





Do More Complete Dissertations' Metadata Get More Engagement?

Behrooz Rasuli ¹, Michael Boock ², Joachim Schöpfel ³, and Brenda Van Wyk ²

¹Iranian Research Institute for Information Science and Technology (IranDoc), rasuli@irandoc.ac.ir

²Oregon State University, michael.boock@oregonstate.edu

³University of Lille, joachim.schopfel@uni-lille.fr

⁴The University of Pretoria, brenda.vanwyk@up.ac.za

Author Note

We would like to thank the MIT Libraries for their help with our research. They answered our questions about the process of depositing dissertations. Additionally, we would like to acknowledge Dr. Amir Hossein Seddighi, an assistant professor at IranDoc, for his contributions to our data collection efforts. Correspondence concerning this article should be addressed to Behrooz Rasuli, Iranian Research Institute for Information Science and Technology (IranDoc), No. 11. Khajeh Nasir Ave. South Felestin St. Enqelab Eslami Ave., Tehran, Iran. Email: rasuli@irandoc.ac.ir

Abstract

This study investigates the role of metadata quality in Electronic Theses and Dissertations (ETDs), focusing on its completeness and its impact on discoverability and user engagement within institutional repositories (IRs). Using DSpace@MIT as a case study, the current research analyzed 22,276 doctoral dissertations to assess metadata completeness and its correlation with the number of views and downloads. Various metadata fields and usage statistics were extracted for detailed analysis. The study identified a moderate positive correlation between the number of unique metadata fields and both the Department Views Ratio (DVR) and Department Download Ratio (DDR), suggesting that enhanced metadata can improve the visibility and accessibility of dissertations. Additionally, the length of abstracts is positively correlated with engagement metrics. In contrast, title length does not significantly influence the visibility. These findings showed the importance of high-quality metadata in enhancing the discoverability of ETDs.

Keywords: Visibility, Bibliometrics, Open Repository, Open Science, Altmetrics, Research Impact.

Introduction

Over the past three decades, higher education institutions (HEIs) have increasingly adopted digital formats for theses and dissertations (TDs) to enhance accessibility, visibility, and impact. These Electronic Theses and Dissertations (ETDs) are now widely discoverable through various channels, including institutional ETD program portals (e.g. Electronic Theses & Dissertations at Johns Hopkins University), national ETD portals (e.g. ShodhGanga in India), regional ETD portals (e.g. DART-Europe E-theses Portal), and Institutional Repositories (e.g. DSpace@MIT). However, regardless of the access point, the quality of an ETD's metadata is crucial for several reasons, including discoverability, interoperability, assessment, and preservation.

Research Problem and Motivation

Repositories employ various methods to ensure ETDs are described thoroughly with quality metadata. These range from requesting researchers to provide more comprehensive information during ETD deposit to policy-driven approaches (Kasonde & Phiri, 2023) and even proposals for automated quality improvement (Choudhury et al., 2023). While a universally agreed-upon definition of metadata quality remains elusive, metadata quality in research and practice is often assessed through completeness, accuracy, consistency, accessibility, conformance, provenance, and timeliness (Kumar et al., 2024). Notably, accuracy, completeness, and consistency are the most emphasized criteria in the literature (Park, 2009). Additionally, Kasonde and Phiri (2023) emphasize the paramount importance of complete metadata for ETDs within IRs.

Despite extensive research on ETD metadata quality and its recognized importance, a gap exists in empirical studies on the impact of metadata completeness on ETD impact. This study aims to bridge this gap by investigating the relationship between ETD metadata completeness and the number of views/downloads in institutional repositories (IRs). The underlying assumption is that more complete metadata enhances ETD discoverability, leading to a potential increase in the number of views and subsequently the number of downloads.

This research will utilize dissertations archived on DSpace@MIT as a case study. Established in the early 2000s, DSpace@MIT is the institutional repository of the Massachusetts Institute of Technology (MIT) and houses scholarly works produced by its affiliated researchers. In addition to providing metadata for ETDs, DSpace@MIT leverages IRUS (Institutional Repository Usage Statistics) to track and report the number of views and downloads for each item within the repository (Roosa, 2024). As of April 26, 2024, DSpace@MIT has 22,353 doctoral dissertations in 30 distinct collections¹.

Purpose of the Present Study

The purpose of the present study is to investigate several key aspects of ETDs within the DSpace@MIT repository. Specifically, this research aims to determine the completeness of metadata for doctoral dissertations, identify the number of views and downloads of ETDs, and explore the relationships between metadata completeness and both the number of views and downloads. Additionally, the study will examine the correlation between the number of views and downloads of ETDs. Finally, the research seeks to identify which metadata fields within ETDs demonstrate consistently higher completeness rates compared to others.

Literature review

ETD was first introduced in the early 1990s to facilitate digital access to students' theses and dissertations (Rasuli et al., 2019). HEIs have increasingly adopted these digital formats to enhance accessibility, visibility, and impact of knowledge assets and scholarly communication availed in open access (Adam & Kaur, 2021; Van Wyk & Mostert, 2014). Its availability and accessibility are by now a well-entrenched practices within research repositories collections and services. Notable examples are institutional ETD program portals such as the ETDs at Johns Hopkins University; the national ETD portals of ShodhGanga in India; regional ETD portals such as the DART-Europe E-theses Portal. Then

¹ <https://dspace.mit.edu/handle/1721.1/131022>

there are also many Institutional Repositories of which the Massachusetts Institute of Technology boasts its DSpace@MIT as a high-functioning example.

Since 2000 several metadata standards were developed (Glogoff & Forger, 2000) offering broad sets of descriptive metadata elements as part of metadata structures for the creation of the database. Used with Web searching tools provide the granularity needed for resource discovery. Underpinned by the principle of openness, these digital collections aim to expand global e-Research networks. Openness is reliant on interoperability which will allow the systems to communicate with each other and communicate information back and forth in a usable format (Adam & Kaur, 2021).

Much of the previous research on ETDs focused on the implementation projects, success factors, and sustainability of ETDs, on a macro level. On a more micro level research on content as described by metadata as a measure of success has not been researched to its fullest. It stands to reason that the quality of an ETD's metadata is pivotal for discoverability, interoperability, assessment, and preservation. The interest in further researching ETDs is growing. However, Choudhury (2023) laments the gaps in research with regard to suitable research and frameworks to extract information from ETDs. In particular, he alludes to insufficient information on ETD segmentation, metadata extraction, metadata quality improvement, and parsing reference strings (Choudhury, 2023). Literature highlights the need for further research in metadata quality, of which the completeness of metadata as an indicator of metadata quality, is an area urgently needing more insights.

Metadata quality in research and practice is often assessed through completeness, accuracy, consistency, accessibility, conformance, provenance, and timeliness (Kumar et al., 2024). Notably, accuracy, completeness, and consistency are the most emphasized criteria in the literature (Park, 2009). Additionally, Kasonde and Phiri (2023) emphasize the paramount importance of complete metadata for ETDs within IRs.

Research is replete of factors that may negatively impact quality of metadata assigned in TDs (Chisale & Phiri, 2023; Kasonde & Phiri, 2023; Osman et al., 2023; Park & Richard, 2011). The many

different fields and research disciplines represented in ETDs often have varying standards, and if standards are not in place, consistency and interoperability may be compromised.

Open access harvesters such as OAI-PMH are crucial protocols for building connected and interoperable digital information infrastructure, and their usefulness largely depends on the quality and consistency of the metadata provided by the repositories. On a global scale metadata harvesting of downstream aggregation services such as the Networked Digital Library of Theses and Dissertations (NDLTD)'s Union Catalog and the Open Access Theses and Dissertations portal cannot harvest poor quality metadata (Chisale & Phiri, 2023).

This study investigates the relationship between ETD metadata completeness and its impact on discoverability within the institutional repository of the Massachusetts Institute of Technology (MIT). As a private land-grant research university MIT was established in 1861 in Massachusetts, USA. Being a research-intensive university with valuable research and knowledge assets in inter alia oceanography, the need to establish an electronic TD repository was realised and planning started in 2000. The DSpace open-access repository was launched in 2002 (Baudoin & Branschofsky, 2003). The adoption of the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) was deployed. The OAI Registry includes DSpace, making its Dublin-Core-formatted metadata available to compatible harvesting code. *"DSpace@MIT contains more than 53,000 selected theses and dissertations from all MIT departments. The DSpace@MIT thesis community does not contain all MIT theses."* (MIT Libraries, 2024a) MIT Libraries encourage students to *"apply appropriate accessibility features and metadata into [their] thesis document."* (MIT Libraries, 2024b)

The staff at the MIT Libraries is responsible for ensuring the accuracy and completeness of the final metadata for all thesis records. Since 2022, MIT has encouraged authors to submit their own metadata to the Libraries when they deposit their theses, although this submission is not mandatory. Even if authors do not provide metadata, the Libraries staff will continue to review and update the records, adding any missing information as necessary to maintain the quality and discoverability of the theses in the repository.

Method

Data Collection

This study collected data on 22,276 doctoral theses from the DSpace@MIT repository. The data collection process involved extracting various metadata fields and usage statistics for each dissertation. MIT was chosen as a case study for several reasons:

- **High Visibility:** As a renowned institute with international reach, MIT dissertations likely attract a significant number of views and downloads, providing a robust dataset for analysis.
- **Controlled Variables:** Focusing on a single institution allows for control over the "reputability of institute" variable, potentially mitigating its influence on the findings.
- **Standardized Format:** Limiting the study to doctoral dissertations ensures a consistent type of work across the sample, minimizing the impact of document type as a confounding factor.
- **Data Availability:** DSpace@MIT offers not only comprehensive metadata for dissertations but also readily accessible usage statistics for each document, facilitating data collection for both metadata completeness and discoverability measures.

The following steps outline the data collection methodology:

1. **Data Source:** The primary source of data was the DSpace@MIT repository, specifically the URL format: <https://dspace.mit.edu/handle/1721.1/XXXXX?show=full>, where XXXXX represents the unique ID for each dissertation. To extract the IDs, the Doctoral Theses collection in DSpace@MIT (<https://dspace.mit.edu/handle/1721.1/131022/recent-submissions>) was browsed, and all submitted records until August 21, 2024, were retrieved.
2. **Metadata Fields:** The study focused on metadata fields that begin with "dc." (Dublin Core), which are standard fields used for describing digital resources.
3. **Data Extraction:** The data extraction process was conducted from August 10 to August 21, 2024. Each dissertation's metadata and usage statistics were retrieved and recorded.

Data Organization

Once the data was collected, it was organized into an Excel file with the following columns for each dissertation:

1. **Record ID:** A unique identifier assigned by DSpace@MIT to each record.
2. **Department:** This refers to the academic department at MIT associated with the dissertation. The original dataset included more than 70 different values for this field. However, some dissertations contained misspellings or variations in the names of specific departments. To address this issue, all department and program names were standardized by consulting the MIT website, resulting in a cleaned and unified dataset.
3. **Date Available:** The date the dissertation became publicly accessible through DSpace@MIT. 2005 is the first year for the availability of dissertations in the current study's dataset. For dissertations lacking a specific "Date Available," the study utilized the "Date Issued" or "Date Copyright" fields as substitutes to ensure completeness in the dataset.
4. **Number of Unique Metadata Fields:** Count of distinct metadata properties filled out, excluding duplicates. For example, if a specific field (e.g. dc.description.abstract) is filled out twice or more for one record, it was considered as one field in the counting process.
5. **Number of Duplicated Metadata Fields:** Total count of all metadata fields filled out, including duplicates.
6. **Abstract Word Count:** Total word count of the dissertation's abstract, considering all fields where the abstract may be recorded.
7. **Title Word Count:** Total word count of the dissertation's title.
8. **Number of Downloads:** Total full-text downloads recorded for the dissertation.
9. **Department Download Ratio (DDR):** DDR is a download-based measure of the impact of one record and it indicates the relative download performance of a dissertation when compared to similarly-aged dissertations in its department. It is calculated by dividing the number of downloads by the geometric mean of downloads for similarly-aged dissertations in the same

department. This indicator ensures the normalization of the number of downloads across different departments and years.

10. **Number of Views:** Total page views recorded for the dissertation.

11. **Department Views Ratio (DVR):** DVR is a view-based measure of the visibility of one record and it indicates the relative page view performance of a dissertation when compared to similarly-aged dissertations in its department. It is calculated by dividing the number of views by the geometric mean of views for similarly-aged dissertations in the same department. This indicator ensures normalization of the number of page views across different departments and years.

Data Analysis

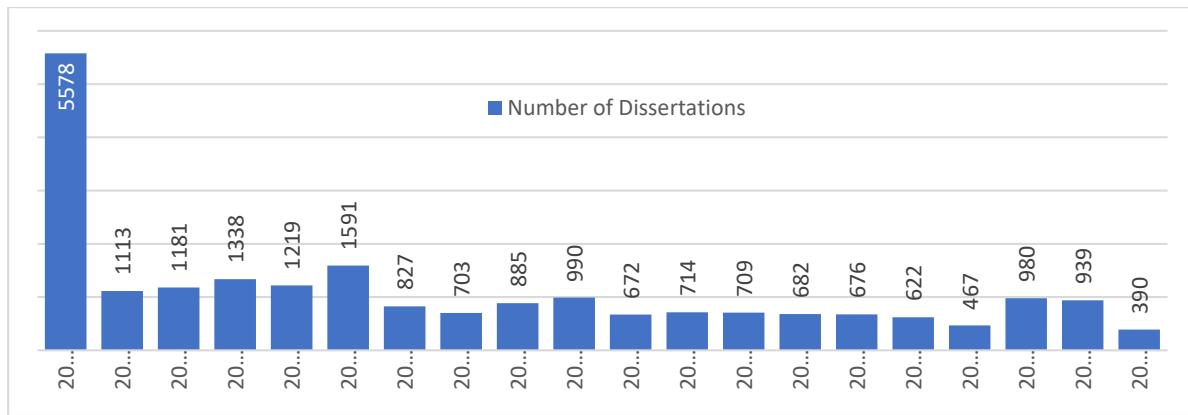
The analysis focused on several key aspects to understand the relationship between metadata completeness and user engagement. First, the study assessed metadata completeness by analyzing the number of unique and duplicated metadata fields for each record across the dataset and through descriptive statistics. Next, the analysis examined the correlation between metadata completeness and usage statistics, specifically the number of page views and downloads. Statistical tests were conducted to determine whether there was a significant relationship between the completeness of metadata and these engagement metrics. This step was crucial for understanding how metadata quality might influence user interaction with the dissertations.

Results

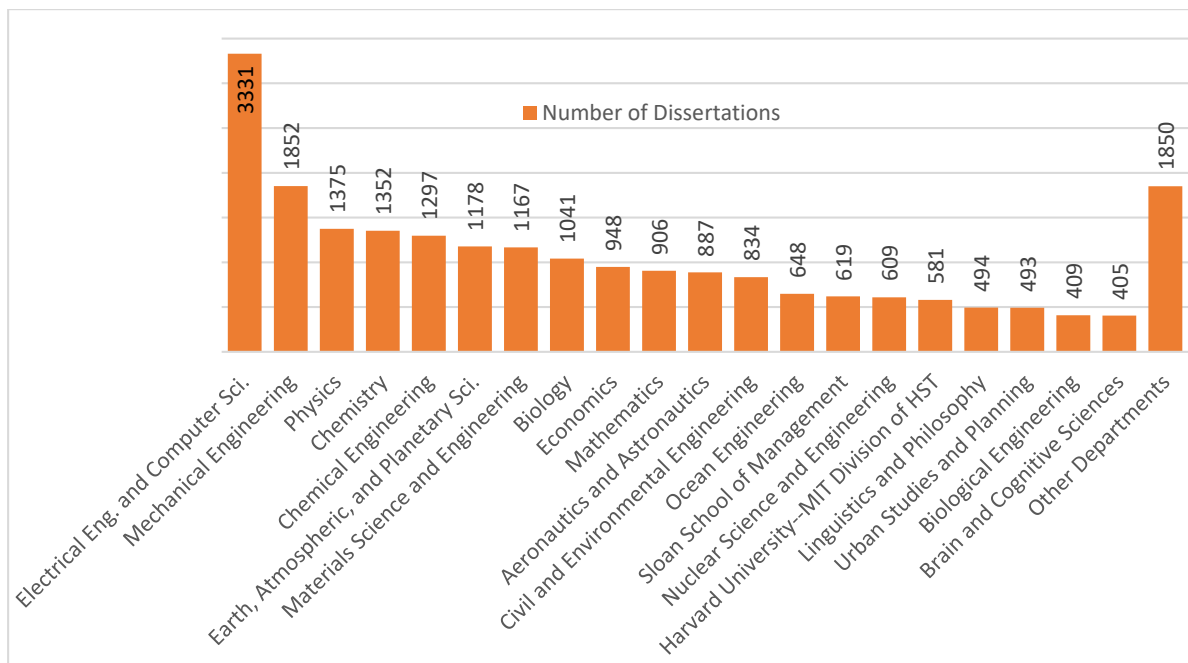
As stated earlier, this study collected data on 22,276 doctoral theses from the DSpace@MIT repository. Figure 1 and Figure 2 illustrate the frequency of dissertations across various departments and years. Notably, the Department of Electrical Engineering and Computer Science stands out with a total of 3,331 dissertations. Additionally, the year 2005 was significant, as it saw the release of 5,578 dissertations through DSpace@MIT.

Figure 1

Distribution of Dissertations by Year at DSpace@MIT

**Figure 2**

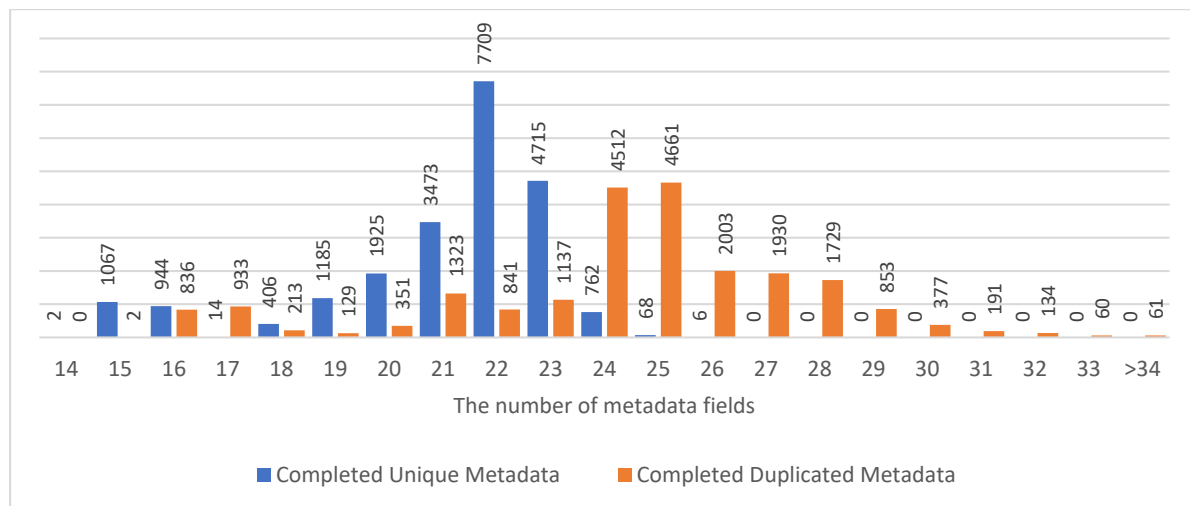
Distribution of Dissertations by Department at DSpace@MIT



The analysis of unique and duplicated metadata fields completed for the doctoral dissertations in the DSpace@MIT repository reveals significant insights into the quality of metadata associated with these academic works. Figure 3 presents the distribution of completed unique and duplicated metadata fields across the dissertations.

Figure 3

Distribution of Completed Unique and Duplicated Metadata Fields in Doctoral Dissertations at DSpace@MIT



According to Figure 3, The minimum number of unique metadata fields completed is 14, while the maximum is 26. However, the majority of dissertations, totaling 7,709, have 22 unique metadata fields completed. On the other hand, the number of duplicated metadata fields varies significantly, with a minimum of 15 fields and a maximum of 41 fields completed. Notably, the highest frequency of duplicated fields occurs at 24 fields, with 4,512 instances recorded. This suggests that many dissertations include multiple entries for certain metadata categories, which may enhance the detail and context of the information provided.

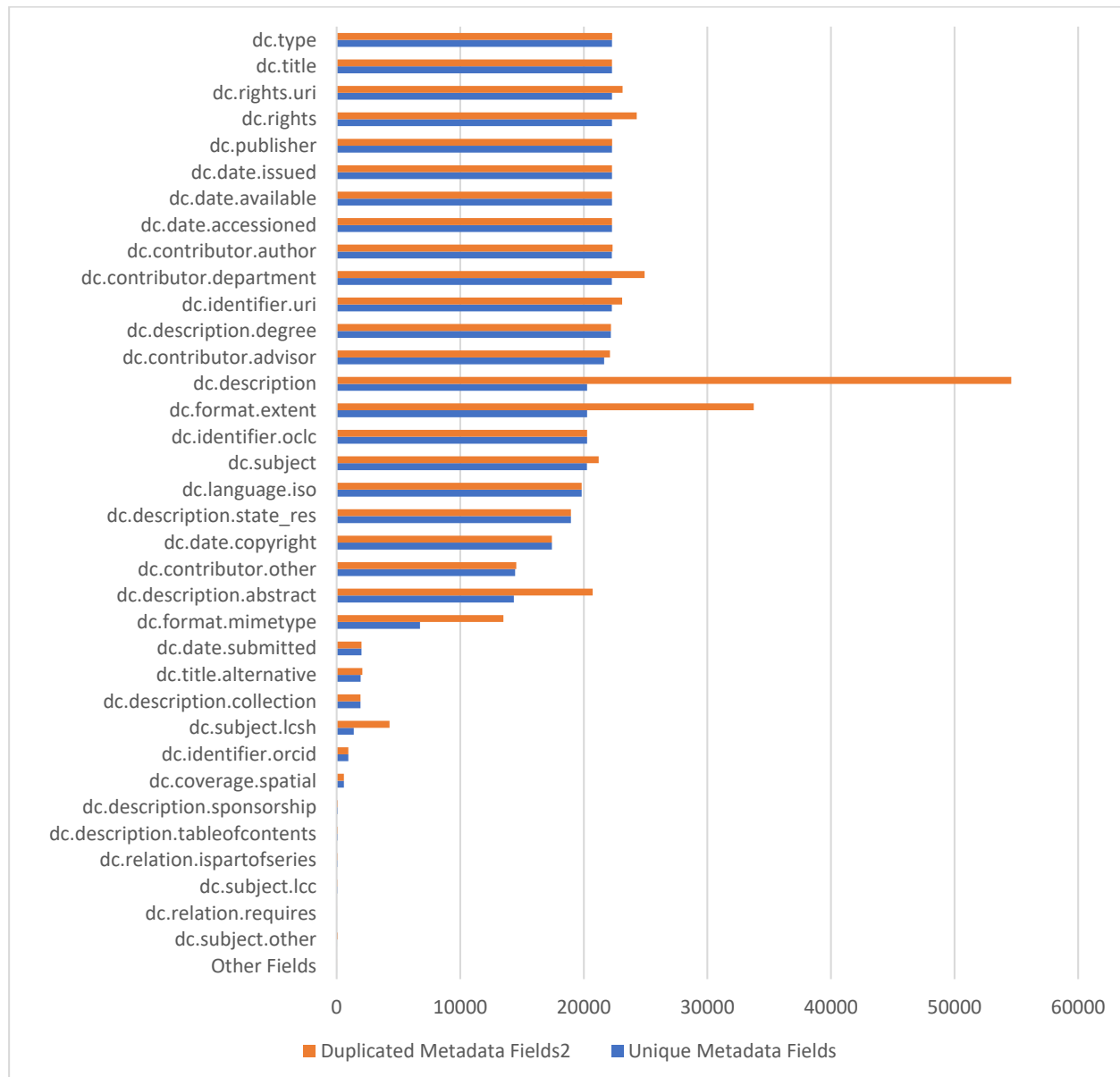
Another important aspect of examining metadata completeness is identifying which metadata fields are filled out most frequently. Figure 4 presents the results of the analysis regarding the frequency of various metadata fields that are completed for the dissertations in DSpace@MIT.

Figure 4 illustrates the frequency of unique and duplicated metadata fields for doctoral dissertations in the DSpace@MIT repository. Each entry represents a specific metadata field, displaying the count of unique entries alongside the total number of duplicated entries for that field. For example, the "dc.description.abstract" field shows a significantly high count of 14,343 unique entries and 20,714 duplicated entries (it means that there are 6,371 records with two or more than two fields for "dc.description.abstract"), indicating its frequent use. In contrast, fields such as "dc.relation.requires" and "dc.subject.other" exhibit lower counts, suggesting they are less commonly filled out. The other fields (e.g. dc.contributor, dc.date, dc.date.created, dc.date.updated,

dc.identifier.govdoc, dc.relation, dc.language, dc.audience.educationlevel, and dc.identifier.other) are filled out for only 27 dissertations.

Figure 4

Frequency of Unique and Duplicated Metadata Field Types for Doctoral Dissertations in DSpace@MIT



The visibility and impact of academic research can be significantly influenced by the number of downloads and page views that dissertations receive.

In order to investigate the possible correlations between dissertation metadata quality and user engagement, statistical analyses are needed. As a first step, examining descriptive statistics provides a foundational understanding of the distribution and variability of key variables across the

dataset. Table 1 provides a detailed snapshot of the dissertations' metadata quality and user engagement levels within the DSpace@MIT repository.

Table 1

Descriptive Statistics of Dissertation Metadata and Access Metrics in DSpace@MIT

Variables	Descriptive Statistics								
	<i>N</i>	<i>Min</i>	<i>Max</i>	<i>Sum</i>	<i>Mean</i>	<i>Median</i>	<i>SD</i>	<i>Variance</i>	<i>SK</i>
Downloads	22276	0	151810	14509563	651.35	403.0	2053.36	4216289.89	41.60
DDR	22276	0	301	31364	1.41	0.918	3.78	14.26	43.52
Page View	22276	0	195261	21646075	971.72	702.0	1868.49	3491265.40	59.96
DVR	22276	0	180	27153	1.22	0.949	1.89	3.55	52.43
Unique Metadata	22276	14	26	470818	21.14	22.00	2.19	4.80	-1.47
Duplicated Metadata	22276	15	41	542847	24.37	25.00	3.37	11.38	-.58
Abstract Word Count	22276	0	1973	4852788	217.85	251.0	195.76	38320.78	.61
Title Word Count	22276	1	43	220604	9.90	9.000	3.96	15.70	.73

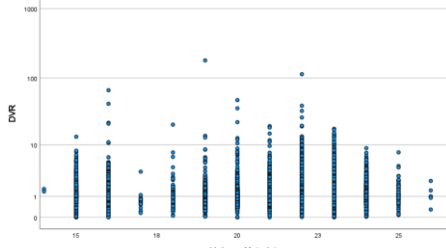
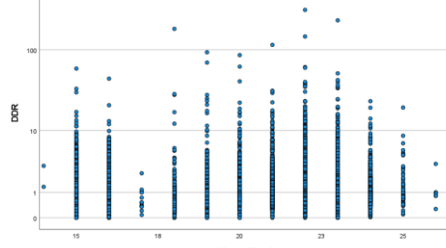
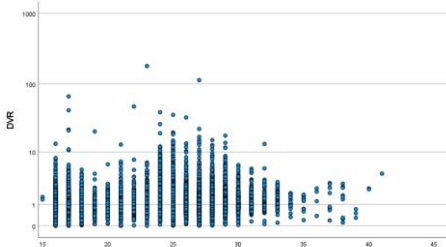
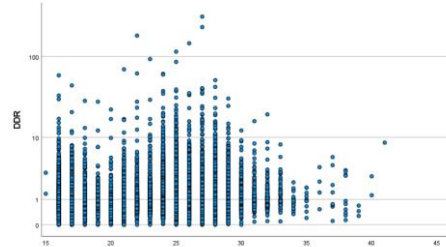
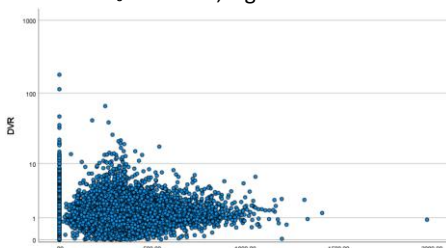
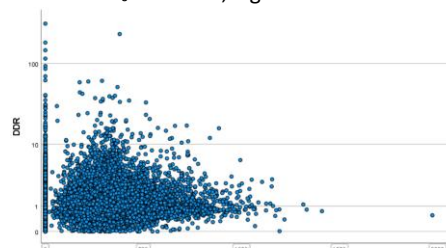
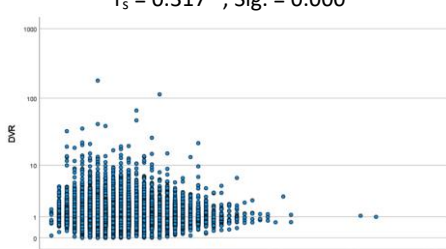
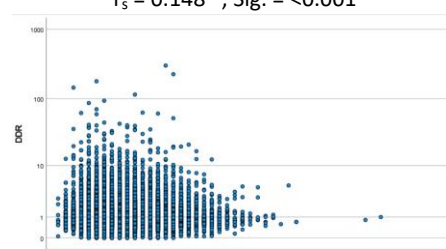
According to Table 1, Downloads range from 0 to 151,810, with a mean of 651.35 and a median of 403.0, indicating a right-skewed distribution (skewness of 41.60) where a small number of dissertations have exceptionally high download counts. Similarly, page views vary from 0 to 195,261, with a mean of 971.72 and a median of 702.0, and an even higher skewness of 59.96. The number of unique metadata fields completed per dissertation ranges between 14 and 26, with an average of 21.14 and a median of 22.00. Document characteristics also exhibit notable findings. The average abstract word count is 217.85 words, with a median of 251 and a maximum of 1,973 words. Titles average 9.90 words in length, with a median of 9 and a maximum of 43 words.

To further explore the relationships between dissertation metadata quality and user engagement, a Spearman rank correlation test was conducted. The results of the Spearman correlation analysis are summarized in Table 2. The table presents the correlation coefficients (r_s) along with their significance levels (p-values [Sig. (2-tailed)]).

DDR and DVR are important metrics used to assess the performance of dissertations within their respective academic departments. Both ratios provide a normalized measure of engagement, allowing for fair comparisons among dissertations that may differ significantly in terms of their age, department, and overall visibility.

Table 2

Correlation Between Dissertation Metadata Quality/Characteristics and Engagement Metrics

Variables	DVR (N = 22,276)	DDR (N = 22,276)
Number of Unique Metadata	 <p>$r_s = 0.204^{**}$; Sig. = <0.001</p>	 <p>$r_s = 0.079^{**}$; Sig. = <0.001</p>
Number of Duplicated Metadata	 <p>$r_s = 0.116^{**}$; Sig. = <0.001</p>	 <p>$r_s = 0.052^{**}$; Sig. = <0.001</p>
Number of Abstract's Word	 <p>$r_s = 0.317^{**}$; Sig. = 0.000</p>	 <p>$r_s = 0.148^{**}$; Sig. = <0.001</p>
Number of Title's Word	 <p>$r_s = -0.009$; Sig. = 0.204</p>	 <p>$r_s = -0.048^{**}$; Sig. = <0.001</p>

****** Correlation is significant at the 0.01 level (2-tailed).

According to Table 2, There is a moderate positive correlation between the number of unique metadata fields and both DVR ($r_s = 0.204$ and Sig. = <0.001) and DDR ($r_s = 0.079$ and Sig. = <0.001). This indicates that dissertations with a greater number of unique metadata fields tend to have higher visibility (DVR) and download performance (DDR) compared to their peers. The stronger correlation with DVR suggests that unique metadata may be particularly effective in attracting views, which could lead to increased downloads. In additions, a positive correlation exists between the number of duplicated metadata fields and both DVR ($r_s = 0.116$ and Sig. = <0.001) and DDR ($r_s = 0.052$ and Sig. = <0.001), although the correlation is weaker than that of unique metadata.

The findings in Table 2 show that the number of words in the abstract shows a strong positive correlation with DVR ($r_s = 0.317$ and Sig. = 0.000) and a moderate correlation with DDR ($r_s = 0.148$ and Sig. = <0.001). This indicates that dissertations with longer abstracts are likely to have higher visibility and download performance. The correlation between the number of words in the title and DVR is not statistically significant ($r_s = -0.009$ and Sig. = 0.204), indicating that title length does not appear to influence visibility. However, there is a weak negative correlation with DDR ($r_s = -0.048$ and Sig. = <0.001), suggesting that longer titles may be associated with slightly lower download rates relative to the number of documents.

Discussion

Metadata serves as a fundamental pillar in our information-centric society, facilitating the efficient organization, retrieval, and utilization of information (Tani et al., 2013). To keep ETDs usable, it is essential to have high-quality metadata for these digital items. The ETD community employs a variety of metadata practices at national, regional, and international levels. The criteria for what defines minimal, good, and optimal metadata—regardless of format or schema—varies based on numerous factors, which can differ between countries and institutions. High-quality metadata is crucial for creating a union catalog, which is essential for library networking and resource sharing. Although ETDs are managed by the institutions that produce them, it is possible to present them as a unified collection through a central search engine that aggregates all the

metadata. When users find a relevant document, they are redirected to the institution that holds it (Alemneh et al., 2014).

The findings revealed that out of the 129 metadata fields in the Dublin Core schema (DCMI Usage Board, 2020), 44 fields are utilized to describe MIT's dissertations. The average completeness of unique fields across the records is 21, while the average for duplicated fields is 24. These results indicate that the metadata for dissertations in DSpace@MIT is not fully complete, suggesting potential areas for improvement in metadata quality.

The moderate positive correlation between the number of unique metadata fields and both DVR and DDR indicates that more detailed metadata can (relatively) enhance the visibility and accessibility of dissertations. Completeness in metadata refers to the inclusion of all necessary information that accurately describes a resource (Ochoa & Duval, 2009). When metadata is comprehensive, it enhances the ability of users to find and access relevant materials more easily. Moreover, complete metadata not only facilitates easier discovery but also contributes to the credibility of the repository (Fear & Donaldson, 2012). When users are provided with sufficient information about a dissertation, they will have a clear understanding of the content of the described resource, potentially leading to more downloads (or helping users determine if the dissertation is worth reading). According to Tani et al. (2013), the quality of metadata directly impacts the discoverability of information resources.

The correlation between abstract length and engagement metrics suggests that more detailed abstracts may attract greater attention, as this factor can predict a higher score of DVR. The Abstract is one of the primary marketing elements of any scientific output (Pottier et al., 2023). When abstracts contain a rich array of pertinent terms and phrases, they become more likely to be indexed effectively by web search engines and the DSpace@MIT repository's search engine. This increased visibility can lead to higher search rankings, making it easier for potential readers to find the dissertation. Moreover, the presence of relevant terminology not only aids in search engine optimization (SEO) but also ensures that the content aligns with the interests and queries of the

target audience. By using specific and widely recognized terms related to their field, researchers can attract a more relevant readership, thereby increasing the likelihood of engagement with their work. Previous studies have found that the words in abstracts correlate with the citations of research outputs, with citation counts increasing steadily as abstract length increases (Robson & Mousquès, 2016; Sohrabi & Iraj, 2017).

Interestingly, the analysis reveals that title length does not significantly impact visibility, and longer titles may even correlate with lower download rates. This finding challenges the assumption that more descriptive titles necessarily lead to higher engagement, suggesting that brevity may be more effective in capturing reader interest. Supporting this finding, research by Paiva et al. (2012) found that articles with shorter titles had higher viewing and citation rates compared to those with longer titles.

In conclusion, this study showed the key role of metadata completeness in enhancing user engagement within the DSpace@MIT repository. We found that well-structured and comprehensive metadata significantly influences the visibility and accessibility of dissertations, ultimately impacting their retrieval and usage.

References

- Adam, U. A., & Kaur, K. (2021). Institutional repositories in Africa: Regaining direction. *Information Development*, 38(2), 166-178. <https://doi.org/10.1177/02666669211015429>
- Alemneh, D., Donovan, B., Halbert, M., Han, Y., Henry, G., Hswe, P., McMillan, G., & (Lucy) Wang, X. (2014). *Guidance Documents for Lifecycle Management of ETDs* (M. Schultz, N. Krabbenhoft, & K. Skinner, Eds.). Educopia Institute. [https://educopia.org/wp-content/uploads/2018/07/Guidance Documents for Lifecycle Management of ETDs_0.pdf](https://educopia.org/wp-content/uploads/2018/07/Guidance_Documents_for_Lifecycle_Management_of_ETDs_0.pdf)
- Baudoin, P., & Branschovsky, M. (2003). Implementing an Institutional Repository. *Science & Technology Libraries*, 24(1-2), 31-45. https://doi.org/10.1300/J122v24n01_04
- Chisale, A., & Phiri, L. (2023). *Towards Metadata Completeness in National ETD Portals for Improved Discoverability* 26th International Symposium on Electronic Theses and Dissertations (ETD2023), Gandhinagar, Gujarat, India. <https://ir.inflibnet.ac.in/handle/1944/2412?mode=full>
- Choudhury, M. H. (2023, June 26–30, 2023). *ETDSuite: An Library for Mining Electronic Theses and Dissertations* JCDL'23, Santa Fe, New Mexico. https://www.cs.odu.edu/~cs_mchou001/website/resources/paper/JCDL_2023_DC-final.pdf
- Choudhury, M. H., Salsabil, L., Jayanetti, H. R., Wu, J., Ingram, W. A., & Fox, E. A. (2023, 26-30 June 2023). MetaEnhance: Metadata Quality Improvement for Electronic Theses and Dissertations of University Libraries. 2023 ACM/IEEE Joint Conference on Digital Libraries (JCDL),
- DCMI Usage Board. (2020). *DCMI Metadata Terms*. Dublin Core Metadata Initiative (DCMI). Retrieved August 20 from <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>
- Fear, K., & Donaldson, D. R. (2012). Provenance and credibility in scientific data repositories. *Archival Science*, 12(3), 319-339. <https://doi.org/10.1007/s10502-012-9172-7>
- Glogoff, S. J., & Forger, G. J. (2000). Metadata Protocols and Standards. *Internet Reference Services Quarterly*, 5(4), 5-14. https://doi.org/10.1300/j136v05n04_03
- Kasonde, C. C., & Phiri, L. (2023). *Assessing and Promoting Metadata Quality for Electronic Theses and Dissertations in Institutional Repositories Using a Policy-Driven Approach* 26th International Symposium on Electronic Theses and Dissertations (ETD2023), Gandhinagar, Gujarat, India. <https://ir.inflibnet.ac.in/handle/1944/2412?mode=full>
- Kumar, V., Chandrappa, & Harinarayana, N. S. (2024). Exploring dimensions of metadata quality assessment: A scoping review. *Journal of Librarianship and Information Science*, 09610006241239080. <https://doi.org/10.1177/09610006241239080>
- MIT Libraries. (2024a). *About MIT Theses in DSpace@MIT*. MIT Libraries. Retrieved September 01 from <https://libguides.mit.edu/dspace/about-theses>
- MIT Libraries. (2024b). *MIT Thesis FAQ: Thesis Checklist*. MIT Libraries. Retrieved September 01 from <https://libguides.mit.edu/mit-thesis-faq/checklist>
- Ochoa, X., & Duval, E. (2009). Automatic evaluation of metadata quality in digital repositories. *International Journal on Digital Libraries*, 10(2), 67-91. <https://doi.org/10.1007/s00799-009-0054-4>
- Osman, R., K., Y. I. A. M., & A, A. (2023). Metadata matters: evaluating the quality of Electronic Theses and Dissertations (ETDs) descriptions in Malaysian institutional repositories. *Malaysian Journal of Library and Information Science*, 28(1), 109-125. <https://doi.org/10.22452/mjlis.vol28no1.7>
- Paiva, C. E., Lima, J. P. d. S. N., & Paiva, B. S. R. (2012). Articles with short titles describing the results are cited more often [10.6061/clinics/2012(05)17]. *Clinics*, 67(5), 509-513. [https://doi.org/10.6061/clinics/2012\(05\)17](https://doi.org/10.6061/clinics/2012(05)17)

- Park, E. G., & Richard, M. (2011). Metadata assessment in e-theses and dissertations of Canadian institutional repositories. *The Electronic Library*, 29(3), 394-407.
<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*, 47(3-4), 213-228.
<https://doi.org/10.1080/01639370902737240>
- Pottier, P., Lagisz, M., Burke, S., Drobnik, S. M., Downing, P. A., Macartney, E. L., Martinig, A. R., Mizuno, A., Morrison, K., Pollo, P., Ricolfi, L., Tam, J., Williams, C., Yang, Y., & Nakagawa, S. (2023). Keywords to success: a practical guide to maximise the visibility and impact of academic papers. *bioRxiv*, 2023.2010.2002.559861.
<https://doi.org/10.1101/2023.10.02.559861>
- Rasuli, B., Solaimani, S., & Alipour-Hafezi, M. (2019). Electronic Theses and Dissertations Programs: A Review of the Critical Success Factors. *College & Research Libraries*, 80(1), 60-75.
<https://doi.org/10.5860/crl.80.1.60>
- Robson, B. J., & Mousquès, A. (2016). Can we predict citation counts of environmental modelling papers? Fourteen bibliographic and categorical variables predict less than 30% of the variability in citation counts. *Environmental Modelling & Software*, 75, 94-104.
<https://doi.org/https://doi.org/10.1016/j.envsoft.2015.10.007>
- Roosa, S. (2024, April 18, 2024). *Institutional Repository Usage Statistics (IRUS) at DSpace@MIT* Third Annual LyrOpen Fair, Virtual Conference.
- Sohrabi, B., & Iraj, H. (2017). The effect of keyword repetition in abstract and keyword frequency per journal in predicting citation counts. *Scientometrics*, 110(1), 243-251.
<https://doi.org/10.1007/s11192-016-2161-5>
- Tani, A., Candela, L., & Castelli, D. (2013). Dealing with metadata quality: The legacy of digital library efforts. *Information Processing & Management*, 49(6), 1194-1205.
<https://doi.org/https://doi.org/10.1016/j.ipm.2013.05.003>
- Van Wyk, B., & Mostert, J. (2014). African Institutional Repositories as Contributors to Global Information: A South African Case Study. *Mousaion: South African Journal of Information Studies*, 32(1), 98-114. <https://doi.org/10.25159/0027-2639/1704>