

Fall 11-10-2023

## Cataloging Electronic Theses and Dissertations: Updates and Perspectives from the Mississippi State University Libraries

Meg C. Wang

Mississippi State University, [mwang@library.msstate.edu](mailto:mwang@library.msstate.edu)

Follow this and additional works at: <https://digitalcommons.kennesaw.edu/seln>



Part of the [Information Literacy Commons](#), [Scholarly Communication Commons](#), and the [Scholarly Publishing Commons](#)

---

### Recommended Citation

Wang, Meg C. (2023) "Cataloging Electronic Theses and Dissertations: Updates and Perspectives from the Mississippi State University Libraries," *The Southeastern Librarian*: Vol. 71: Iss. 3, Article 9. Available at: <https://digitalcommons.kennesaw.edu/seln/vol71/iss3/9>

This General Interest is brought to you for free and open access by the Active Journals at DigitalCommons@Kennesaw State University. It has been accepted for inclusion in The Southeastern Librarian by an authorized editor of DigitalCommons@Kennesaw State University. For more information, please contact [digitalcommons@kennesaw.edu](mailto:digitalcommons@kennesaw.edu).

## **Cataloging Electronic Theses and Dissertations: Updates and Perspectives from the Mississippi State University Libraries**

Meg C. Wang<sup>a</sup> ([mwang@library.msstate.edu](mailto:mwang@library.msstate.edu))

<sup>a</sup>Mississippi State University, Starkville, Mississippi, USA

### **ABSTRACT**

Over the last two decades, electronic submission of theses and dissertations has become more common, and cataloging processes have evolved as well. At the same time, the Resource Description and Access (RDA) cataloging guidelines, which are better designed to describe digital resources, were widely implemented in 2013 in order to replace the Anglo-American Cataloging Rules (AACR2), which emphasize the description of a physical item. These changes have brought unceasing challenges into cataloging workflows. Especially with the development of linked data and the semantic web, catalogers consistently need to adapt measures to local conditions. This paper describes the recent issues that have occurred at the Mississippi State University Libraries and how local practices addressed these concerns to enhance bibliographic data and authority data for better discovery of Electronic Theses and Dissertations (ETDs).

### **KEYWORDS**

Cataloging, ETDs, ORCID

### **OVERVIEW**

More and more universities have turned their theses and dissertations into digital formats often accompanied with electronic submission systems to which students can upload files and provide their own bibliographic data. According to an online shared list accessed in February 2023, there are at least 123 academic institutions in North America that require Electronic Theses and Dissertations (ETDs) submissions (Texas ETD Association, n.d.). A survey of Texas colleges and universities showed progressive growth of ETDs between 2009 and 2020, with the percentage of institutions not considering/accepting ETDs dropping significantly from 20% to zero (Texas ETD Association, 2021).

Theses and dissertations are research outcomes for graduate students as well as important assets for universities. These works often represent cutting-edge or pioneering research in the academic disciplines. To achieve wider dissemination and discovery of such valuable scholarship, the notion of ETDs was discussed in 1987 in a meeting with participants from the University of Michigan, Virginia Tech, and two software companies, SoftQuad and Arbor Text. By 1996, Virginia Tech, with the support of the Southeastern Universities Research Association (SURA), developed the software ETD-db as a complete ETD submission package (Networked Digital Library of Theses and Dissertations, 2022). ETD-db is a MySQL open-source database system programmed by a series of Perl scripts that uses the Common Gateway Interface to deliver an ETD, serving as both a submission portal and a repository (Atkins, 2004). Through this electronic submission software, authors and librarians can easily create and manage an ETD using any computer with an internet connection and a web browser.

Mississippi State University (MSU) started using the ETD-db system in 2007. However, ETD-db did not have a feature for communication between students and librarians or the capability for batch processing or pulling cataloging information. With support from the Graduate School and campus administration, the Office of Thesis and Dissertation Format Review at the University Libraries began to look for a new ETD submission process in 2019. The online learning management system, Canvas, which is equipped with email and notification functions, stood out in comparison to other document management systems, such as Vireo and eForms or in-house software. Accordingly, a new course (LIB9010) offered by the Office of Thesis and Dissertation Format Review is integrated into the University's Canvas

courses system. All graduates completing a thesis or dissertation must enroll in this free online course through the University's Office of the Registrar and complete it before receiving their degrees.

Before RDA was implemented, Hoover and Wolverton gave an overview of thesis cataloging using AACR2 based on a survey of 171 institutions of higher education in the United States (Hoover & Wolverton, 2003). As the cataloging rules gradually shifted from AACR2 to RDA, the adoption of RDA into thesis cataloging and ETDs submission have been concurrent challenges for catalogers. Ashman (2013) gave a detailed discussion on MARC fields for cataloging ETDs using RDA. In terms of time- and labor-saving, two main ways of creating bibliographic records for ETDs are harvesting author-supplied metadata automatically and using OCLC constant data records (McCutcheon, 2011; Robinson et al., 2016).

At MSU, the OCLC constant data record method is employed for cataloging theses and dissertations. Catalogers created a template for cataloging ETDs using RDA and saved it as a constant data record in the online cataloging tool, OCLC Connexion. Separated from ETDs, printed theses and dissertations and architecture bachelor's theses have their own templates for cataloging. Constant data auto-fills certain fixed information from the designated template when creating a new bibliographic record, without entering repeatedly the same information in each new ETD record, such as 33x fields, the university name in the 264 and 710 fields, or academic theses as genre/form in the 655 field. This function helps to reduce typos and reduces processing time. The master template needs to be evaluated on a regular basis to verify that the constant data is still accurate. For instance, dates in fixed field and 090, 264, 502 fields will vary when working on ETDs published in different years, and the prefix URL in the 856 field would vary according to the type of permanent link used.

## RECENT CHALLENGES IN CATALOGING ETDs

### URL Changes for Better User Experience

During 2021 and 2022, data migration at MSU Libraires caused several ETD access issues, which are discussed in the following four phases.

#### ***Phase 1: The Combination of Redirect URLs and New URLs, December 2021***

Prior to 2020, the storage of MSU ETDs moved from the ETD-db system to the Handle server, becoming part of the MSU institutional repository via DSpace. However, the MSU Libraries experienced several unexpected Handle server crashes, leading to the decision in December 2020 to migrate data from Handle to a more robust and reliable system. Digital Commons, by bepress, was chosen, and both the current institutional repository and the ETD collection were migrated to this new cloud-hosted platform. The new repository was renamed Scholars Junction. During this data migration, the Library Web Services staff assisted with redirecting URLs for the existing 5,006 ETDs bibliographic records. URL redirects can forward searching traffic from one URL to another automatically when the old URL no longer functions. This kept catalogers from having to revise the URLs manually in each record, which totaled over 10,000 records between the Library Catalog and WorldCat.

Before the old Handle server was officially phased out in December 2021, there were 38 ETDs remaining in the to-be-cataloged list. Processing this group of ETDs required not only putting new Scholars Junction URL links (<https://scholarsjunction.msstate.edu/td/example>) in both backend cataloging systems, OCLC Connexion and Sirsi Workflow, but also retrieving author-supplied metadata and full-text PDFs from the old Handle server. This metadata information was used to create bibliographic and name authority files for the ETDs. This process was difficult as the Handle server went down continually.

#### ***Phase 2: DOIs Instead of URLs, May 2022***

Broken links are a common issue in the digital landscape, and these can impair the discovery of ETDs. DOIs were suggested as a solution to this issue. Unlike URLs, DOIs are static. Once they are assigned, they will not change, which means it is very easy to locate the document at any future time, even

after it has been moved. Beginning with the next group of ETDs to be cataloged, catalogers were able to retrieve author-supplied metadata and full-text PDFs from bepress. ETD URLs were replaced with DOIs assigned by the Scholarly Communication Librarian. Unlike some OCLC online bibliographic records showing DOIs in the 024 field as identifiers, cataloging librarians at MSU entered DOIs in the 856 field as electronic location and access.

### ***Phase 3: DOIs Go Optional, June-July 2022***

After the process of replacing URLs with DOIs in the ETDs' bibliographic records was in place, a large backlog of verifying DOIs of cited journal articles in the ETDs came up. As the position of Scholarly Communication Librarian was vacant at that time, this issue required the attention of the librarian at the Office of Thesis and Dissertation Format Review. The time it was taking to assign DOIs for ETDs became a concern and brought a halt to the use of DOIs as part of the cataloging workflow. This eventually led to the decision to have DOI creation as optional, only at the request of the authors of theses or dissertations.

Afterwards, cataloging librarians went back to using Scholars Junction URL links in the 856 field for ETDs that had not yet been cataloged, instead of waiting for DOIs to be assigned.

### ***Phase 4: URL forwarding failed, July-August 2022***

Once the URL redirection was implemented in phase one, the ETD records containing the old Handle URLs should have automatically forwarded the user to the new Scholars Junction links. However, when processing ETD bibliographic records with either Scholars Junction URLs or DOIs during the period of Phase 2 and Phase 3, the cataloger noticed there were some random dead links on the old Handle URLs which would lead the user to the 'record not found' webpage, a typical 404 error response from the server. To resolve the dead link issue, the cataloger manually deleted the old Handle URL and inserted a valid Scholars Junction URL on each broken link record in both cataloging systems: OCLC Connexion and Sirsi Workflow. In July 2022, the cataloger discovered that all of the old Handle URLs became dead links, resulting in poor user experience.

Replacement of the broken Handle URLs to the valid ones was considered for solving the issue, but this would involve going into over 10,000 records to manually remove broken Handle URLs and add valid Scholars Junction URLs. It might be a safe choice to update the URLs to fix the 404 errors, but that can be very time-consuming. While solidifying alternative solutions with the Web Services unit, the Systems staff managed to figure out the root cause for the broken link issue. The forwarding fails occurred when the Handle server had a bad shutdown because of a lightning strike which caused a power outage over the July 4<sup>th</sup> weekend. As this scenario is extremely rare, when Systems does a power recovery in the future, they will test the Handle server to prevent unnoticeable server issues from occurring. The broken link issue was resolved by simply restarting the Handle server.

### **Using ORCID for Better Author Identity Management**

Name authority control is another challenge for cataloging ETDs. After graduation, ETD authors often continue pursuing their careers in the same academic disciplines and continue to publish in scholarly resources in the future. Creating their name authority files is a way to provide an access point among various bibliographic resources published at different times. Since ETD authors are mostly in their early stage of academic life, it is relatively easy to evaluate whether any NACO authority file has been already established for a different entity.

In a name authority file, typically information about date/month/year of birth is provided in the 040 field, and the title of ETDs is provided in the 670 field. Along with the name of the author, this information is usually sufficient to distinguish between authors with the same names. Recently, the concept of identity management, such as adding an Open Researcher and Contributor ID (ORCID) identifier or a Wikidata identifier in the 024 field, has been discussed by Stalberg et. al. (2020) and the PCC URIs in the MARC Pilot NACO Subgroup (2022). ORCID is a unique identifier for academic authors. The ORCID platform is an easy way to document an author's research activities and makes a connection between

authors, publishers, employers, and funders. The downside is that authors may not maintain their own ORCID profile or may create more than one profile. ORCID's purpose of linking data for identifying authors does not always function correctly. In that case, using ORCID as an author identifier becomes meaningless and creates more confusion.

The MSU Libraries was hesitant about adding ORCID into bibliographic descriptions since ORCID is not required when students submit their ETDs. Students may be resistant to providing this additional information. The Office of Thesis and Dissertation Format Review has already received complaints about too many fields to fill in the ETD submission form. However, more and more academic journals are now requiring it when manuscripts are submitted. Through an outreach program in library instruction, such as a research guide (*ORCID Identifiers*, 2022, August 15), ORCID registration among faculty and students has been gradually accepted. After the MSU institutional repository switched to Scholars Junction, ORCID has become one of the optional fields for students to enter in the bepress platform. As a result, author-supplied metadata for cataloging ETDs started to show ORCID even though it is not mandatory.

Following the PCC URIs in the MARC Pilot NACO Subgroup release of the revised *NACO 024 Best Practices Guidelines* in May 2022, our local practice is to add ORCID identifiers in the 024 field (first indicator =8) with no subfield \$2 in the name authority file, not in the bibliographic record. The clickable link in subfield \$1 helps the cataloger to check the validity of the author's ORCID link. Causes for broken ORCID links vary: the name of ORCID is misspelled as "orchid", the first part of URL is duplicated (<https://orcid.org/orcid.org/example>), the identifier contains more than 16 characters, or it does not have dashes to delimit the identifier into 4-digit groups.

### **Interconnection Between Author-Supplied Metadata and Cataloger-Created Metadata**

Author-supplied metadata for cataloging ETDs comes from two sources: one is ETDs in full-text PDF files and the other is the University's ETDs submission form. Because the Office of Thesis and Dissertation Format Review helps our graduate students with getting their theses and dissertations formatted correctly, bibliographic information included in the resulting PDF document is the most accurate source of information for cataloging. However, information inserted in the electronic submission form by students occasionally contains errors that might lead to inaccurate cataloger-created metadata. Two of the most common errors submitted are the student's department and document type. It is surprising that this information would be incorrect as one would assume that students would correctly indicate what department they are part of and if their submission is a thesis or dissertation.

One reason for students confusing the document type is that in other countries a "thesis" is a product of a PhD degree, and a "dissertation" is a product of a master's degree, which is the opposite of the usage of the terms at MSU, as well as generally in US universities. If a student were to select the wrong document type, then this would affect the accuracy of the cataloger-added data in the 655 field (second indicator =7) in local practice and the item level details in Sirsi.

When filling out the submission form, students are asked to choose their college and their department in two separate fields. There are times when catalogers see the name of the same college in both fields without the name of department selected, or the name of two different colleges might be shown without the name of department selected. Other times, the correct name of the college is chosen but the incorrect name of the department is provided. These types of errors can cause the catalogers to assign an inaccurate call number, as they are assigned based on the student's department. An incorrect call number can cause the cataloger to inaccurately assign the 5-character cutter number and add unneeded extra letters to the cutter to distinguish between authors with the same last name.

### **DISCUSSION AND CONCLUSION**

A broken link in the 856 field can cause a serious access issue for born-digital materials, such as ETDs, which lead to poor user experience. URL forwarding and manual updates to valid URL links are both implemented to resolve broken link issues at the MSU Libraries. Furthermore, reaching out and

communicating with different library departments, such as Web Services and Systems, can often tackle the root causes in a more efficient way.

While a URL identifies the web address or location of a unique source, ORCID is another type of uniform resource identifier (URI). ORCID is a string of characters that generally identifies any web resource by using a name. A Wikidata identifier is also a type of URI frequently used for cataloging. Implementing ORCID for academic authors and URIs for catalog librarians are new trends that are being explored. ORCID identifiers can link a person's research, and URIs in NACO enhance discovery and identification for all users. Since most ETD authors are young and junior researchers, their ORCID profiles may be blank, but at times the profiles contain information that connects all of the author's research activities. For both authors and catalogers, adding the ORCID in the 024 field helps to incorporate a name authority file into the identity management environment.

Author-supplied metadata is important for cataloging ETDs. Once students submit their ETDs, they prefer to have their ETDs published in the MSU institutional repository as soon as possible so they can have a link to their ETDs on CVs for job applications. To-be-cataloged ETDs are given to catalogers 3 times a year and tend to already have been published for access. If author-supplied metadata appears to be incorrect, the catalogers have been given administrative permission to make the necessary corrections. While catalogers are correcting information, the incorrect data has been accessible for potentially 4 months. Inaccurate bibliographic data associated with academic research outputs and publications may have a negative effect on both the author and university. To avert this, a better checking mechanism needs to be in place during the ETD submission process that does not rely on corrections during the cataloging process. Using OCLC constant data records to catalog ETDs at the MSU Libraries has been an effective working model despite all the above-mentioned challenges.

OAI-PMH, a working model of harvesting metadata descriptions of records, is under investigation even though there is no concern of backlog cataloging for ETDs. Catalogers are facing more and more tasks of working on non-MARC metadata nowadays. Apart from harvesting ETD metadata, based on the Dublin-Core scheme in the bepress system, the OAI-PMH method can also be utilized to harvest finding aids metadata, based on the Encoded Archival Description (EAD) scheme in the ArchiveSpace platform. Once the programming tasks in the MarcEdit software are established, the OAI-PMH working model can be applied to batch process ETDs and finding aids cataloging in a more efficient way than the constant data method.

MSU Libraries is exploring retrospectively digitizing printed theses and dissertations. If pursued, all related departments will need to collaborate to establish a workflow and lifecycle of the newly scanned ETDs. In the MSU Libraries, the hard copies and digital copies are cataloged on separate bibliographic records. The tangible theses and dissertations, often considered as manuscripts, live in Archives and Special Collections or in Circulating Collections while ETDs, special kinds of e-books, live virtually in Scholars Junction. Even without the standard publishing process, ETDs from state universities in the United States are considered state government publications (McCutcheon, 2015). Once our Archives and Special Collections digitization team starts scanning printed theses and dissertations, the workflow for submission and cataloging the newly scanned documents needs to be updated. Open communication between cataloging and Special Collections is a big part of this process of becoming successful.

## REFERENCES

- Ashman, A.B. (2013). A brief look at how RDA is being used to catalog electronic theses and dissertations. *Kentucky Libraries*, 77 (3), 16-23.
- Atkins, A. (2004). ETD-db. In E. A. Fox, S. Feizabadi, J. M. Moxley & C. R. Weisser (Eds.), *Electronic theses and dissertations: A sourcebook for educators, students, and librarians* (pp. 163-167). CRC Press. <https://doi.org/10.4324/9780203021279>
- Hoover, L., & Wolverson, R.E. (2003). Cataloging and treatment of theses, dissertations, and ETDs. *Technical Services Quarterly*, 20, 3-57. [https://doi.org/10.1300/J124v20n04\\_02](https://doi.org/10.1300/J124v20n04_02)
- McCutcheon, S. (2011). Basic, fuller, fullest: Treatment options for electronic theses and dissertations. *Library Collections, Acquisitions, & Technical Services*, 35, 64-68. <https://doi.org/10.1016/j.lcats.2011.03.019>

- McCutcheon, S. (2015, November 2). *Cataloging ETDs in RDA*. [Video]. Youtube. <https://www.youtube.com/watch?v=i4d-Aq3saRg>
- Networked Digital Library of Theses and Dissertations. (2022). *Mission, goals, and history*. <https://ndltd.org/mission-goals-and-history/>
- ORCID Identifiers. (2022, August 15). Mississippi State University Libraries. Retrieved February 9, 2023, from <https://guides.library.msstate.edu/orcid>
- PCC URIs in MARC Pilot NACO subgroup. (2022, May 5). *NACO 024 best practices guidelines*. <https://www.loc.gov/aba/pcc/naco/documents/NACO-024-Best-Practices.pdf>
- Robinson, K., Edmunds, J. & Mattes, S.C. (2016). Leveraging author-supplied metadata, OAI-PMH, and XSLT to catalog ETDs. *Library Resources & Technical Services*, 60 (3), 191-203. <https://doi.org/10.5860/lrts.60n3.191>
- Stalberg, E., Riemer, J., MacEwan, A., Liss, J.A., Ilik, V., Hearn, S., Godby, J., Frank, P., Durocher, M. & Billey, A. (2020). Exploring models for shared identity management at a global scale: The work of the PCC task group on identity management in NACO. *Cataloging & Classification Quarterly*, 58 (3-4), 428-437. <http://doi.org/10.1080/01639374.2019.1699880>
- Texas ETD Association. (n.d.). *Add your institution and maintain your ETD information on the shared List of Institutions Requiring Electronic Thesis/Dissertation Submissions*. <http://txetda.org/index.php/outreach-and-communication/>
- Texas ETD Association. (2021, January 13). *Results from 2020 TxETDA Survey - Growth of ETDs in Texas Colleges and Universities*. [Video]. Youtube. <https://www.youtube.com/watch?v=87vYvZBsFCQ>