Cooperative e-book cataloging in the OhioLINK Library Consortium

Abstract

Since 2004, members of OhioLINK’s Database Management and Standards Committee have worked together to produce and distribute bibliographic records for over 44,000 electronic books. Using historical evidence, as well as the personal experience of key personnel, this paper examines the ways in which division of labor, cataloging standards, and procedures are negotiated within the consortium. Two case studies illustrate the ways in which cooperative e-book cataloging projects are created, developed, and adapted in response to changing circumstances. Challenges to current practices are discussed, and recommendations are offered to other libraries and consortia preparing to embark on cooperative cataloging projects.

Introduction

The Ohio Library and Information Network (OhioLINK) is "a consortium of 88 Ohio college and university libraries, and the State Library of Ohio, that work together to provide Ohio students, faculty and researchers with the information they need for teaching and research." OhioLINK’s major activities include hosting a consortial union catalog with patron-driven interlibrary loan functionality; providing a platform for members’ institutional repositories and local digitization projects; negotiating consortium discounts from e-content providers and library service vendors; purchasing or leasing e-content (such as electronic books, electronic journals, digital videos, and databases) for consortium-wide use; and
hosting purchased content on OhioLINK-administered Web sites (such as the Electronic Journal Center, Electronic Book Center, and Digital Media Center).

As of February 2011, OhioLINK has purchased nearly 50,000 e-books and leases over 10,000 additional e-books from a variety of vendors. Public service and collection development staff active in OhioLINK’s advisory committees express a strong desire for immediate, title-level access to e-books via the consortial union catalog and the catalogs of individual member institutions. OhioLINK’s unique e-content hosting situation (described more fully below) renders most vendor-supplied bibliographic records unusable in the out-of-the-box state. Therefore, OhioLINK has turned to cooperative cataloging, organized by the members of its Database Management & Standards Committee, as a means to provide catalog representation for e-books.

Electronic books have long been recognized as an increasingly significant component of academic library collections. Though access to e-books can be provided in a variety of ways, by 2004, a major survey found "general consensus that the OPAC should be the primary means of access to e-journals and e-books." By this time, practical guidance was available as to which MARC fields should be included in bibliographic records for e-books and what those fields should contain. However, e-books are frequently purchased in collections so large that individual attention to each record may be incompatible with timely completion of cataloging. This problem has frequently been addressed by applying some combination of vendor-supplied bibliographic data (MARC records or metadata spreadsheets), cataloging expertise (creation of cataloging standards or quality evaluation of vendor-supplied bibliographic records), and computer expertise (adapting existing technological tools or writing original programs to create or modify bibliographic data).

Authors have detailed procedures for creating, evaluating, improving, and incorporating e-resource records into their catalogs using tools such as Perl, MarcEdit, Microsoft Office software, and global update functions within the integrated library system. The labor
involved in designing and implementing such processes, as well as their frequent application to collections purchased by many libraries, would seem to make them strong candidates for cooperative efforts.

Authors using the phrase "cooperative cataloging" typically focus on national organizations such as the Program for Cooperative Cataloging, or on major bibliographic utilities such as OCLC. However, cooperation -- "activity shared for mutual benefit" is not restricted to activities on a national or international scale involving dozens or thousands of participants. Smaller groups of catalogers, frequently working within library consortia, have cooperated in such activities as establishing local cataloging standards, providing authority control for records used in a consortial environment, and requesting improvements to bibliographic records supplied by a vendor.

For purposes of this paper, "cooperative e-book cataloging" is defined as any effort in which

- catalogers from one or more institutions work together voluntarily,
- with some level of formal agreement as to assignment of tasks,
- to create or edit title-level bibliographic records for e-books,
- with finished records being freely distributed in batch-loadable form to all participating institutions.

This activity can be as limited as editing specific fields in a file of vendor records, or as extensive as full-level original cataloging; most projects involve some combination of traditional piecemeal cataloging activities and computer programming tasks such as writing scripts to edit records en masse. Rather than focusing on the technical procedures utilized for e-book cataloging at OhioLINK (which frequently parallel those described by the authors cited above), this paper seeks to explore the ways in which the labor of cooperative cataloging has been organized, negotiated and divided among project participants. OhioLINK is uniquely suited for such an examination: the sheer quantity of e-books
purchased and leased, combined with technical issues which render vendor records
unusable without significant modification, have necessitated ambitious cooperative
cataloging projects involving a variety of stakeholders.

Utilizing the historical meeting minutes of OhioLINK’s Database Management &
Standards Committee and other consortial advisory bodies, the archives of the Committee’s
electronic discussion list, personal communications with key personnel, and the author’s
own experience, this paper explores how OhioLINK cooperative e-book cataloging projects
are organized, and the ways in which that organization changes in response to factors such
as data availability, member needs, and changing cataloging standards. Two case studies
provide a more concrete picture of the manner in which OhioLINK cooperative cataloging
projects are initiated, negotiated, and stabilized over time. Difficulties and areas for
improvement are addressed, both organizational issues arising from within the consortium,
and external challenges presented by changing cataloging standards. Finally, lessons of the
OhioLINK experience are presented that can be utilized by other libraries and consortia
embarking on cooperative cataloging projects.

Background

OhioLINK’s consortial union catalog (known as the OhioLINK Library Catalog)
functions by accepting and merging bibliographic and item data from the local online public
access catalogs of the 89 member institutions, utilizing Innovative Interfaces’ Inn-Reach
software. Thus, in order for a consortially purchased e-book to be represented in the
OhioLINK Library Catalog, at least one member institution's OPAC must contain a
bibliographic record for that e-book. Since some librarians and patrons prefer to search their
local institutions’ OPACs before searching the OhioLINK Library Catalog, many member
institutions wish to have title-level bibliographic records for consortially purchased e-books
in their local OPACs as well. OhioLINK’s cooperative cataloging efforts have thus focused on
two goals: achieving title-level e-book representation in the OhioLINK Library Catalog, and
distributing batch-loadable copies of bibliographic records to interested member institutions.
The architecture of the OhioLINK Library Catalog necessitates a "separate records" approach to e-book cataloging.\textsuperscript{16} since the system architecture makes it difficult to predict which member institution’s version of a bibliographic record will appear in the consortial catalog, alterations to records for the print manifestations of books would have to be separately performed in every member institution’s local catalog, an impractical undertaking that would defeat the purpose of cooperative record creation.

OhioLINK services are maintained by a small consortial staff advised and supported by several working committees comprised of staff from member institutions. One such committee, the Database Management & Standards Committee (DMSC), describes itself on its Web site as a body that

Maintains quality standards for the [consortial union] Catalog. Creates metadata policies and procedures for all members to follow. Organizes cooperative projects to catalog consortial resources. Consists of cataloging librarians.\textsuperscript{17}

As of January 2011, DMSC has 25 official members (24 from member institutions plus one consortial staff liaison), but all technical services staff at member institutions are welcome to participate in its activities. The Committee holds bimonthly in-person meetings, which since 2008 have been live-streamed over the World Wide Web to interested member institution staff. Between meetings, the Committee communicates via an electronic discussion list which any staff member at any OhioLINK institution may join.

Discussion of e-resource cataloging can be found in DMSC meeting minutes as early as March 1996.\textsuperscript{18} Experiments with e-book cataloging are mentioned as early as June 1999, but these efforts were generally undertaken by single institutions, and there is no clear evidence that complete record sets were produced or distributed throughout the consortium in an organized manner. OhioLINK’s first major e-book purchase was a collection from NetLibrary in 1999. Title-level bibliographic records for this collection were purchased via OCLC’s PromptCat service rather than produced cooperatively, but this experience acquainted DMSC members with the potential pitfalls of purchased records, including
upfront financial costs, delivery time, and potential data quality issues. (It may also have helped solidify the belief within the consortium that e-books could and should be represented in the catalog in a manner basically similar to print monographs.)

In 2004, DMSC ratified the first version of its *Standards for Cataloging Electronic Monographs*. Originally intended as a standard for vendor records to be used by collection development staff negotiating e-book purchases, the *Standards* ultimately became much more important as a guide for DMSC members engaged in cooperative cataloging. The development of the *Standards* in 2003-2004 coincided with the earliest modern-style DMSC cooperative e-book cataloging efforts.

**Why not use vendor records?**

Many e-book vendors are able to supply customers with prefabricated, batch-loadable sets of bibliographic records. It might seem that this should obviate the need for cooperative production of bibliographic records. However, OhioLINK’s unique situation complicates its’ members ability to utilize vendor records. Vendor records generally contain the URLs used to access e-books on the vendor’s own Web site in an electronic location/access field (MARC tag 856), reflecting the typical agreement in which a vendor grants its customers access to its content on its own site. However, OhioLINK’s vendor agreements often do not conform to this model. Rather, a copy of the full text of e-books purchased (rather than leased) by OhioLINK is typically hosted on a separate, OhioLINK-run site known as the Electronic Book Center (http://ebooks.ohiolink.edu/). Purchase agreements often provide for this "local load" in lieu of access at the vendor’s site. Since an e-book’s URL in the Center is different from, and bears no relationship to, its URL on the vendor’s site, vendor records for titles housed in the Center require, at minimum, labor-intensive record-by-record replacement of the URL in MARC field 856. Since consortial public service staff expect bibliographic records for e-books to provide a direct link to the full text, vendor records are not considered usable without this customization.
Furthermore, concerns about bibliographic record accuracy, retrievability, and adherence to cataloging standards often inspire DMSC to go beyond this minimum necessary customization when working with vendor records. Although practicality prevents DMSC from insisting on perfect adherence to its *Standards for Cataloging Electronic Monographs* in all the records it accepts and distributes, DMSC strives to provide high-quality metadata. If vendor records lack important access points, fail to conform to recognized cataloging standards and conventions, or contain inaccurate data, DMSC may choose to enhance the records, or to abandon them and create higher-quality records based on data obtained elsewhere. DMSC’s complete record standard is available on the World Wide Web and in the interests of space will not be reproduced here. However, some specific violations of the standard encountered in past sets of vendor records can be enumerated:

- **Lack of Library of Congress Subject Headings**, considered by many members to be vital to the retrievability of non-fiction books
- **Lack of Medical Subject Headings**, highly desired in records for books of appropriate subject matter by member libraries that support medical schools
- **Author name headings not in authorized form**, subverting the high value many members place on authority control
- **ISBNs of original print manifestations** (which facilitate, among other things, the linking of bibliographic records in the consortial and member institution catalogs to a bibliographic enrichment service that provides reviews, author biographies, and additional content) not present anywhere in the record
- **Cataloging of serial issues as monographs**, leading to multiple near-identical records that can be confusing to public service staff and patrons
- **Prior to 2009, cataloging of electronic monograph reproductions** as if they were born-digital resources, with information about the original print manifestation relegated to a note field (a reversal of the then-standard application of Library of Congress Rule
Interpretation 1.11A to electronic reproductions). The reasons this was considered a significant problem are not explicitly recorded, but seem to have included a desire for consistency with other e-book records, good cataloging community citizenship through conformity to standards, and in at least one case (that of the Chadwyck-Healey collection discussed below) a belief that information about the print manifestation would be particularly salient to the type of researcher likely to use the material.

The organization of cooperative e-book cataloging at OhioLINK

When a cooperative cataloging project is undertaken, participants are organized by DMSC, and labor is generally either performed by or supervised by DMSC members at their local institutions. The "who" and "how" of each project is negotiated among DMSC members during bimonthly meetings. These negotiations often determine how the Standards for Cataloging Electronic Monographs will be applied to the specific project; for example, whether to devote specific effort to ensuring the presence of MARC fields listed as optional in the standards. Decisions are made informally based on member consensus or majority opinion, taking into account factors such as:

- whether vendor records are available, the quality of those records (in terms of features like presence of access points, conformity to standards, and accuracy of data); and terms and conditions for their use;
- whether records exist in OCLC WorldCat, either for the e-books themselves or for alternate manifestations such as print versions; if so, the quality of those records and the ease of retrieving and editing them;
- the nature of the material being cataloged (for example, Library of Congress Subject Headings may be considered more important to the retrieval of non-fiction than of literary titles); and
• the likely enduring value of the material being cataloged (for example, e-books that are leased by OhioLINK on a temporary basis might warrant less cataloging time and effort than e-books permanently purchased and archived).

The amount of time and labor required to obtain a desired level of record quality or enhancement must often be balanced against the level of demand for immediate title-level representation from public service and collection development staff.

If the project is large enough to require some form of batch processing, an initial cataloging procedure is designed by a DMSC member skilled in the application of technological tools. If vendor records are available, those records may be handled in a variety of ways. Records that do not require URL replacement (i.e., represent leased titles accessed via the vendor's Web site), have OCLC numbers, and are otherwise judged to be of sufficient quality may be distributed with little editing. If the records lack OCLC control numbers, DMSC strongly prefers to convert the records to OCLC records to facilitate interoperability with various Web sites and services (most notably the OhioLINK Library Catalog, which uses OCLC control numbers as its primary means of determining when to treat records contributed by different member libraries as equivalent). If vendor records are of generally high quality and terms of use permit, they may be uploaded to OCLC, edited as required, then contributed to OCLC WorldCat before being distributed to OhioLINK member libraries. If vendor records lack required access points, contain inaccurate data, or their formatting departs widely from cataloging standards, it may be less labor-intensive to abandon them entirely and find or create higher quality records in WorldCat.

[FIGURE 1: Possible pathways for cooperative creation and distribution of e-book records]

Once a procedure has been determined, volunteers are sought during DMSC meetings or, in cases where only a few members are known to have the required technological skills for a project, approached privately by the committee's consortial staff liaison. The division of labor for each project is heavily dependent on the chosen procedure. For example, if e-book titles are to be cataloged piecemeal in OCLC WorldCat, all DMSC
members possess the necessary skills to participate, and large collections can easily be broken into smaller groups of titles for division among multiple volunteers. If, on the other hand, collections are to be batch-processed using tools such as MarcEdit, computer scripting languages, or custom record load profiles in the local OPAC, the number of potential volunteers is limited by the lack of widespread familiarity with these tools, and the procedure may "scale" well enough that division of the collection would offer little benefit or even be detrimental. Other projects lend themselves to assembly line-style procedures: for example, time-consuming piecemeal OCLC searching of a large collection may be divided among several institutions, who then funnel their search results to a single technologically skilled volunteer for batch processing.

[FIGURE 2: Possible divisions of labor for cooperative cataloging projects]

Finished records are distributed to member institutions via OhioLINK- or member-maintained FTP servers, coordinated by DMSC’s OhioLINK staff liaison. For each project, a single institution -- often, but not always, the one that had the largest hand in creating the bibliographic records -- is assigned responsibility for loading records into its local OPAC and ensuring that the records are transmitted from the local OPAC to the OhioLINK Library Catalog. Other institutions are free to choose whether or not to load particular sets of records into their local catalogs and, if so, whether or not to transmit information about these "holdings" to the union catalog. Spot checking of e-book records in the union catalog shows that a typical cooperatively produced e-book record set is loaded into the local catalogs of at least 20-25 institutions (probably more, since anecdotal evidence suggests that some institutions load the records locally but do not transmit holdings information to the union catalog).

The following case studies serve to illustrate the wide variety of procedures and techniques utilized in DMSC’s cooperative e-book cataloging projects.

Case Study 1: Chadwyck-Healey literary e-books
In 2003, OhioLINK acquired a large collection of literary e-books from Chadwyck-Healey (a subsidiary of ProQuest). The collection consists of over 11,000 titles in 15 sub-collections, mostly electronic reproductions of books of poetry and drama originally published in print between the sixteenth and mid-twentieth centuries. The full text of the books, with descriptive metadata provided by the vendor, was loaded onto the OhioLINK Electronic Book Center server, and the Center Web site was to serve as the exclusive access point for OhioLINK patrons. Public service staff from OhioLINK member institutions expressed immediate interest in title-level cataloging: August 2004 DMSC meeting minutes state that the committee "has been asked several times to try to [provide] analytic records for these sets."

Bibliographic records were provided by ProQuest, but the records were not immediately usable as the linking URLs they contained led to the vendor’s site rather than to the Electronic Book Center. DMSC initially decided to disregard the vendor records and create original records based on existing OCLC records for the print manifestations. The reasons for this decision are not clearly recorded, but it is known that despite the fact that the vendor explicitly described the e-books as reproductions of print originals, the vendor records did not adhere to then-current standards for cataloging electronic monograph reproductions. Rather, the records treated the e-books as born-digital items, placing information about the e-book in the imprint (MARC 260) and physical description (300) fields while relegating information about the print original to a note field. This departure from standards may have led DMSC members to be skeptical of the overall quality of the vendor records, even without concrete evidence to support that skepticism. The author remembers concern being (verbally) expressed about the possibility that the vendor records might lack correct or complete information about the print originals, information which might be particularly salient to researchers utilizing one-hundred-plus-year-old books.
It was initially determined that descriptive metadata for one large sub-collection, "English Poetry," would be extracted from the Electronic Book Center by OhioLINK information technology staff and imported into Microsoft Excel spreadsheets. Available metadata included author names, titles, and notes containing imprint and physical description information for the original print manifestations; these notes would be crucial for identifying the correct print manifestation records, since most print manifestations were too old to have unique identifiers such as ISBNs. DMSC volunteers would use the available metadata to search OCLC WorldCat for records for the individual print manifestations, and add the records’ OCLC control numbers to the spreadsheets. It is not clear to what degree the committee worried about the possible inclusion of pre-AACR2 records in the search results, but the occasional instances of, for example, pre-ISBD punctuation that can be found in the finished record set suggests that some such records were accepted.

Another volunteer with high technology skills would then retrieve the records from OCLC and use batch processes to add MARC fields and subfields then standard in records for electronic monograph reproductions, such as the General Material Designation (245 subfield h), Additional Material Characteristics (006), Physical Description Fixed Field (007), Reproduction Note (533), and Electronic Description and Access (856). The edited records would be contributed to OCLC WorldCat as new records representing the e-book manifestations found in the Electronic Book Center. Calls for volunteers were made at DMSC meetings in late 2004 and early 2005, while DMSC’s OhioLINK staff liaison worked with OhioLINK information technology staff to extract metadata. In May 2005, spreadsheets were produced and distributed to several volunteer institutions via the DMSC discussion list, but the initial spreadsheets were corrupted, requiring an additional waiting period for a second, cleaner data extraction. After this hurdle had been cleared, instructions given to the volunteer searchers were clarified, and adapted to unusual cases, via discussion on the discussion list, and the project was able to proceed. However, due to the inherent slowness of title-by-title searching without unique identifiers, and perhaps the reliance on volunteers
to pace themselves (meeting minutes and discussion list archives show no evidence of a timeline), the completed records for the first sub-set of the English Poetry sub-collection were not released until November 2006. By this time, comparisons of the vendor records to the corresponding print manifestation records suggested that the note fields in the vendor records did contain sufficient information about the original print manifestations.

By April 2007, DMSC was ready to reconsider the idea of generating OhioLINK-usable records from the records supplied by ProQuest. By the June 2007 meeting, a technologically skilled volunteer had developed a batch procedure for moving information about the print manifestations from the note fields of the vendor records to the 260 and 300 fields, while adding an appropriate 533 field describing the reproduction. This procedure greatly reduced processing time, and OCLC staff declared the resulting records acceptable for contribution to WorldCat. However, delays in the release of finished record sets remained, likely in part because the technology skill requirements and lack of documentation of the new procedure effectively limited the number of volunteers to one. In December 2007, though most Chadwyck-Healey sub-collections remained uncataloged, a moratorium on Chadwyck-Healey cataloging was declared, in part due to the purchase of the Springer e-book collection (described in Example 2 below) to which at least some consortial stakeholders assigned a higher priority.

In spring 2009, after procedures for cataloging the Springer e-books had been developed and stabilized, OhioLINK’s reference-focused User Services Committee was asked to prioritize the cataloging of the remaining Chadwyck-Healey sub-collections. DMSC resumed work on the Chadwyck-Healey project, applying the lessons learned during earlier attempts. DMSC’s OhioLINK staff liaison once again called for volunteers, then -- rather than dividing a generic task equally among them, as had been done with OCLC searching under the initial procedure -- distributed sub-collections with close regard to available time and technological skill. Small sub-collections, which could quickly be cataloged piecemeal using skills common to all catalogers, were assigned to volunteer institutions with smaller
staffs, less available time, and/or lower technological skills. Higher-skilled volunteers could then concentrate on batch processing of sub-collections comprising hundreds or thousands of titles.

The developer of the method for transforming ProQuest vendor records into electronic monograph reproduction records shared information about his method with another volunteer (the author). She, in turn, developed a modified version of this method using the MarcEdit cataloging utility, Innovative Interfaces record load profiling, and OCLC Connexion Client cataloging software, and documented her method with a written procedure that could be followed by individuals possessing basic familiarity with these tools. She also provided technological support to other volunteers via email. The combination of more rational division of labor and more streamlined, well documented procedures allowed for faster progress in producing usable records. The final record sets were distributed in March 2010, less than a year after resumption of work on the project.

The Chadwyck-Healey project is one of the most ambitious so far attempted by DMSC cooperative cataloging volunteers, both in terms of the number of titles and the number of volunteer institutions involved. It also illustrates the process of testing, refinement, negotiation and compromise characteristic of DMSC’s cooperative cataloging projects. DMSC volunteers began by testing an elaborate procedure involving multiple institutions, with all final processing done by a single volunteer using technology skills not possessed by the other participants. When this procedure proved labor-intensive and its quality dividends less than expected, DMSC changed gears, finding ways to improve the initially unacceptable vendor records, and dividing the project in a manner that allowed institutions at all skill levels to participate.

Case Study 2: Springer-Verlag e-book backfile

In late 2007, OhioLINK negotiated the purchase of a 13,000+-title collection of ebooks from Springer-Verlag, plus an ongoing accumulation of newly published titles. The books were to be accessed by OhioLINK patrons via the Electronic Book Center. At first,
DMSC hesitated to initiate a cooperative cataloging project due to announcements that OCLC was already working with Springer to produce records. However, by spring 2008, both delay in the availability of these records, and potential problems with their use -- including the difficulty of isolating records for the exact portion of the collection included in OhioLINK’s purchase, and their lack of Electronic Book Center URLs -- led DMSC to conclude that cooperative cataloging was warranted.

Cataloging of newly added Springer titles was undertaken by a single, large volunteer institution willing to devote sufficient human resources to piecemeal-cataloging of these additions. However, due to its sheer size, the backlog of older titles could only be efficiently handled using batch processes. Vendor records were available, but they did not conform to DMSC standards and lacked Library of Congress subject headings, a feature DMSC considered important to the retrievability of this collection.

After negotiating record quality requirements and possible procedures, DMSC elected to generate records for the OhioLINK Electronic Book Center manifestations from existing OCLC records for the print manifestations. On the surface, the Chadwyck-Healey cataloging experience described above would seem to recommend against this procedure. However, the Springer backfile titles differed from the Chadwyck-Healey literary titles in important respects. First, their print manifestations had been published recently enough to have unique numeric identifiers (ISBNs). Second, being the output of a major scientific publisher, most of the print manifestations had been cataloged in OCLC WorldCat by highly trusted cataloging agencies such as the Library of Congress and Program for Cooperative Cataloging participants. The combination of unique identifiers and OCLC search qualifiers that limited results to records produced by trusted agencies offered the potential to quickly isolate high-quality print manifestation records via batch searching.

The Springer e-book backfile was cataloged using a multi-stage process. In preparation, metadata supplied by the vendor for the Electronic Book Center load (including author names, titles, publishers, publication dates, and some ISBNs) was extracted from the
Center into spreadsheets. Support staff and student workers at one volunteer institution augmented the spreadsheets manually, adding data such as certain print manifestation ISBNs which had not been included in the Center metadata. Once the spreadsheets were complete, another volunteer (the author) used computer scripts to transform the data in the spreadsheets into OCLC Connexion search keys incorporating ISBNs, cataloging agency preferences, and other data. She then ran batch searches and cleaned the search results, ensuring that one record per book was included and noting the occasional item that lacked an existing print manifestation record of the desired quality. Retrieved print manifestation records were exported into a MARC file which was passed to a third, highly technologically skilled volunteer, who used custom load profiles in his institution’s Innovative Interfaces OPAC to successively add, remove and rearrange data, transforming print manifestation records into electronic monograph records for the manifestations found in the Electronic Book Center. These records were then contributed to OCLC WorldCat.

The first record sets created by the above process were distributed in summer 2008. As of November 2010, the last few thousand titles in the Springer backfile have entered the final phase of the cataloging process. The Springer project could be completed more efficiently than earlier efforts such as the Chadwyck-Healey project because of the knowledge base -- of both technical procedures and project management strategies -- developed via trial and error during those earlier projects.

Organizational challenges

Evidence from DMSC’s electronic discussion list and meeting minutes suggests that DMSC’s cooperative e-book cataloging activities have stabilized since 2008. From 2004 through winter 2008/2009, cooperative e-book cataloging projects are the subject of debate and discussion. However, from spring 2009 through 2010, cooperative e-book cataloging is mentioned mostly in simple announcements of who has volunteered to take on specific projects, or of the availability of completed record sets. Probably this is partly due to certain members having become accustomed to volunteering repeatedly, thereby developing a
repertoire of standards and procedures that can be applied to new record sets with little need for modification.

However, there is still some variety among volunteers. As of October 2010, at least six different OhioLINK institutions have volunteered to the extent of bearing principal responsibility for the cataloging of one or more collections, and several additional institutions have participated (sometimes quite extensively) without being designated the lead institution for an entire collection. Notably, not all of the six volunteers responsible for entire collections are members of DMSC’s "Priority 1" group, the institutions whose bibliographic data is preferred for display in the union catalog due to their relatively strong subject analysis and authority control policies supported by relatively large and skilled cataloging staffs. Volunteers come from institutions with a wide range of cataloging staffing levels.

All of the above suggests that DMSC has succeeded in developing a workable system of cooperative e-book cataloging. However, certain challenges remain to be negotiated as DMSC continues its current cataloging projects and initiates new ones. Foremost is the issue of fairness in an all-volunteer system. Only a small minority of OhioLINK’s 89 member institutions participate in cooperative cataloging projects. Meanwhile all member institutions’ public service staff and patrons benefit from access to e-books via the OhioLINK Library Catalog, and at least a quarter of member institutions load cooperatively created records into their local catalogs.

This approach, in which the efforts of the few benefit the many, is not necessarily objectionable on the basis of "unfairness" alone. It is similar in spirit to the cooperative nature of bibliographic utilities such as OCLC WorldCat: members that contribute little original cataloging may be said to benefit disproportionately from membership in these utilities, but even high-contributing members tend to find the arrangement beneficial. Moreover, at least some DMSC volunteers (including the author) hail from institutions where
public service staff demand for cataloging of OhioLINK-purchased e-books is so high that they would likely be producing bibliographic records for their local catalogs in any case. Thus, it can be argued that they lose nothing by sharing those records with the consortium. Furthermore, it might be argued that, for example, community college libraries with one or no professional catalogers cannot reasonably be expected to contribute on an equal basis with Association of Research Libraries members employing several professional catalogers.

However, if ethical concerns can be dismissed, practical concerns remain. DMSC’s dependence on specific institutions and/or individuals becomes salient when their home institution workloads or other issues delay the completion of cooperative cataloging assignments. Large collections can be efficiently cataloged only by using sophisticated batch processes requiring technical skills which few DMSC members possess, leading records to bottleneck on the desks of a few highly skilled, and proportionately busy, individuals. The actual procedures utilized in a batch processing project are often understood only by the individuals who designed them, leaving DMSC with little idea of what specific work is being done or how long it takes, and therefore no basis on which to monitor and manage progress.

Perhaps the single most effective action DMSC could take to improve the effectiveness of the cooperative cataloging program would be to provide technology training to its members. Many cooperative cataloging projects can only be performed efficiently with the use of tools such as MarcEdit, OCLC Macro Language, custom data profiling in the local OPAC, or non-cataloging-specific computer scripting languages. Many potential cooperative cataloging volunteers are unfamiliar with these tools. Training these individuals could allow the burden of record creation to be spread more widely, as well as alleviating the fragility of projects dependent on single individuals, and allowing for succession planning. It would also provide the trainees with valuable skills that could be applied to cataloging work at their home institutions. Clear and complete documentation of procedures, made easily accessible
through a venue such as the DMSC Web site, would be necessary to ensure the success of the training program.

In addition to technology training, DMSC might consider introducing greater clarity and formality to the volunteering process. Having gained experience over the course of several projects, DMSC should now be able to estimate the amount of time required of volunteers, and should state those time requirements upfront. Expected project completion dates should be established and communicated, particularly in cases where collection development and/or public service staff have expressed strong desire to see a project completed. Volunteers -- and where necessary, their supervisors at their home institutions -- must be realistic about the commitments they make. Since DMSC has no real authority over its members -- many of whom juggle numerous job responsibilities -- time commitments and deadlines must remain flexible, and formal written agreements might not be considered necessary or appropriate. However, clear communication of expectations would help facilitate completion of cooperative cataloging projects.

Finally, more formal recognition of contributions to cooperative cataloging projects might increase incentive among member institutions to participate. Volunteers are acknowledged verbally at DMSC meetings and in writing via the emails informing committee members of the availability of completed record sets. However, more formal, public recognition, including details such as the names of volunteers, the number of hours spent on projects and number of records created, and shared with groups such as the OhioLINK Library Advisory Council (a body of library directors) -- might both reward volunteers, and promote consortium-wide recognition of their value of their efforts, and of cataloging in general. Presentation of data such as the cooperative cataloging participation level by different Priority 1 institutions, or the participation level of all institutions in relation to their Electronic Book Center usage statistics, might even inspire a positive spirit of competition to contribute among directors.

Evolving cataloging standards
Provider-neutral e-monograph records

In addition to the organizational challenges described above, ongoing changes in cataloging standards will require evolving responses from DMSC’s cooperative e-book cataloging program. Through 2009, original records contributed to OCLC WorldCat generally described e-books as electronic monograph reproductions, applying Library of Congress Rule Interpretation 1.11A in the manner common at the time. Information in fields such as the Restrictions on Access Note (MARC 506), Reproduction Note (MARC 533), added entries for provider names (MARC 710) and Electronic Location and Access (MARC 856) reflected the e-book manifestation accessible to OhioLINK members, whether that was one hosted in the OhioLINK Electronic Book Center or one hosted on the vendor’s web site.

In 2009, the Program for Cooperative Cataloging (PCC) released the Provider Neutral E-Monograph MARC Record Guide. This document presents a model of e-book cataloging distinctly different from that previously espoused by DMSC. Some cooperative cataloging volunteers, most notably the PCC-member institution assigned to catalog newly published Springer e-book titles, wished to begin cataloging e-books according to the provider-neutral standard.

In 2010, DMSC’s Standards for Cataloging Electronic Monographs were revised to largely reflect PCC provider-neutral guidelines. DMSC’s standard retains certain MARC fields proscribed by the PCC’s Guide, largely in response to some members’ strong beliefs that these fields (MARC 506 and MARC 710 containing a provider entry, for example) are useful to patrons and staff at their institutions. However, these fields are intended to be added to records in batch just prior to their distribution to OhioLINK members, rather than being included in OCLC master records. Since some of OhioLINK’s e-book collections are regularly expanded with newly published works and/or expressions, DMSC volunteers have begun to encounter opportunities to create original provider-neutral e-monograph records for expressions previously unrepresented WorldCat, which non-OhioLINK libraries can later utilize by adding 856 fields representing the manifestations they purchase or lease. The
adoption of provider-neutral practice thus promises to expand the cooperative benefits of DMSC’s activities beyond the consortium to all OCLC member libraries.

However, provider-neutral records are problematic for the DMSC cooperative cataloging program in ways that go beyond record standards. Prior to the introduction of provider-neutral records, OhioLINK member institutions had generally been able to load records distributed via the cooperative cataloging program into their local catalogs without fear that the records would match and overwrite existing local records. Now that provider-neutral records may be included in cooperatively produced record sets, it is possible for a member institution to purchase a manifestation of an e-book, catalog that manifestation locally, then have that record overwritten by a cooperatively produced version of the same OCLC record containing the URL for a different manifestation purchased by OhioLINK. To the extent that local institutions value separate hyperlinks to their locally purchased manifestations, this is a problem for the local institutions. It is also a problem for cooperative cataloging volunteers, who may need to revise records distributed in earlier record sets with additional URLs for later-acquired manifestations, then re-distribute the record sets.

For the time being, e-mails announcing the availability of new cooperatively produced record sets include a blanket disclaimer instructing institutions that catalog locally purchased e-books to proceed carefully in loading the records. Since all OhioLINK institutions share the same integrated library system, Innovative Interfaces’ Millennium, it may behoove DMSC to provide centralized training to interested member institutions in how to anticipate loading problems and alter load profiles to preserve existing URLs in the Millennium interface.

**Resource Description and Access**

Widespread adoption of Resource Description and Access (RDA), a standard developed to replace the Anglo-American Cataloging Rules, will also necessitate alterations
to DMSC’s *Standards for Cataloging Electronic Monographs*, as well as possible DMSC-organized training for cooperative cataloging volunteers. In December 2010, DMSC designated a subcommittee of members to investigate the potential impact of RDA on the consortium and member libraries, presumably including the impact on cooperative cataloging projects. One OhioLINK member institution is a United States RDA test site and others have begun experimenting with RDA during its U.S. testing period, so the knowledge and experience required to update standards and provide training should be readily found within the DMSC membership.

**Recommendations to others**

The history of cooperative e-book cataloging at OhioLINK reveals some of the complications and pitfalls that can occur when groups of libraries collaborate to produce and distribute bibliographic records. Based on lessons learned from the DMSC experience, several recommendations can be offered to entities embarking on cooperative cataloging projects, whether they are consortia, other official organizations, or less formally assembled groups of two or more libraries.

- Establish clear expectations about volunteer time commitments and project completion dates at the outset, and require volunteers and their supervisors to acknowledge these expectations. If necessary, perform pilot studies to provide reasonably accurate estimates of the time required to complete the project.
- Recruit at least one cataloging technology "guru" who is experienced in, or willing to explore, tools such as MarcEdit and computer scripting languages. Then, train other volunteers so that the cooperative is not dependent on specific individuals to implement these technologies.
- Establish clear but flexible bibliographic record standards. Realize that more elaborate standards may mean additional labor devoted to a project, and balance the value of the added data against the demand for immediate catalog representation of the materials.
Consider multiple methods of creating records. If vendor-supplied records are available but unacceptable in their initial state, can they be revised to an acceptable level with a reasonable amount of time and labor? Do records exist in OCLC WorldCat or other available databases that could either be used as is, or used to derive new records? In some cases, revising vendor-supplied records up to a desired standard may be more labor-intensive than finding or creating records in a bibliographic utility. Test different methods to determine which is most efficient for a particular project.

- Be flexible. Test cataloging standards and procedures with real-life examples before launching large-scale projects with multiple volunteers. If cataloging to a certain standard or using a specific procedure turns out to be more difficult or less beneficial than expected, consider revising the standard or procedure.

- Clearly and accurately document the procedure used for each project. This provides for the continued smooth operation of a project if original volunteers must withdraw before it is completed, and facilitates the use of procedures created for one project on other, later projects.

Notes


2. Title counts based on bibliographic data extracted from Ohio University Libraries’ integrated library system, accessed February 28, 2011. Titles counted included 33,554 OhioLINK Electronic Book Center and Oxford Reference Online titles (purchased by OhioLINK and cooperatively cataloged), 10,530 Safari Books Online and LearningExpress Library titles (leased by OhioLINK and cooperatively cataloged), and 14,835 NetLibrary titles (purchased by OhioLINK but cataloged via a commercial service). The number of titles cooperatively cataloged is given elsewhere as 44,000 (rather than the approximately 59,000 sum of all of the above numbers) due to the fact that the NetLibrary titles were not cooperatively cataloged.


9. Sanchez and others, "Cleanup of NetLibrary cataloging records."


18. All references to DMSC meeting minutes are to "DMSC minutes," OhioLINK Database Management and Standards Committee, accessed January 28, 2011, http://platinum.ohiolink.edu/dms/DMSdocs/DMSminutes.html.


24. Information on the process provided by Jeffrey Trimble, in discussion with the author, November 2010. For an outline, see Jeffrey A. Trimble, "MERC: managing electronic record cataloging" (lecture presented at the Northern Ohio Technical Services Librarians 2010 Fall Meeting, Parma, Ohio, November 19, 2010).


FIGURE 1: Possible pathways for cooperative creation and distribution of e-book records
<table>
<thead>
<tr>
<th>Procedure</th>
<th>Typical division of labor</th>
<th>Example collections</th>
</tr>
</thead>
</table>
| Catalog piecemeal in OCLC WorldCat | Assign cataloging to a single institution  
OR  
Divide collection among multiple institutions | ABC-CLIO; Gale Virtual Reference Library; Oxford Reference Online; Springer-Verlag (ongoing acquisitions) |
| Use vendor records with little or no enhancement | Assign record processing (if any) and contribution to union catalog to a single institution | Learning Express Library; NetLibrary |
| Enhance vendor records using technological tools (with or without eventual contribution to OCLC WorldCat as new records) | Assign record processing to a single volunteer  
OR  
Divide collection among multiple volunteers with moderate to high tech skills; one or more high-skilled volunteers provide technological support to moderate-skilled volunteers | Chadwyck-Healey (phase two); Safari books online |
| Search for metadata in OCLC WorldCat or other sources, then enhance or derive records using technological tools | Assign searching phase to one or more institutions, then funnel search results to one or more high-tech skill volunteers | Chadwyck-Healey (phase one); Springer-Verlag (backfiles) |

FIGURE 2: Possible divisions of labor for cooperative cataloging projects