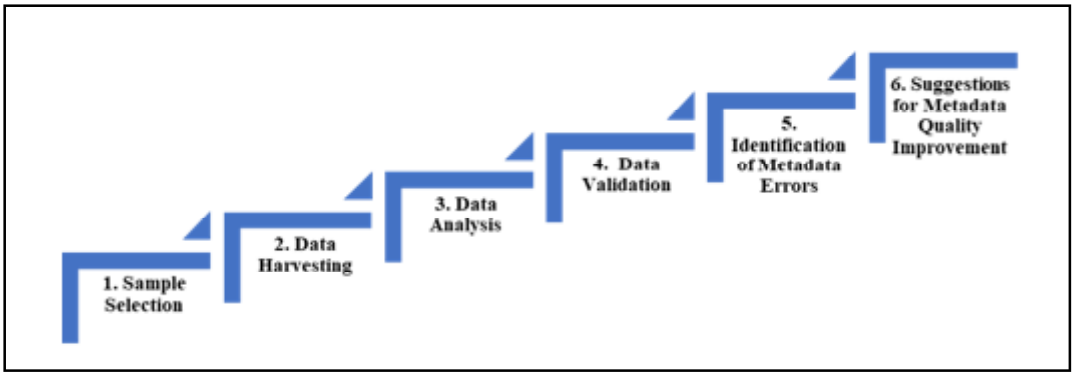
Metadata is defined as 'data about data'. The quality of a repository is often evaluated based on its metadata quality, as metadata facilitates the search and retrieval efficacy of its resources (Ramero-Palaez et al., 2018). The relevance of metadata quality has lately emerged as a significant critical problem in the literature on digital libraries, even though research and best practices manuals have been published on various metadata-related topics. An adequate study on ensuring the metadata quality in digital bibliographic data is required, highlighting the necessity for a comprehensive analysis of metadata quality issues from various perspectives and approaches.

Corresponding Author: Aditi Roy and Dr. Saptarshi Ghosh

Guy, Powell, and Day (2004) described metadata quality in terms of "functional requirements" or "fitness for purpose". Therefore, the evaluation criteria directly related to the functional metadata quality perspective. According to Park's study, **consistency, correctness, and completeness** are the most critical factors in determining metadata quality (Park, 2009). When metadata is comprehensive, each item is documented using all the metadata components necessary for full access to that object's content in digital repositories. Another study was carried out in 2022 that investigated how metadata quality could be improved throughout the different phases of metadata in a big data environment (Elouataoui et al., 2022). A framework named MetaEnhance is proposed that is used to improve the quality of scholarly metadata by auto detecting the errors, correction and canonicalisation (Choudhury et al.,2023). "Metadata Quality Assessment tool developed by the consortium of data.europa.eu to study the quality of metadata harvested by data.europa.eu", the report prescribes weightage to each quality parameter. Finally, it provides a ranking quality based on the total score obtained. The scores are calculated in 5 parameters which are Findability (100), Accessibility (100), Interoperability (110), Reusability (75) and Contextuality (20), with a total score of 405 (Metadata quality assurance,2023).

The study aimed to focus on identifying the following objectives,

1. To identify the primary criteria that can be used to measure metadata quality.
2. To measure the metadata quality of ETDs in selected International Institutional Repositories.
3. To recognise the significant issues encountered in ensuring metadata quality.
4. To find out the primary mechanisms that can be used to improve metadata quality.



2. Methodology

2.1. Sample Selection

For this study, we have considered two types of population: International Institutional Repositories and International Libraries.

International Institutional Repositories	International Libraries
➢ The number of Open Access Repositories enlisted in OpenDOAR is 5981, out of which our study limited the population using the following parameters, Type of repository- Institutional, Software used - Dspace (as it is the most preferred software according to the statistics provided in OpenDOAR), Content type - Theses and Dissertation, Subject - Social Science (as Library and Information Science belongs to this discipline), which lead to a total number of repositories of 1327. Out of this, we selected our sample size using the following formula for finite population, $n' = \frac{n}{1 + \frac{z^2 \times p(1-p)}{\epsilon^2 N}}$ Where, z = Z Score, ϵ = Margin of Error, p = Population Proportion, N= Population Size For this study, we restricted the study with a Confidence level of 70%, a Margin of error of 10%, and a Population Proportion of 5%, which stands the sample n equals 6(six). Samples for the study are selected using the Simple Random Sampling method.	➢ Whereas in the case of International Libraries, the exact population is unknown. So, to calculate the sample size we have used the sample size calculation formula for an infinite population, $n = \frac{z^2 \times p(1-p)}{\epsilon^2}$ Where z = Z Score, ϵ = Margin of Error, p = Population Proportion. For this study, we restricted the study with a Confidence level of 75%, a Margin of error of 10%, and a Population Proportion of 5%, which stands the sample n equals 7(seven). Samples for the study are selected using the Simple Random Sampling method.

2.2. Data Harvesting

❖ For harvesting data, we have used MarcEdit 7.6.3 tool by Terry Reese.

❖ For retrieving the theses metadata from Institutional repositories, we have used the OAI Harvester plug-ins in MarcEdit; 4 sets of Theses metadata are harvested from each repository.

Name of the Repository	OAI-PMH URL	Count of Records
Agder University Research Archive (AURA)	https://uia.brage.unit.no/uia-oai/request	112
Brock University Digital Repository	http://dr.library.brocku.ca/oai/request	3310
Brunel University Research Archive (BURA)	https://bura.brunel.ac.uk/oai/request	386
Cranfield CERES	https://dspace.lib.cranfield.ac.uk/oai/request	877
DARIUS	https://darius.hbu.edu/oai/request	36
Bogor Agricultural University Repository (IPB Repository)	http://repository.ipb.ac.id/oai/request	602

While retrieving records of Libraries, Z39.50/SRU client plug-ins in MarcEdit are used; the keywords used for database search are- Theses, Dissertations, and academic to collect theses metadata.

Name of the Library	Host	Database	Port	Count of Records
Library of Congress	lx2.loc.gov	LCDB	210	170
British Library	z3950cat.bl.uk	ZBLACU	9909	5000
Trent University Library	ca01.alma.exlibrisgroup.com	01OCUL_TU	1921	1536
Virginia University Library	virgo.lib.virginia.edu	Unicorn	2200	5000
University of Exeter	lib.ex.ac.uk	INNOPAC	210	1000
University of Northern Colorado	source.unco.edu	INNOPAC	210	1000
University of Maryland	alephprod.umd.edu	CP	210	3408

2.3. Data Analysis: The present paper employed data analysis with Interpretive content analysis using four tools,

- MarcEdit- Using MarcEdit, we first converted the DCXML records of the Institutional Repositories to MARC21 records.
- Marc Report - We analysed each MARC21 record of IRs and Libraries using Marc Report utility plug-ins - Verify a MARC file and MARC Analysis.
- Metadata-Analyzer- A metadata-analyser is developed for the study to calculate the automated score of each IR and Library depending on the weightage of score distributions as prescribed on Metadata Quality Assessment provided by Consortium of data.europa.eu in July 2023. At first, we converted the XML files to CSV using OpenRefine. Then we imported the CSV files into a Relational Database Management System, and finally, the score was calculated using Database Query Analysis. European Commission provided five parameters to calculate the final score, which is calculated based on the field count and presence percentage using the following formula,

$$\frac{Query - based\ Record\ Count}{Global\ Record\ Count} \times Score\ prescribed\ by\ European\ Commission$$

- Lastly, a java based pre-compiled program by Peter Király has been used in this study. The program "Metadata-Quality analysis-Marc" is available on GitHub (https://github.com/pkiraly/metadata-qa-marc#configuration-1). Completeness of metadata in each thesis catalogue, Thomson-Trail completeness, and functional analysis of metadata is calculated for interpretive content analysis.

2.4. Data Validation: This step interprets and validates all the results.

2.5. Identification of Metadata Errors: Based on the results retrieved using the tools' errors are identified and presented in tabular form in the results and discussion section.

2.6. Suggestions for Metadata Quality Improvement: This section deals with suggestive remedial measures for enhancing metadata Quality based on contemporary revealing.

3. Results and Discussion

Table 1: Institutional Repository Record Structure

Institutional Repository	Total MARC record count	Average record length	Mean Average record length	Shortest record length	Longest record length	Number of records with 040 $e = 'rda'
Agder	112	2964	7945 (1 record with this length)	613 (record number 35)	7945 (record number 1)	0
Brock	3310	2368	2508 (11 records with this length)	456 (record number 626)	14684 (record number 605)	0
Brunel	386	3688	2889 (3 records with this length)	426 (record number 98)	10326 (record number 231)	0
Cranfield	877	3012	3702 (2 records with this length)	593 (record number 768)	7717 (record number 445)	0
Darius	1	2084	2084 (1 record with this length)	2084 (1 record with this length)	2084 (1 record with this length)	0
IPB	602	3545	7045 (3 records with this length)	445 (record number 470)	11857 (record number 381)	0

The Marc Report first verifies the Theses or Dissertation metadata of each International Institutional Repositories and Libraries and then analyses. The Average Record length is maximum in Brunel Institutional Repository among the Institutional Repositories and maximum in the University of Exeter in the case of libraries.

Table 2: Library Record Structure

Libraries	Total MARC record count	Average record length	Mean Average record length	Shortest record length	Longest record length	Number of records with 040 $e = 'rda'
British Library	5000	1301	574 (15 records with this length)	442 (record number 3492)	13748 (record number 861)	660
Library of Congress	170	1916	1576 (2 records with this length)	745 (record number 116)	6155 (record number 161)	14
University of Colorado	1000	3254	9780 (1 record with this length)	991 (record number 247)	11857 (record number 32)	930
University of Exeter	1000	4092	3569 (5 records with this length)	1056 (record number 292)	13060 (record number 656)	591
University of Maryland	3408	2164	2607 (10 records with this length)	537 (record number 2448)	17206 (record number 132)	235
Trent University	1536	1415	1452 (9 records with this length)	905 (record number 1226)	4430 (record number 1339)	3
University of Virginia	5000	1543	4663 (3 records with this length)	707 (record number 24)	18016 (record number 62)	399

The provided data outlines key characteristics of several institutional repositories based on their MARC record counts and various record length statistics. These statistics offer insights into the repositories' data distribution, average lengths, and outliers. The comparative analysis of each repository shows that the total number of record counts is much higher in the case of the libraries as it was downloaded using z39.50/SRU client. In contrast, in the case of repositories, only four sets of data are harvested using OAI-PMH. The 040 $e=rda is absent in all records of the institutional repository metadata.

In contrast, in some records of the library, the field is present with the highest records in the University of Colorado, 930 out of 1000 records. These statistics provide valuable information about the distribution of MARC records' lengths within each repository. They also highlight potential outliers or instances where certain records significantly deviate from the average. The presence of records with specific characteristics, such as records with 040 $e = 'rda', indicates adherence to cataloguing standards. These insights can be utilised by librarians, catalogers, and repository managers to understand the composition of their collections further and identify areas for improvement or refinement in metadata management practices.

Table 3: MARC tag Structure in Libraries and Institutional Repositories

Libraries	Total number of tags	Most repeated tag	Tags present in every record	Total number of subfield codes	Most repeated subfield code
British Library	121582	650 (41 times in record number 664)	001 005 008 245	214894	505 $t (96 times in record number 564)
Library of Congress	5516	991 (20 times in record number 61)	001 005 008 010 040 050 245 650	13105	850 $a (49 times in record number 64)
University of Colorado	34827	020 (14 times in record number 131)	008 040 245 300	67182	505 $t (58 times in record number 550)
University of Exeter	39860	653 (100 times in record number 570)	008 245 907	68354	505 $t (55 times in record number 774)
University of Maryland	116547	852 (76 times in record number 2021)	001 005 008 245	275134	505 $t (154 times in record number 2263)
Trent University	41256	992 (32 times in record number 1404)	001 008 035 245	300 852 992 993	78913 040 $d (25 times in record number 1492)
University of Virginia	134484	700 (90 times in record number 62)	001 008 245 926	222624	040 $d (149 times in record number 4697)
Agder	2700	787 (12 times in record number 80)	024 042 245 260 546 655 720 856	2970	024 $a (1 time in record number 1)
Brock	49887	653 (15 times in record number 2153)	024 042 245 260	59800	024 $a (1 time in record number 1)
Brunel	7627	856 (52 times in record number 293)	024 042 245 260 655	9057	024 $a (1 time in record number 1)
Cranfield	14653	856 (30 times in record number 792)	024 042 245 260 655 720	19182	024 $a (1 time in record number 1)
Darius	10	260 (3 times in record number 1)	024 042 245 260 520 546 720 856	11	024 $a (1 time in record number 1)
IPB	14443	856 (19 times in record number 146)	024 042 245 260 720 856	16368	024 $a (1 time in record number 1)

Table 3 shows that only 008 (Fixed Length Data Elements) and 245 (Title Statement) are present in each library record we studied. In the case of Institutional Repositories, 024 (Other Standard Identifier), 042 (Authentication Code), 245 (Title Statement), and 260 (Publication Statement) are present in each record of each institutional repository. Though in the case of theses, metadata publication statement is a mandatory

field to identify the publisher, which is not given in each record of the libraries. The record structure varies in the case of each record of the libraries, but the record structure is nearly identical in all records in the Institutional repositories. The University of Virginia has the highest number of tags used in its records, and the University of Maryland has the highest number of Subfield codes present in its records.

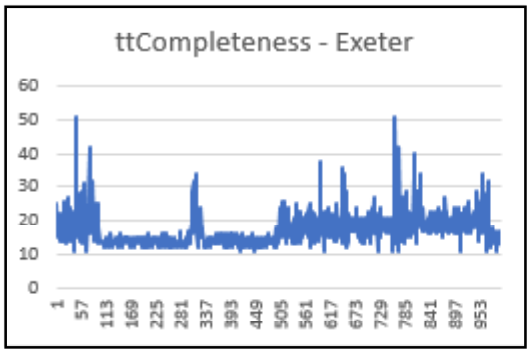
3.1. Match Key analysis:

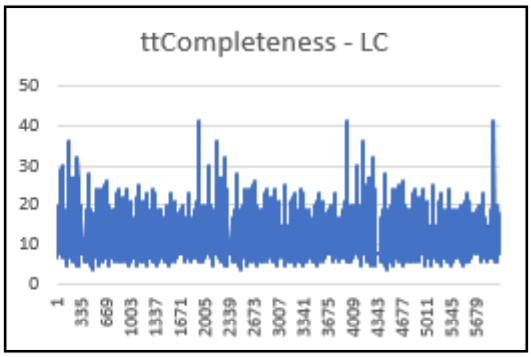
Table 4: Match Key Analysis of Library Metadata

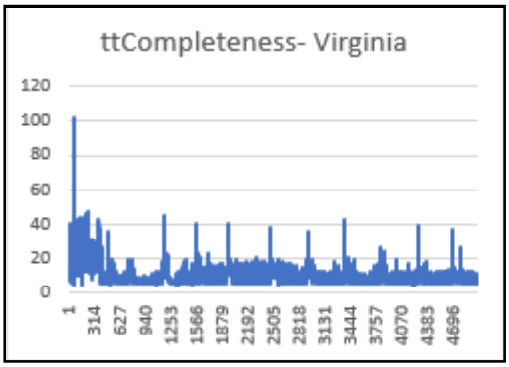
Libraries	Records without any Match Keys	Records with an LCCN	Records with an ISBN	Records with an ISSN	Records with an OCLC
British Library	1463	1145	3215	87	375
Library of Congress	0	170	30	33	90
University of Colorado	0	144	613	0	503
University of Exeter	2	33	998	0	169
University of Maryland	1271	1731	1441	112	515
Trent University	9	15	437	0	1470
University of Virginia	15	340	739	4	4753

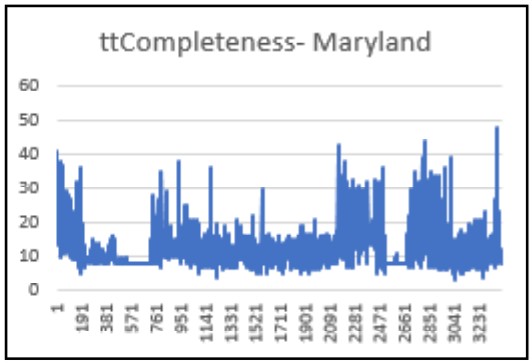
From the study, we found that default Match Keys are- LCCN (010a), ISBN (020a), ISSN (022a), and OCLC Number (001/035a), which are absent in the case of International Institutional Repository records but present in Library Marc records as shown in Table 4. Although in the case of Theses and Dissertations, the presence of match keys is not supposed to be present, the library's metadata provides those tags.

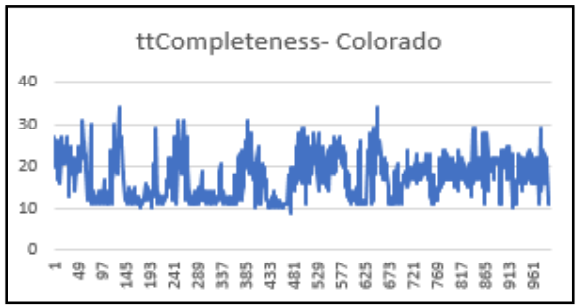
3.2. Thompson Trail Completeness (tt-Completeness)

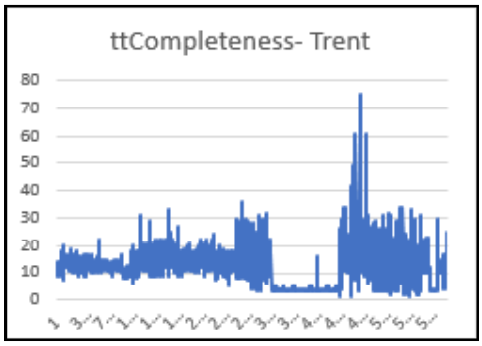


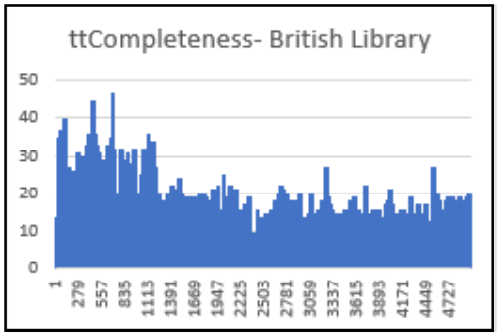












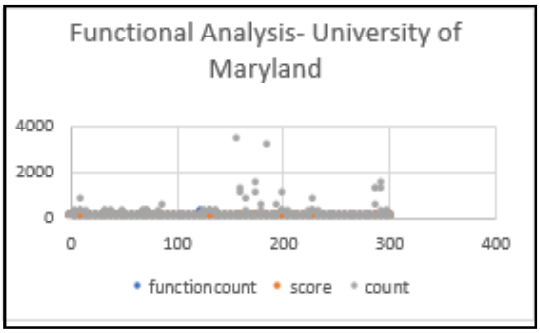
According to Thompson and Trail, the completeness of a record is calculated using the total score weightage given to certain tags of the MARC record. Thompson Trail's tt-Completeness metric provides a systematic approach for assessing the completeness of a MARC (Machine-Readable Cataloging) record based on various tags and fields within the record. The evaluation criteria are designed to gauge the presence of essential elements in the record, contributing to its overall quality and usability. The criteria focus on capturing critical bibliographic information to ensure the record is well-rounded and informative.

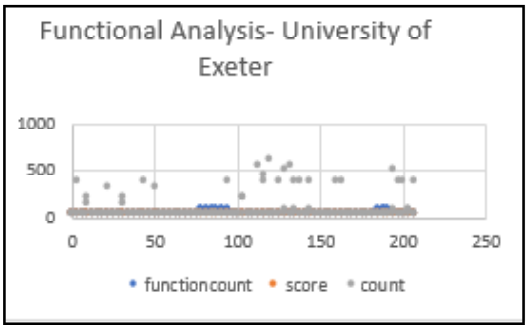
The evaluation criteria are as follows:

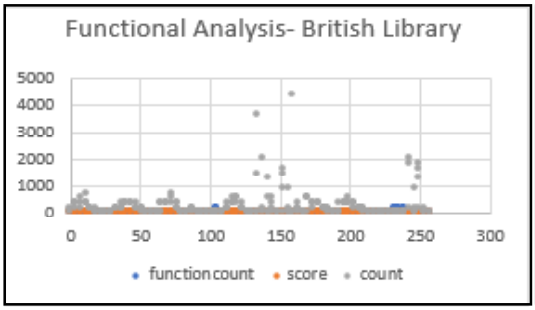
- ISBN: Assigns 1 point for each occurrence of the ISBN field (MARC 020).
- Authors: Assign 1 point for each occurrence of author-related fields (MARC 100, 110, 111).
- Alternative Titles: Assigns 1 point for each occurrence of the alternative titles field (MARC 246).
- Edition: Assigns 1 point for each occurrence of the edition field (MARC 250).
- Contributors: Assigns 1 point for each occurrence of contributor-related fields (MARC 700, 710, 711, 720).
- Series: Assigns 1 point for each occurrence of series-related fields (MARC 440, 490, 800, 810, 830).
- Table of Contents and Abstract: Awards 2 points if both fields (MARC 505, 520) exist; 1 point if either field exists.
- Date (MARC 008): Awards 1 point if valid coded data exists within the specified range (008/7-10).
- Date (MARC 26X): Awards 1 point if a 4-digit date exists within the specified field (260$c or 264$c) and matches the 008 date.
- LC/NLM Classification: Awards 1 point if any classification field (MARC 050, 060, 090) exists.
- Subject Headings: Awards 1 point for each relevant field based on specific indicators and subfields, depending on the classification scheme (LC, MeSH, FAST, Other).
- Description: Awards 2 points if both specified elements are present (008/23=o and 300$an “online resource”); 1 point if either element exists.
- Language of Resource: Awards 1 point if a likely language code exists within the specified range (008/35-37).
- Country of Publication Code: Awards 1 point if a likely country code exists within the specified range (008/15-17).
- Language of Cataloging: Awards 1 point if no language is specified or if English is specified in the 040$b field.
- Descriptive Cataloging Standard: Awards 1 point if the value in the 040$e field is “rda”.

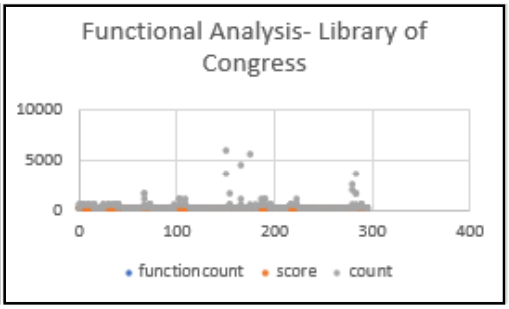
By applying these criteria, the tt-Completeness metric provides a quantifiable measure of the richness and completeness of a MARC record, aiding librarians, catalogers, and information professionals in evaluating and enhancing the quality of bibliographic data. It ensures that crucial information is captured accurately, improving the discoverability and Accessibility of resources within library collections.

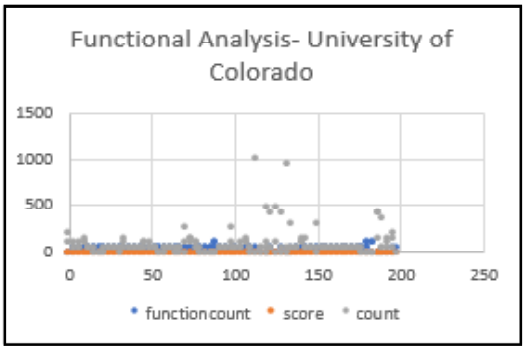
3.3 Functional Analysis

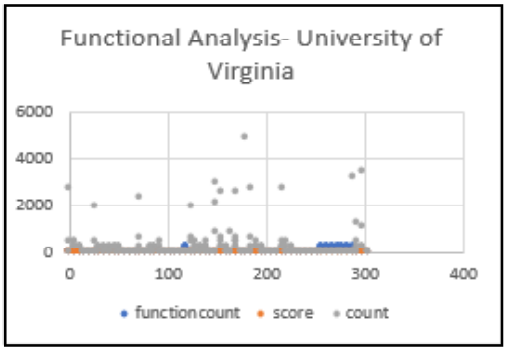


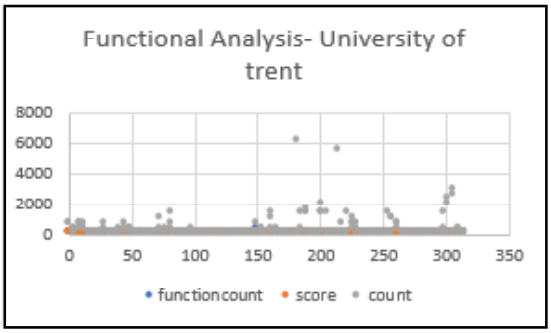












The Functional Requirements for Bibliographic Records (FRBR) document's central part defines the primary and secondary entities which became famous as FRBR models. Functional analysis of library metadata represents how metadata elements like title, author, ISBN/ISSN, and publication date contribute to identifying and differentiating resources from one another. The graphical representations provide detailed information about the content and characteristics of resources. This includes discovery search, discovery identify, discovery select, use restricted, use manage, use interpret, management identify, management process, management sort and management display. The analysis involves understanding how these elements help users assess the relevance and scope of a resource. The analysis includes evaluating how metadata aids users in navigating through collections by presenting hierarchical structures, related resources, and links. The functional analysis examines how metadata authority control mechanisms are implemented to prevent variations in author names (use restrict, use manage), subject headings (discovery identify), and other controlled vocabulary terms. The analysis also evaluates how metadata standards (e.g., MARC, Dublin Core, MODS, BibTeX) promote interoperability (management process, management identify) among library systems, databases, and repositories. Hence, functional analysis examines how metadata elements related to copyright, access restrictions, and licensing contribute to managing resource use's legal and ethical aspects. The functional analysis considers mechanisms for enriching metadata, such as annotations, user-generated tags, and linked data connections, to enhance resource descriptions and discoverability.

3.4 Score Analysis of IIR and IL

Table 5: Metadata Score Calculation

	Findability (100)				Accessibility (100)		Interoperability (110)				Reusability (75)			Contextuality (20)		Total (405)
Repository	Keyword Presence Score	Category Presence Score	Geo-Search Presence Score	Time Presence Score (20)	URL accessibility (100)		Format (50)	Media Type (15)	Media variation (15)	Machine Readability (30)	Access Restriction (-5)	Creator (20)	Publication (10)	Rights (10)	Date (10)	Score
Agder	21.21	14.38	1.36	6.24	43.19	25.97	50	10	3.97	30	13.04	12.29	3	2.9	3.12	**240.67**
Brock	27.38	7.82	20	13.7	68.9	26.11	50	10	10	30	11.68	5.44	2.54	2.6	6.85	**293.02**
Brunel	26.92	6.46	4.09	12.87	50.34	22.2	13.09	10	10	30	45	4.72	2.19	10	6.43	**254.31**
Cranfield	9.46	20.27	20	16.17	65.9	31.36	13.13	10	10	30	10.94	4.7	2.35	2.43	8.08	**254.79**
Darius	30	30	20	20	100	33.33	17.15	10	10	30	45	6.67	10	10	10	**382.15**
IPB	11.79	2.84	20	6.43	41.06	6.51	40.96	10	10	30	45	2.11	1.34	10	3.21	**241.25**
Libraries																
Library of Congress	30	30	20	20	100	100	50	10	10	30	45	20	10	10	10	495
British Library	21.21	14.38	1.36	6.24	43.19	25.97	50	10	3.87	30	13.04	12.29	3	2.9	3.12	**240.67**
University of Exeter	22.59	8.94	20	1.27	52.8	20.71	0.64	10	10	30	4.45	8.4	0.28	0.99	0.64	**191.71**
University of Trent	16.52	10.89	1.32	5.84	52.57	17.8	50	0.88	2.41	30	4.29	7.59	2.07	1.06	1.92	**183.10**
University of Virginia	10.08	8.37	1.13	4.31	23.89	14.45	50	10	1.65	30	45	4.81	1.41	10	2.15	**217.25**
University of Maryland	21.12	14.78	3.27	5.26	44.43	21.07	4.29	0.4	1.09	30	5.29	5.11	2.52	1.18	2.63	**162.44**
University of Colorado	17.33	9.82	20	1.74	48.89	21.23	50	10	1.08	30	1.72	5.5	1.87	0.38	0.87	**220.43**

The table comprehensively evaluates various repositories based on multiple metadata quality criteria. The assessment covers keyword presence, category presence, geo-search presence, time presence, URL accessibility, format, media type, media variation, machine readability, access restriction, creator information, publication details, rights information, and data accuracy. Each repository has been assigned a score based on these attributes. Several key observations can be drawn from the analysis:

Diversity of Metrics: The evaluation encompasses a wide range of metrics, reflecting the multifaceted nature of metadata quality. Attributes such as Accessibility, format, media type, and machine readability shed light on the technical aspects of metadata. At the same time, elements like keywords, categories, and geo-search presence highlight contextual relevance.

Variability in Scores: Repositories exhibit diverse levels of metadata quality across different attributes. Some repositories, like "Dairus," demonstrate strong scores across multiple categories, while others show varying strengths and weaknesses. For example, the "Library of Congress" and "Brock" repositories seem to have consistently high scores across most attributes.

Impact on Overall Quality: Each attribute contributes to the overall score, reflecting the holistic nature of metadata quality assessment. Attributes like URL accessibility, format, and media type are crucial for ensuring the availability and usability of data. At the same time, metadata completeness and accuracy are highlighted through attributes like keywords and time presence.

Contextual Considerations: Some attributes, such as access restriction and rights information, consider the repository's policies and governance. These aspects contribute to data security, sharing, and compliance, enhancing the overall quality of the repository's metadata.

Opportunities for Improvement: Repositories with lower scores in specific attributes have opportunities for improvement. Enhancing attributes like machine readability, media variation, and time presence can contribute to more comprehensive and valuable metadata.

Precisely, this metadata quality analysis underscores the importance of robust metadata management for effective data utilisation. Repository managers can use this evaluation as a roadmap to enhance metadata quality, promote data discoverability, ensure accurate interpretation, and foster more informed decision-making processes. It also emphasises the need to continuously monitor and refine metadata practices to adapt to evolving data needs and technological advancements.

Table 6: Overall Metadata Quality Score of International Institutional Repositories and Libraries

Repository	Total Score (405)	Rating
Agder	197.48	Sufficient
Brock	224.12	Good
Brunel	203.97	Sufficient
Cranfield	188.89	Sufficient
Darius	282.15	Good
IPB	200.19	Sufficient
Library of Congress	395	Excellent
British Library	197.38	Sufficient
University of Exeter	138.91	Sufficient
University of Trent	150.59	Sufficient
University of Virginia	193.36	Sufficient
University of Maryland	118.01	Bad
University of Colorado	171.54	Sufficient

The analysis of the provided metadata quality data showcases varying levels of data quality across different entities. The assessment indicates that institutions like the Library of Congress and Brock have demonstrated excellent and good metadata quality, respectively. These institutions have effectively maintained accurate, complete, and relevant metadata, enhancing the trustworthiness and utility of their data.

On the other hand, entities like the University of Maryland appear to have metadata quality concerns, as evidenced by the categorisation as "Bad". This suggests potential issues with their metadata's accuracy, completeness, and consistency, which could impact their data-driven activities and decision-making processes.

For many institutions, such as Agder, Brunel, IPB, British Library, University of Exeter, University of Trent, University of Virginia, and the University of Colorado, the metadata quality falls under the "Sufficient" category. While not optimal, this rating still indicates a certain degree of adherence to metadata quality standards. However, there might be room for improvement in refining metadata attributes to ensure better accuracy, consistency, and completeness.

4. Conclusion

In conclusion, the analysis of metadata quality plays a pivotal role in ensuring data accuracy, reliability, and usefulness in various domains. Through this comprehensive examination, organisations can assess metadata's completeness, consistency, accuracy, and relevancy, impacting the overall data integrity and decision-

making processes. A high metadata quality facilitates efficient data discovery, enhances data interoperability, and supports meaningful analytics, leading to better insights and informed decision-making.

This study provides a brief comparative account of Electronic Theses and Dissertation Metadata structure of Institutional Repositories and Libraries. As all the required tags are not present in the case of Institutional Repository theses metadata, the java-based pre-compiled program by Peter Király cannot be used in those records to determine the tt-completeness test and functional analysis of the records.

The process of metadata quality analysis involves a combination of automated tools and human expertise, ensuring a comprehensive evaluation of metadata attributes and relationships. Organisations must invest in continuous monitoring and improvement of metadata quality, as data environments are dynamic and constantly evolving.

By maintaining robust metadata quality, organisations can mitigate the risk of erroneous interpretations and foster greater trust in data-driven initiatives. This, in turn, enhances collaboration across departments, aids compliance with regulations, and contributes to developing more accurate and valuable data assets. In a data-driven world, where the quality of information is paramount, metadata quality analysis is a fundamental pillar for successful data management and utilisation. This metadata quality analysis underscores the importance of maintaining high-quality metadata across institutions to ensure reliable and meaningful data utilisation. The assessment can serve as a foundation for targeted improvements in metadata management practices, ultimately contributing to more effective data-driven operations and decision-making.

References

Choudhury, M. H., Salsabil, L., Jayanetti, H. R., Wu, J., Ingram, W. A., & Fox, E. A. (2023). MetaEnhance: Metadata Quality Improvement for Electronic Theses and Dissertations of University Libraries. http://arxiv.org/abs/2303.17661

Day, M., Guy, M., & Powell, A. (2004). Improving the quality of metadata in eprint archives. Ariadne, 38.

Elouataoui, W., El Alaoui, I., Gahi, Y. (2022). Metadata Quality in the Era of Big Data and Unstructured Content. In: Maleh, Y., Alazab, M., Gherabi, N., Tawalbeh, L., Abd El-Latif, A.A. (eds) Advances in Information, Communication and Cybersecurity. ICI2C 2021. Lecture Notes in Networks and Systems, vol 357. Springer, Cham https://doi.org/10.1007/978-3-030-91738-8_11

Király, P. (2019). Measuring metadata quality. http://hdl.handle.net/21.11130/00-1735-0000-0003-C17C-8

Metadata Quality Assurance. https://doi.org/10.08.2023

Park, J.-R. (2009). Metadata Quality in Digital Repositories: A Survey of the Current State of the Art. Cataloging & Classification Quarterly, 47(3-4), 213–228. doi:10.1080/01639370902737240

Romero-Pelaez, A., Segarra-Faggioni, V., & Alarcon, P. P. (2018). Exploring the provenance and accuracy as metadata quality metrics in assessment resources of OCW repositories. ACM International Conference Proceeding Series,

Spencer, S., & White, H. (2019). Automated Techniques for Measuring Metadata Quality. https://zenodo.org/record/3612497