CHAPTER 4

One Library's Workflow and Processes for Record Management

LESLIE A. ENGELSON

Keeping track of files of bibliographic records, and procedures and decisions related to those records, can sometimes feel like herding cats. Trying to keep them together and moving in the same direction can be frustrating. Developing a data management plan to facilitate this process provides several benefits for the library and makes managing the metadata for those resources less daunting. Krier and Strasser (2014) speak to these benefits when they state that:

> Data management plans save time for the researcher over the long term. Effort spent before data collection begins can be focused on the wider context of the project rather than the details of a specific task or item. This ensures that the decisions made about data organization, management, and preservation are beneficial to long-term goals. Less time is spent rearranging, renaming, searching for, or otherwise

> handling files and data sets if their organization and management are thought out well in advance.
>
> Plans prevent upheaval brought about by staffing changes. ...The fact that staff come and go has no effect on the accessibility or usability of that data. The knowledge about data management, organization, and archiving stays within the group. (19–20)

While Krier and Strasser are speaking specifically about research data, their arguments for these plans apply equally to metadata related to bibliographic or authority records. Without a plan on how to manage metadata, "...the varied data management practices that result from [an] ad hoc practice can create a lack of continuity and lead to missing or incomprehensible data when a research assistant [or staff member] leaves the project. Data is easier to retrieve and use, whoever produced it, when it is managed properly" (2014, 7).

Following the concepts supporting research data management (RDM), metadata management plans should include:

- **Sources** of metadata
- **Tools** used to process metadata
- **Organizing strategies** for metadata

In this chapter, I discuss these elements of the metadata management plan as related to records for resources acquired by the Murray State University (MSU) libraries. I cover where our record files come from and list additional record sources, the tools I use for managing metadata, principles for organizing files, and additional information helpful to managing the full metadata lifecycle. For each source of files, I record the metadata management plan in a document called the Procedure.

Sources of metadata

While metadata describing information resources can come in a variety of schemas (MARC, MODS, Dublin Core, XML, etc.), MSU uses only MARC records in our catalog. Most records are acquired through OCLC's WorldCat database. When we cannot find a record

in WorldCat, we create one that conforms to international cataloging standards and add it to the WorldCat database.

A second source of records is vendors. When records are acquired from vendors, the Procedure documents the URL or ftp address, login information, and contact information for the vendor representative. Two of our vendors provide records that require minimal processing. The other vendors' records vary in quality and completeness. Details about record quality are documented in the Procedure. Because of the variance in quality of vendor records, I determine on a vendor-by-vendor basis whether or not the records will be used or if each resource will be manually cataloged, and note it in the Procedure. Some factors informing that decision are how many records we are acquiring at one time, if we are acquiring them one-by-one or from a package, and if data points are provided that facilitate matching records in WorldCat using a batch search. The use cases below demonstrate these factors and the decisions I made regarding whether or not, and how, vendor records are used.

- Use case 1. The vendor provides WorldCat records for both print and electronic records. The records are high quality and require minimal processing. The vendor notifies us when new and changed records are available on a monthly basis. We download the records, import them into our library management system (LMS), and batch add our holdings to WorldCat.

- Use case 2. The vendor provides basic records for electronic resources including the URL for accessing the resource. These records are fairly accurate but do not include unique identifier numbers (OCNs) for WorldCat records. I download records several times a year from the vendor's website for new and deleted resources. Using batch searching capabilities in Connexion, OCLC's cataloging tool, I match WorldCat records and add holdings, process the records adding the URL, and then import them into our LMS. I batch delete deleted records from our LMS and capture the OCNs to batch delete those holdings from WorldCat.

While MSU does not use Z39.50, this is a very useful way of acquiring records for many libraries that cannot afford a WorldCat subscription. Z39.50 is an international communication protocol

that allows for searching library databases and retrieving records. Searching and exporting records is limited to one record at a time. Any cost is tied directly to staff time and internet usage, so it is essential that library databases selected for searching are vetted for their ability to provide the maximum number of records. A first choice would be a national library database, as they tend to make their catalogs available for searching via Z39.50 and have significant holdings. These include:

- Library of Congress (United States)
- Bibliothèque et Archives (Canada)
- Biblioteca Nacional de España (Spain)
- Deutsche Nationalbibliothek (Germany)
- Bibliothèque nationale de France

An additional choice for acquiring MARC records for theological and religious libraries is via library catalogs at institutions connected to religious studies and theology. Check with the libraries to determine if they have made their databases available for Z39.50 searching.

Some LMSs provide another way to access records for electronic resources. Utilizing a centralized knowledge base available to all libraries using a specific LMS, a library can activate records for resources, making them available for use. Availability and access are dependent upon the relationship of the LMS vendor with the resource vendor. The MSU libraries utilize the knowledge base associated with our LMS to activate records for electronic resources we either subscribe to or have available through a demand-driven acquisition plan, but not for purchased resources.

Another source of metadata for records is an institutional repository (IR) from which metadata can be harvested and converted to MARC records. Finally, we create brief records for resources that we do not lend. These records are for equipment that can be checked out such as cameras, laptops, and study room keys, and are either created manually or converted to MARC from a spreadsheet.

Tools

Website links to the resources listed below are included in Appendix 4A.

The basic tool necessary for working with MARC metadata is the LMS. These systems vary in capability and facility for editing bibliographic records. When a group of records is acquired, many of the changes that need to be made to one record need to be made to many records in the group. One tool that I use for batch editing thousands of records at a time is MarcEdit. In addition to bulk editing MARC records, it can convert other schemas and data in spreadsheets to MARC. This tool is very powerful, has incredible functionality, and is free. Its creator, Terry Reese, is continually updating the tool with new functionality. He has provided documentation and YouTube videos to help you learn how to use it.

Another tool recently added to my cataloging toolbox is ChatGPT for creating tables of contents notes. Any AI tool for text generation can likely be used for this purpose. The challenge is developing a set of prompts to get results that require only minimal manual adjustments. This tool has the potential to save a significant amount of time and minimize inputting errors.

Included in the Procedure for each vendor is a list of edits to make to their records. These are considered guidelines as each file should be assessed for necessary edits due to changes vendors make in their metadata processes. Past practices do not always carry forward from file to file as vendors respond to customer feedback and changes in cataloging standards.

In addition to batch editing tools, cataloging tools for description, classification, subject analysis, and configuration within the MARC schema are necessary. Many are available freely on the internet; some require a subscription fee. Resource Description & Access (RDA) guidelines are the most current international descriptive cataloging guidelines and are available through the RDA Toolkit for an annual subscription fee based on the number of concurrent users. The Toolkit is used in conjunction with the RDA Registry, a free resource that "contains linked data and Semantic Web representations of the entities, elements, and terminologies approved by the RDA Steering Committee" (American Library Association, Canadian Federation of Library Associations, and CILIP: Chartered Institute of Library and Information Professionals 2024). The concepts underlying the RDA

guidelines are challenging to understand, and there is a steep learning curve associated with using the Toolkit. To help with this, the Program for Cooperative Cataloging (PCC) developed training for using this resource. Tools for help with the MARC schema include MARC21 Format for Bibliographic Data and OCLC's Bibliographic Formats and Standards, both freely available.

The main subject heading thesaurus is *Library of Congress Subject Headings* (LCSH). These headings can be found in downloadable PDFs or in a searchable database. This database also includes authorized headings for names, titles, genre terms, and more. When using this database, it is important to understand that unauthorized headings are also indexed. Their presence in the database does not indicate that they are the authorized term. To inform your usage of the terms in this database, identify the thesaurus from which the term is obtained, understand the structure of narrower and broader terms as well as "see" and "see also" references, recognize the 1XX tag used, and read scope notes. Buttons on the main page and inside the database for help for a specific page are invaluable.

In addition to this authority database, the Library of Congress has also provided access to another database of searchable ontologies and controlled vocabularies called *id.loc.gov*, in which LCSH and name authorities, genre terms, relators, and RDA terminology can be searched. This service intends to provide uniform resource identifiers (URI) for use in linked data and it is freely available. PDFs of the Library of Congress' Subject Heading Manual are available for help in building subject heading strings through the use of subdivisions and pairing headings to best reflect the topic of the resource.

The two main classification systems used by libraries are the Library of Congress Classification System (LCC) and the Dewey Decimal System (DDC). The LCC is available in PDFs, searchable through *id.loc.gov* and The Cataloging Calculator, and can be accessed by subscription through Classification Web. The DDC is owned by OCLC and can be accessed through a subscription to the WebDewey database.

Finally, The Cataloging Calculator is a free resource that provides searching of LCSH, LCC, standard subdivisions, and several other cataloging resources. Additionally, geographic and LC Cutters can be easily calculated using this tool.

Organizing strategies

Once you have determined the source of your records, what editing is needed, and how the editing will happen, it is time to design a plan for organizing the record files. Whether downloading files to an individual's computer if they are the only ones who need access to them, or to a networked folder if more than one person needs to access them, it is crucial that the file naming convention be shared with, understood, and implemented by each person accessing them.

Because our library receives MARC records for a variety of projects, the initial organization of computer folders is by project. For projects that acquire MARC records from multiple vendors, I create folders for each vendor. Within each vendor or project folder, I include a Procedure document and any other documentation associated with the vendor such as vendor-specific access and editing requirements, licenses, and technical specifications.

I then create an additional folder for records. This provides a dedicated area in which I can organize the various record files as I work through the editing process. This folder contains a document for keeping notes to track the editing progress and problems encountered for follow-up. This document is extremely helpful for keeping me on track in the event of interruptions. I include chain of custody information such as the date and initials of who updated the file. Finally, if it is important to archive old files, I add an archive folder and move the files I wish to retain to that folder to make the records folder available for the next load of records.

Downloaded files will have a vendor-supplied name and, depending on how informative that name is, you may want to keep it. More likely, you'll want to determine a naming convention by considering who will access the files and what information needs to be conveyed by the name. Computer scientists have determined best practices for naming versions of files, and a quick internet search results in a slew of resources on file naming conventions. Some of the main considerations are:

- Use a unique and descriptive identifier. Keep the name short and easy to understand. Since files are sorted by the first few elements, start with more general components and move to more specific ones.

- Avoid special characters and spaces. The only non-alphabetic and non-numeric characters that should be used in file naming are hyphens (-) and underscores (_). Avoid using spaces. While some recommend not using capital letters for consistency, it can be helpful to use them when identifying the beginning of a new word.

- Numbering. Use leading zeros in numbering for the computer to sort sequentially. Additionally, when using dates, use the format YYYYMMDD.

- Version control. Whether it is a file of MARC records or a spreadsheet of data, always keep the original file unchanged. When saving each edited version, add the letter v and a number to the end of the file name, e.g. v01, v02, etc. Resist using the word "final" as the inevitable need to edit the final version might result in some odd naming contortions to distinguish the different "final" versions.

Finally, include in the Procedure information about the lifecycle of the files. The determination of how long to keep metadata files is often based on the project, the comfort level of those who work with the files, and computer storage space. Once MARC files are imported into our LMS, we delete most of them from the computer. Some files are kept until the next batch is run as a way of documenting when they were last processed. Some spreadsheets that document decisions are kept long-term in the event they need to be referenced later. Reasons for archiving metadata should be documented in the Procedure and the decision reviewed periodically.

Conclusion

Managing metadata can sometimes be confusing and chaotic with numerous files and versions of files, multiple vendors, a variety of workflows, and various procedures to keep track of and implement. However, by utilizing some documentation practices and organizing strategies, managing metadata can become much more manageable. By developing a metadata management plan for each source of files, documenting that plan, and assigning a meaningful name to the files, I have a tidier list of files, I know what needs to be done with them

and how often, and, most importantly, our users have timelier access to resources available from the MSU library.

References

Krier, Laura and Carly A. Strasser. 2014. *Data Management for Libraries.* Chicago: ALA TechSource.

American Library Association, Canadian Federation of Library Associations, and CILIP: Chartered Institute of Library and Information Professionals. 2024. “RDA Registry.” Accessed May 29, 2024. *https://www.rdaregistry.info/*.

Appendix 4A: Links to Resources and Tools

Cataloging tools:

Bibliographic Formats and Standards: *https://www.oclc.org/bibformats/en/home.html*

Classification Web: *https://classweb.org/*

id.loc.gov Linked Data Service: *https://id.loc.gov/*

Library of Congress Authorities: *https://authorities.loc.gov/*

Library of Congress Classification: *https://www.loc.gov/catdir/cpso/lcco/*

Library of Congress Subject Headings: *https://www.loc.gov/aba/publications/FreeLCSH/freelcsh.html#About*

MARC Format for Bibliographic Data: *https://www.loc.gov/marc/bibliographic/*

RDA Registry: *https://www.rdaregistry.info/*

RDA Toolkit: *https://www.rdatoolkit.org/*

Subject Heading Manual: *https://www.loc.gov/aba/publications/FreeSHM/freeshm.html*

The Cataloging Calculator: *https://calculate.banerjee.site/*

WebDewey: *https://dewey.org/webdewey/login/login.html*

Editing tools:

ChatGPT: *https://chatgpt.com/*

MarcEdit: *https://marcedit.reeset.net/*

Tutorials:

MarcEdit tutorials: *https://www.youtube.com/@tpreese*

PCC Introductory RDA Training: *https://www.loc.gov/catworkshop/rda/index.html*