Metadata

Acquiring Metadata for Your Library Resources
What to look (and look out) for

By Eric Parker and Jacqueline Magagnosc

With law libraries' increasing purchases of, and reliance on, electronic information resources, there is a growing need to provide users with as much high-quality metadata as possible to enable resource discovery. Gone are the days when we could just point users to a particular set of shelves and let them browse for information pertinent to their needs. These days, our holy grail is to make all of the resources we provide easily discoverable through one, hopefully simple, search interface.

As fewer staff resources are available to keep up with the growing need for metadata, libraries must rely more and more on metadata supplied by outside sources. This article aims to provide the reader with: (1) an overview of the different sources for such metadata; (2) the unique issues to be aware of when reviewing a contract for metadata services; (3) specific points to consider when determining data quality; and (4) issues related to the mechanics of making the metadata available in a catalog or other discovery platform. While some of the details, particularly those related to data quality, can quickly become quite complex, especially for non-catalogers, the target audience for this article is the

law librarian who may not have a high level of expertise in cataloging and metadata but who wishes to provide his or her users with the optimum level of resource discoverability. For the purposes of this article, we are limiting our discussion to MARC records, though we use the terms "metadata" and "cataloging" interchangeably.

Sources of Metadata

Currently, the marketplace for MARC records for legal and law-related resources is a rather complicated one, with records being made available through a number of different approaches.

Sets of MARC records are sometimes included in the purchase price of an electronic resource. Some recent examples of this practice include Oxford University Press resources, such as *Oxford Legal Research Library* and *Oxford Scholarly Authorities on International Law.* The March 2014 Report of the AALL Technical Services Special Interest Section's (TS-SIS's) Task Force on Vendor-Supplied Bibliographic Records Creation and Distribution Models (tinyurl.com/pa9g5be) recommends that vendors build the expense of MARC record creation into their costs for

overhead and development, since it is in their long-term interest to provide records to increase use of products. As of this writing, it remains to be seen whether this approach will become a best practice among legal and other information providers.

In other instances, vendors provide MARC records for individual titles contained in their products, but at an additional charge beyond the price of the resource content itself. Examples of this practice include MARC records for titles in the various components of Gale Cengage Learning's *Making of Modern Law.*

In some cases, cataloging is provided by third parties and must be purchased separately from the resource itself. Examples of this approach include Cassidy Cataloguing's MARC record sets for individual titles contained within the various libraries in HeinOnline, for the titles in Wolters Kluwer's Intelliconnect service, and for individual titles contained within WestlawNext and Lexis Advance.

Finally, there is the OCLC WorldShare Metadata Collection Manager service (formerly WorldCat Collection Sets). Through the WorldCat knowledge base, the service provides metadata for more than 11,000 content

collections from nearly 6,000 providers. Once a library has set up its profile, the service will allow the library to keep its data current via regular updates. Although it is not particularly heavy on law titles, there are a few law-specific collections, including records for titles in CALI, LLMC Digital, and some military legal resources. It is possible to download a spreadsheet, updated monthly, of the collections contained in the WorldCat knowledge base, **www.oclc.org/support/ services/collection-manager/release-notes.en.html**. OCLC plans to add user-defined collections to this service in the relatively near future.

Issues to Be Aware of in Contracts

Depending on the source for your records, you may or may not have a separate contract to negotiate for your metadata. But no matter the situation, it is helpful to clarify as many issues with your cataloging/metadata provider as you can. Many issues will be familiar to you if you have experience negotiating or reviewing license agreements for electronic resources. For example, contracts for catalog records or metadata need to spell out such issues as who the contracting parties are, what is to be provided, the term length of the contract, terms of payment, termination of contract, and severability.

There are some aspects to cataloging/ metadata contracts that are unique:

- The contract may specify how records are to be provided (e.g., via FTP, web-based download interface, or email). Be sure that you can easily retrieve the records using whatever approach the vendor uses to provide them.
- The contract should contain language specifying the level of quality and the standards that are to be applied to the records. If you already have a catalog or discovery interface consisting of high-quality records, you will want your new records to integrate well with those you already have for seamless discovery of resources. You will want to address these issues before the contract is signed, while you are in the driver's seat!
- Will the records conform to MARC21, RDA, and/or other nationally or internationally recognized standard(s)?
- Will records be provider-neutral? For those unfamiliar with provider-neutral records, the basic idea is that a record will describe the resource without referring to a specific provider or platform for an e-version

Figure 1: Detail from a provider-neutral catalog record as displayed in an OPAC.

in the descriptive portions of the record. Using a provider-neutral record, the same resource that may be available via multiple providers can have one record in your catalog, with links to the different versions, rather than having multiple separate records, each for a different version of the same thing. Issues associated with management of provider-neutral records will be discussed further under manipulation and loading of records. See Figure 1 for an example.

- Will the records use controlled access points and authorized forms, e.g., from the Library of Congress Name Authority File, for all names?
- Will the records use, and use correctly, standard subject schemes such as Library of Congress Subject Headings?

Because of its importance, we discuss data quality in greater detail below:

- What can the library do with the records it is purchasing? It may seem obvious to you that you are going to use the records in your catalog to allow people to find and access resources, but it is helpful to have spelled out whether they can be used in a discovery layer or uploaded to a utility such as OCLC, and, if not, what happens if they are, even accidentally. On a related note, unlike license agreements for e-resources, contracts for records do not tend to define authorized users since the records will go into an OPAC or other discovery layer where they may be retrieved and viewed by just about anyone.

- How much customization, if any, will the vendor provide? One useful piece of customization to specify is having the vendor customize URLs to include your proxy server prefix (if you have one) or use of customized text to appear as linked text in your catalog or discovery layer. If the vendor cannot provide customizations for you, you will likely have to do some processing on your end prior to loading the records (see below for additional details).

- Contracts for catalog records or metadata can be for a one-time purchase, but they can also include updates to records for a specified period of time. The updates can include records for new titles added to a database or revisions to existing records. They can also include notifications about records to remove from your database. If you are contracting for updates, it will be important to have the language specify the length of the term (often one year) and how frequently updates will be provided. Some databases that experience significant and frequent turnover of their content benefit from monthly MARC record updates.

As with many electronic resource purchases, it is advisable to ask your provider for a small sample set of records that exemplify those they will provide you. You can analyze the sample records to see whether they will be easy to load, will meet your quality standards, and will comply with contractual specifications. You can also determine what changes

may be needed before they can be added to your library's catalog.

Data Quality

Quality of vendor records is a long-term issue of concern, and the problematic issues have remained consistent over time. Much of the library literature concerning quality of vendor-supplied cataloging focuses on issues associated with specific collections.

What defines quality cataloging? Quality cataloging consists of adequate, accurate, and consistent metadata. Additionally, records should adhere to the specifications of the purchasing library as outlined in the purchase contract. As mentioned earlier, in order to prevent problems in the catalog and discovery layer, it is important that vendor records be consistent with standard cataloging protocols such as AACR2/RDA, MARC21, NACO (Name Authority Cooperative Program, to establish and maintain standardized forms of personal, corporate, and conference names) and SACO (Subject Authority Cooperative Program, to establish and maintain standardized subject headings for the Library of Congress Subject Headings). If a vendor's records are out of sync with cataloging standards, they will not behave in the same way as standards-compliant records when they are loaded into a local system. These kinds of issues can lead to a "split file" with works not correctly collocated by author or subject, making research more difficult and time-consuming.

The types of errors commonly found in vendor records fall into a number of categories: lack of completeness, noncompliance with standards, and inaccuracy of the records themselves. Errors in these areas affect resource discoverability and access; incomplete or inaccurate representation of collections undermines patron confidence in library services and generates frustration.

The first potential problem area is completeness of the record set. The number of records provided may not equal the number of titles in the collection. Titles may be missing, or titles not associated with the purchased collection may be included. Sometimes a record will be included for each volume of a multivolume set: for example, individual records for state code volumes instead of one record for the entire state code as is typically the case for the print. If you do not want your library's catalog and discovery layer to present results like this, it is best to pursue correction of this kind of problem before any more time is devoted to the record set.

Record metadata may not be compliant with national standards.

Publishers are not catalogers and they may not understand the significance of correct application of MARC21 or the importance of using controlled NACO/SACO-compliant headings, e.g., why the form used for an author's name or a subject heading should be the one given in the Name Authority File (NAF) or LCSH subject authority file, respectively. Some vendor record sets contain brief, generalized subject metadata which will also hamper patron discovery. If records come from OCLC, they are more likely to be within expected cataloging norms, but if they are derived from a publisher's internal metadata, they may differ from those norms significantly.

Vendor records may contain many different kinds of errors; those that impede patron access to materials are show stoppers. An example of this kind of error would be a malformed URL that does not connect to the resource described. A record for an online resource may be derived from the record for the print version without thorough adjustment of format-specific metadata. The fixed fields (see Figure 2) in the bibliographic record may not be coded correctly, leading to other potential problems since many integrated library systems and discovery layers draw from this data to differentiate between print and online materials or monographic and serial works.

Are accurate and consistent numerical identifiers included in the records? Identifiers, such as OCLC numbers, ISBNs, or ISSNs, are useful in collection maintenance, especially for nonstatic collections where content changes over time, necessitating addition, deletion, or updating of records. Conversely, incorrect use of identifiers, such as using the print ISBN for an e-book, can make management of a record set more difficult.

Recently, one of us dealt with a group of record samples that provided examples of many of these problems. One record option consisted of free records based on publisher metadata. The sample title was a state code, and each volume of the code had a separate record. The access points (forms and choices of personal and corporate names) in the records were completely different from those present in the record for the print version of the publication, and there was no link to the record for the print; subject metadata was brief and incorrectly coded. These records would provide patron access to a known title, so they are better than nothing, but it's unlikely that they would be included among subject-based search results.

Two price points were presented for fuller MARC records. Both options provided records derived from records describing a print publication, but each option had errors in either the fixed fields or variable fields of the bibliographic data. The less expensive option sample record had fixed field coding suitable for serials in a record describing a monograph; the more expensive option provided linking fields to records for the print publications—a desirable feature—but contained other

Treaties and laws of the Osage Nation, as passed to November 26, 1890...

000 01825cam a2200397Ia 450
001 4793004
005 20090312132746.0
006 m d
007 cr bn|300auaba
008 070914s1895 ksu s 001 0 eng d
035 __ |a (OCoLC)171296961
035 __ |a (OCoLC)ocn171296961
035 __ |a (IEN)LLMC-Digital D071213.B0035469.bin 00000012
037 __ |a 71-291 |b LLMC
040 __ |a LLMC2 |c LLMC2
043 __ |a n-us-ok
110 1_ |a Osage Tribe, Oklahoma.
240 10 |a Laws, etc. (Treaties and laws)
245 10 |a Treaties and laws of the Osage Nation, as passed to November 26, 1890 |h [electronic resource] / |c compiled by W.S. Fitzpatrick.
260 __ |a Cedar Vale, Kan. : |b Press of the Cedar Vale Commercial, |c 1895.
300 __ |a [22], 103 p. ; |c 24 cm.
500 __ |a Includes the Constitution of the Osage Nation.
500 __ |a Includes index.
530 __ |a Also issued in microfiche.
533 __ |a Electronic reproduction. |b [Washington, D.C. : |c Library of Congress, |d 2007?] |f (LLMC-Digital). |n Digitized image of the original. |n Subs... World Wide Web.
539 __ |a s |b 2007 |d dcu |e n |g s
500 __ |a Available from anywhere to all members of the Northwestern University community.
650 _0 |a Law |z Osage Tribe, Oklahoma.

Figure 2: Correctly coded fixed fields within a MARC record as displayed via OPAC; these often control displays of information to the public.

errors and omissions. Both record set options would require some correction and manipulation before they would be suitable for addition to a library catalog.

The Task Force on Vendor-Supplied Bibliographic Records Creation and Distribution Models has been reconstituted as the TS-SIS Vendor–Supplied Records Advisory Working Group (VRAG). This new group's mission is to work with vendors and libraries to improve the quality of record sets. Members of the group have evaluated and analyzed many law-centric record sets so that we do not have to. A full description of the work of this group and a spreadsheet presenting their analysis of different record sets can be found at **www.aallnet.org/sections/ ts/committees/Cataloging/Working- Groups/Vendor-Supplied-Records**.

Manipulation and Loading of Records

Analyzing a record set to determine the types of problems that may be present may seem daunting but can be managed using MarcEdit and Excel spreadsheets. MarcEdit, freely available at **marcedit.reeset.net**, is a program designed to facilitate editing and manipulation of MARC record metadata. The program parses MARC into a human-friendly format. Using this program and Excel, you can sort, count, and globally update fields. For example, using the MarcEdit "Count Fields" function, you can easily determine whether the number of records provided in a set equals the number of titles contained in the collection. For an overview of batch record analysis, see Yael Mandelstam's *Demystifying Batchload Analysis* PowerPoint presentation slides at **tsvbr.pbworks. com/f/batchload_analysis.ppt**.

You can also determine whether all required fields are present in the records by comparing field counts. For example, if the total number of subject heading fields is less than the number of records, some records must lack subject headings. One can check consistency of access points, such as author names, by exporting data to Excel, sorting, and then "eyeballing" the data. One can also add proxy server prefixes to URLs in record sets en masse if a vendor cannot provide this customization. Additionally, MarcEdit has tools for validating MARC records and verifying URLs.

Another cataloging choice to be considered is whether to present different versions of a work on one or multiple records. The RDA WEMI (Work, Expression, Manifestation, Item) construct requires separate records for

the print and electronic versions of a work since they are different expressions of that work. From the user side, one record presenting all access options seems best because it is less confusing. The Program for Cooperative Cataloging's *Provider-Neutral E-Resource MARC Record Guide* and *Provider- Neutral E-Monograph MARC Record Guide* document for publishers and catalogers the requirements of provider- neutral records. On the other hand, in the absence of consistent and reliable identifiers, separate records for each electronic version of a work are easier to manage for collections that change over time and so require addition and removal of records.

Management of records in batch is very different from the piece-by-piece cataloging normally used for print materials. Consistent record keeping and workflows are essential. Each record set will have its quirks, and these quirks may not be consistent between collections from the same vendor. You will want to keep a record of any problems and the metadata manipulations used to correct them. Some collections, such as *Making of Modern Law: Legal Treatises, 1600- 1926,* are static, but many reference collections have titles cycling in and out of them as they are kept up to date. It is essential to document identifiers and procedures used to keep records for these types of collections up to date.

For each record collection it is helpful to keep a record of:

- Data used to identify records for the set in your system, which can include both publisher-provided data and locally defined fields
- Location/source of records
- Updated schedule if the collection is not static
- Contractual limitations (OCLC, use in discovery layers)
- Data manipulations needed before loading records into your local system.

It is important to keep in mind that all vendor record sets are likely to require some editing before they are loaded into a local catalog. In an ideal world, with unlimited time, we could make all the checks above. In practice, depending on your particular situation, you may wish to focus on particular issues, such as identifying and correcting errors that most impede patron access. This category would include incorrect or inconsistent forms of personal and corporate names, typos in title fields, and URLs that do not work. If the problem fits a consistent pattern, you

can use batch update capabilities, either within MarcEdit or in your local system, to make corrections. Remember to get the provider of your cataloging to make corrections in the records that are out of compliance with your contract.

Conclusion

We have attempted to lay out some of the specific issues that libraries and librarians may face when purchasing, and then loading and using, catalog records or metadata that come from outside sources. While the issues can get complex, we believe that they can be addressed successfully by the information professional. Ultimately, our fellow librarians and our patrons benefit from improved accessibility of the resources we provide to them. ■

Eric Parker (**ecp278@law. northwestern.edu**), Associate Director for Collection and Bibliographic Services, Pritzker Legal Research Center at Northwestern University School of Law, Chicago

Jacqueline Magagnosc (**jkm95@cornell.edu**), Continuations Management Librarian, Cornell University Law Library, Ithaca, New York

Further Reading

Selected recent writings on vendor- provided cataloging data:

Roman S. Panchyshyn, "Asking the Right Questions: An E-Resource Checklist for Documenting Cataloging Decisions for Batch Cataloging Projects," *Technical Services Quarterly* 30 (2013): 15-37.

Stacy Traill, "Quality Issues in Vendor- Provided E-Monograph Records," *Library Resources & Technical Services* 57 (2013): 213-26.

Li Zhang and Miao Jin, "Cataloging E- Books: Dealing with Vendors and Various Other Problems," *The Serials Librarian: From the Printed Page to the Digital Age* 57 (2014): 76-80.