Metadata Management
Keeping Your Cataloging House in Order

by Leslie A. Engelson

It is very gratifying to bring order to a chaotic space. At home my fabric collection is sorted by color and theme and folded into same-sized bundles, my closet is sorted by type of garment and then color, even the containers in my kitchen cupboard are stacked by shape then size with the lids sorted so that I can easily match container with lid. Unfortunately, those spaces don't stay orderly when new items are added, old items are removed, or when items are used. The same can be said of our databases. Day-to-day activities, changes in standards, and major projects impact the consistency, accuracy, and quality of our metadata. Metadata management is never done; ongoing tidying and cleaning are required to maintain order.

A key element to keeping the database tidy is identifying and minimizing the factors that contribute to messiness. Not all factors that impact metadata can or should be eliminated. Just like I'm not going to stop eating so that I don't have to wash dishes, libraries are not going to stop implementing new standards, innovating, and developing just so we don't have to clean up data. However, measures can be put in place that can eliminate some messes, reduce the possibility of others, and make cleaning up the messes that do happen a little easier.

This article will discuss factors that contribute to messy data, identify high-value metadata elements in order to target cleanup efforts, and highlight tools that can be used to facilitate major database clean-up projects as well as the day-to-day chores of database maintenance and integrity.

A quality catalog starts with quality people who create that catalog. Too often short-sighted administrators think eliminating cataloging staff positions or using lower-paid staff in cataloging departments will minimize the negative impact of budget cuts on library users. This is far from true. Just because cataloging work is not as obvious as direct interaction with the public doesn't mean it is any less impactful. Cataloging work requires high-level critical thinking and decision-making. Salaries for these positions need to be high enough to attract and retain staff who have the appropriate level of analytical and evaluative ability.

Based on my experience, the biggest factor that contributes to messy metadata is poorly trained (though well-intentioned) staff. I cannot emphasize enough how important it is to have well-trained staff who know what the standards are, are aware of changes to standards, and know how to interpret and apply those standards. Funding for ongoing training and professional development for these staff needs to be included in the budget. Additionally, time in cataloging staff's schedules should be allocated in order for them to take advantage of freely available professional development opportunities such as reading listservs and participating in online training and webinars.

Vital in any metadata environment is consistency and accuracy. Therefore, it is essential that decisions on how cataloging standards are applied in your environment are well documented in order to inform future decisions and ensure consistency. Without ongoing training and clear documentation, major metadata integrity issues can arise, negatively impacting the ability to run accurate reports and find resources and rendering the most important and expensive tool used to support discovery unreliable.

Changing standards, such as new cataloging rules or the recent decision by the PCC to eliminate terminal punctuation, contribute to inconsistencies in the database. While records produced under

Leslie A. Engelson is the Metadata Librarian at Murray State University.

different standards are not incorrect, they do impact the search results, reporting, and database performance.

Essential to database integrity is clear communication with the cataloging department. When Research and Instruction Librarians decide to weed something or move it to a different collection, it is essential to communicate that decision to cataloging staff so the change can be reflected in the catalog. Better yet, including cataloging staff in meetings where projects that impact their work are discussed provides opportunity for those decisions to be informed by their impact on the catalog and cataloging staff as well as users.

Vendor records vary in quality and, while I have seen some improvement in vendor records over the years, vigilance about the quality of these records is essential. Additionally, vendors often contribute records to WorldCat with screen-scraped contents and summary notes. These can be particularly problematic as they don't include appropriate punctuation to show hierarchy and they are often incomplete or even pointless, such as those that have only the letters of the alphabet or chapter numbers. I have seen many summary notes that break off mid-sentence or even mid-word. Additionally, I have concerns about summary notes that are biased and serve only to sell the resource, as that runs counter to cataloger's efforts to eliminate, as much as possible, our own biases when we're cataloging.

Of course, limited time and staff impact our ability to ensure that the records that go into the database are high-quality. While batch loading does help with getting records into the database quickly, because of staffing limits it can also be a quick way to upload errors as well as records that fall short of the quality standards we would like.

As if maintaining a quality database wasn't enough, it's also important, when considering metadata management, to keep an eye on the future. Cataloging managers should consider how the metadata might be used in the future when making decisions about how they manage metadata now and on which elements to focus limited time and attention.

The question about what constitutes quality cataloging has been under discussion for some time. Two articles that are useful for understanding which elements to consider when determining cataloging quality as well as how to think about quality cataloging both for today as well as the future are:

Snow, K. 2017. "Defining, Assessing, and Rethinking Quality Cataloging." *Cataloging & Classification Quarterly* 55, nos. 7–8: 438–455.

Schultz-Jones, B., K. Snow, S. Miksa and R.L. Hasenyager, Jr. 2012. "Historical and Current Implications of Cataloguing Quality for Next-generation Catalogues." *Library Trends* 61, no. 1: 49–82.

These articles can help inform decisions about cataloging and metadata production both now and in consideration of moving to a linked-data environment. Determining high-value elements of data will be different for every library as we all serve unique communities and the most important element that determines quality cataloging is how it reflects the needs of our community. According to the *Statement of International Cataloguing Principles*, "convenience of the user" is the highest principle of cataloging.

Having said that, it is important to keep in mind how next-generation catalogs and discovery layers use the data. With the use of icons based on coding, facets to help narrow or focus searches,

relevancy rankings, de-duplication algorithms, and linked data-connecting resources, consider as important data integrity issues:

- Complete and correct coding (fixed fields, tags, and subfields)

- Complete and correct summaries and contents notes

- Correct authorized access points

Now that we have discussed the factors that impact cataloging quality and high-value metadata elements, let's turn our attention to tools that can assist our efforts to both maintain metadata integrity and cleanup inconsistent and messy data. The following is a list of commonly used tools, but it is not exhaustive and more tools are being developed all the time.

- MarcEdit – use for cleaning up records, verifying the MARC structure, validating access points, providing URIs for linked data, RDA processing

- OpenRefine – use for cleaning up metadata

- Macros – use for adding and deleting metadata; assists with consistency and efficiency

- Notepad++ and Regular Expressions – use for cleaning up metadata, facilitating authority work

- What Unicode Character is This? – use for identifying problem characters

- Batch processing in ILS – use for cleaning up records, deleting and importing records, updating access points

- Vendors – use for RDA processing, authority processing, cataloging of foreign language or unfamiliar formats; BackStage Library Works, Marcive

- Professional development – an absolute imperative!

- Student workers

- Help from experts – listservs, AskQC, webinars, etc.

Database integrity is essential for providing a useful tool for our users. Now, more than ever, as we turn our attention toward a linked data environment, I am starting to see people outside the cataloging community understand the importance of viewing cataloging not just as a task to acquire records but as a process of developing and curating a database that provides a reliable search experience for our users, quickly and efficiently connecting them to the resources they need. While staffing levels are still a challenge, more and more tools that can assist us in this effort are being developed and made available. Although the cleanup work will never be done, it can be managed with high-quality staff and targeted efforts.