Notes on Operations

E-Book Cataloging Workflows at Oregon State University

Richard Sapon-White

Among the many issues associated with integrating e-books into library collections and services, the revision of existing workflows in cataloging units has received little attention. The experience designing new workflows for e-books at Oregon State University Libraries since 2008 is described in detail from the perspective of three different sources of e-books. These descriptions highlight where the workflows applied to each vendor's stream differ. A workflow was developed for each vendor, based on the quality and source of available bibliographic records and the staff member performing the task. Involving cataloging staff as early as possible in the process of purchasing e-books from a new vendor ensures that a suitable workflow can be designed and implemented as soon as possible. This ensures that the representation of e-books in the library catalog is not delayed, increasing the likelihood that users will readily find and use these resources that the library has purchased.

The increase in the availability of e-books in the past decade has transformed how our society communicates information. As publishers transition from print to electronic, or at least provide their publications in both print and digital form, readers are changing their preferred reading formats.[1] E-books have affected the academy as well. Many university bookstores now sell e-textbook access to students while some universities have experimented with using textbooks on e-readers.[2]

Academic libraries have been affected as well, with 95 percent of American university libraries purchasing e-books.[3] Both scholarly and popular titles are increasingly available from a variety of publishers and e-book vendors. Internet-based e-books have the advantages of being accessible anywhere, require no shelf space, never need mending, and have features that are not available with print, such as keyword searching of text. They also have disadvantages when compared to print, including greater management for licensing and technological requirements, as well as restrictions on printing, interlibrary loan, course reserves, and the number of simultaneous users.

Libraries are making e-books accessible to their users via online public access catalogs (OPACs) and, for monographic series, indirectly through library websites, often referred to as "A to Z lists." (A third, future possibility for user discovery of e-books would involve the selection of e-book titles from a knowledge base and subsequent reliance on a discovery service to index the e-books.) OPAC access integrates e-books with a library's other resources, providing an advantage over website access as users can retrieve metadata on both electronic and print resources with one search. Research has shown that loading bibliographic records for e-books into the OPAC increases their discovery and use.[4] To facilitate cataloging and promote the use of their e-books, many vendors and publishers provide MARC bibliographic records for their titles. This is an especially useful service because many publishers sell their e-books in packages containing hundreds or even thousands of titles; cataloging such sets title-by-title is beyond the capacity of many cataloging departments. However, the poor quality of vendor

Richard Sapon-White is Head of Cataloging and Metadata Services, Oregon State University, Corvallis, Oregon

Submitted March 21, 2013; tentatively accepted June 24, 2013, pending modest revision; revision submitted August 22, 2013; accepted for publication October 1, 2013.

records is often a concern; alternate sources can provide high-quality bibliographic records for e-books.

To catalog e-books efficiently, traditional technical services workflows for print materials need to be modified. A major difference from print is that no physical piece arrives at the library to trigger the processing of an e-book. Without boxes to open, barcodes and call number labels to apply, or security strips to insert, e-books "arrive" ready to catalog as soon as a cataloging unit has been notified that an e-book has been purchased. For e-books, the principal tasks for technical services staff include licensing and purchasing, downloading bibliographic records, and, often, batch editing of those records. Since e-book publishers provide bibliographic records to libraries in a great variety of ways, cataloging workflows may need to be created *de novo* for each vendor.

At Oregon State University Libraries (OSUL) before 2008, e-book packages were rarely purchased, and firm orders for individual e-book titles were seldom sent to vendors, primarily because of their limited availability and lack of adequate platforms for user access. When vendors began to offer discounted packages, especially of science and technology e-books, OSUL began actively acquiring them from publishers and vendors. This led the libraries, in 2011, to set a policy favoring the purchase of electronic over paper resources whenever possible. Since then, OSUL has purchased e-books from many different vendors, resulting in different workflows for each vendor. These workflows differ according to the sources of bibliographic records, the quality of those records, the methods used to load records into the catalog, and the staff involved in e-book processing. Established workflows are documented on the OSUL wiki.[5] This paper examines three e-book workflows at OSU in detail to highlight the successes and difficulties encountered in their design. Through this case study of cataloging unit workflow design, other academic libraries can better understand potential problems and solutions. Since the focus is on cataloging workflow, aspects of e-book acquisitions such as licensing and negotiating fees are not covered in this paper.

Literature Review

In the context of a library cataloging and/or metadata department, workflow is the sequence of steps applied to an information resource from the point of acquisition by the library to the point where the resource is represented by a record in the library's catalog. A description of the cataloging workflow for e-books might include the processes that the e-books undergo (i.e., downloading bibliographic records into the local catalog, editing URLs to incorporate proxy information, or editing records to ensure they meet local cataloging standards), the staff responsible for each step, and the time taken for staff to perform each step. This literature review examines various aspects of e-book workflows, although few articles address workflow directly.

Wu and Mitchell describe workflows for collections of e-books at the University of Houston Libraries (UHL), which has relied heavily on vendor-supplied bibliographic records.[6] They state that the problem for catalogers is not that e-books are difficult to catalog but that their workflow is difficult to manage because of the unpredictability of vendor record quality. While the PCC's guidelines for provider-neutral records has simplified cataloging at the individual MARC field and record level, it does not address the problems faced when batch processing vendor records.[7] This is because e-book bibliographic records come from many sources and follow few standards. Purchasing individual e-books, although likely to follow a traditional workflow, presents the challenge of being difficult to track through the process. Workflows for e-books at the UHL vary depending on whether purchased singly or in batches. Open access titles are handled differently from reference works. The management of batches of records requires different skills of their catalogers than traditional cataloging of tangible books. They also discuss the problem of not showing UHL e-book holdings in WorldCat because of their use of vendor records and lack of an easy step for identifying matching bibliographic records. The importance of communication between the acquisitions and the cataloging department concerning the arrival of new e-resources is also discussed.

Dealing with the problems encountered when loading vendor-supplied records is a common theme in e-book literature. Martin and Mundle discuss the cataloging of Springer e-book collections at the University of Illinois at Chicago (UIC).[8] They describe the problems with using vendor-supplied MARC records in three areas: poor record quality (i.e., lack of authoritized headings), loading problems, and access problems (i.e., broken links). They conclude that while vendor records may be provided at no charge, costs are incurred by the staff time and effort to revise those records. While workflow is not described, these problems needed resolution to establish a smooth workflow for e-books.

The library also decided to load vendor records into their OPAC to facilitate access to a package of 12,000 e-books.[9] They charged a task force of three catalogers to evaluate vendor records for quality. Task force members required that vendor records be provided in MARC format, adhere to national cataloging standards for full records, and include subject headings—either Medical Subject Headings (MeSH) or Library of Congress Subject Headings (LCSH). Problems

encountered again included quality issues (presence of extraneous and unnecessary fields), load issues (improper formulation of the MARC 001 and 003 fields needed for overlaying records), and access issues (lack of dual URLs in each record as desired). Through examination of sets of vendor records, they identified problems with vendor-supplied MARC records and developed guidelines for other libraries to use to incorporate vendor-supplied records into their catalogs. Bulk corrections were made using MarcEdit, a popular freeware tool capable of batch editing MARC records,[10] but some records needed individual attention. In addition to addressing vendor-supplied MARC records, the authors describe the original cataloging of health science e-books, which lacked vendor-supplied records.

Other articles focus on the poor quality of vendor-supplied bibliographic records.[11] They typically describe the procedures used to revise records to meet cataloging standards. Global editing functions in integrated library systems (ILS), and utilities such as MarcEdit have facilitated such database maintenance.

Libraries have experienced difficulties when loading vendor-supplied records into their catalogs. Victoria University provided access to e-books from NetLibrary by loading vendor-supplied MARC records into its catalog.[12] The records were of high quality, but the process of loading them was problematic, including the unwanted generation of order records. Although reporting on various technical problems during two record loads, no information is given about which staff were responsible for e-book processing, how often loads occurred, or how much time elapsed before records became available to the library.

Gedeon and Meyer described the acquisition and cataloging of e-books at Western Michigan University.[13] Focusing on seven e-book vendors, they noted how titles are acquired, including both subscription and free. Some vendors provided package deals while others supplied works title-by-title. They also used vendor-provided records for e-books, but the records lacked quality and authority control. Loading the records into their catalog required the collaboration of a systems librarian, a programmer, and a catalog librarian. Revision of the records by cataloging staff was an additional expense of the e-book purchase and delayed loading of the records into their ILS.

At the University of Colorado Denver's Auraria Library, a study was conducted to assess bibliographic records for titles freely available on the Internet.[14] Before loading, a sample of the MARC records was examined for quality. Although record quality was determined to be poor and diacritic incompatibility with the local system caused the insertion of extraneous characters, the library decided to proceed with loading the records. Making the records available through the catalog, as opposed to merely linking to the entire collection from the library's webpage, was considered to be valuable because the catalog provides aggregation and the e-books collection helped fill gaps in the library's physical collections. As more open access, digitized books—such as the more than five million available from the Hathi Trust (www.hathitrust.org)—become available, libraries will need to plan and manage how best to provide access to them.

Vendor-supplied records can require significant editing, especially when access points need to be revised to match authority records. At Virginia Tech, incoming vendor-supplied records are first sent to the library's authority control vendor for processing. Based on the headings report from the vendor, unmatched headings are edited using MarcEdit, with some additional revisions done via their ILS. The report does not identify who at their institution performs this work, but it does provide the sequence of quality control editing.[15] Similarly, many libraries are concerned about the vendor records supplied with packages of e-books as the records generally do not meet standards for quality control, both in description and in authorized headings. Some have devised means of cleaning up the records, using batch editors such as MarcEdit.[16]

Since both the Anglo-American Cataloguing Rules, 2nd edition (AACR2) and Resource Description and Access (RDA) require separate records for each e-book publisher or distributor, an e-book available from multiple vendors is often represented in OCLC by a plethora of records. Guidelines for e-book cataloging changed significantly in 2009 with the publication of the Program for Cooperative Cataloging (PCC) guidelines on the creation of provider-neutral bibliographic records for e-books. The guidelines address the issue of multiple records for the same intellectual content by removing all provider-specific data from the bibliographic record. This single generic e-book record has multiple MARC 856 fields, each with a provider-specific URL.[17]

Oregon State University Libraries Overview

Oregon State University is a land, sea, sun, and space grant institution with approximately 26,000 students and 3,500 faculty. The OSUL holdings include more than 1.4 million volumes, 14,000 serial subscriptions, and more than 500,000 maps and government documents. A main library and veterinary medicine library on the Corvallis, Oregon, campus are complemented by two branch libraries serving remote facilities of the university.

The libraries acquire approximately 15,000 monographs annually in addition to receiving about 5,000 government documents. Of the firm-ordered and approval plan

monographs, approximately 95 percent have copy available in OCLC (either Library of Congress– or member-contributed) with full-level cataloging, including call numbers and subject headings. Until recently, most cataloging copy at OSUL was downloaded in a "fast-cat" process by a support staff member. Since 2010, almost all firm orders and approvals arrive as shelf-ready books with an associated file of bibliographic records. The head of cataloging is the only professional cataloger in the unit, with time split between administration of the unit and providing original and complex copy cataloging. Until recently, other members of the unit included 7.5 FTE library technicians responsible for cataloging monographs and serials. In 2012, a reorganization of the Center for Digital Scholarship and Services, which includes the cataloging unit, reduced the unit size to four FTE. The cataloging unit participates in the Name Authority Cooperative Program (NACO), the Subject Authority Cooperative Program (SACO), and the Cooperative Online Serials Program (CONSER). OSUL uses Innovative Interfaces' Millennium ILS.

OSUL belongs to the Orbis Cascade Alliance (OCA), a consortium of thirty-seven academic libraries in the Pacific Northwest. OCA provides many services, including consortial pricing on databases, a union catalog (Summit), and a patron-driven acquisitions (PDA) program.

OSUL purchases e-book packages in the humanities, life sciences and physical sciences. This e-book purchasing policy complements one of purchasing e-serials whenever possible as the libraries transition from a print institution to a predominantly electronic one.

Workflows

Before 2010, print resources moved through OSUL's technical services in a workflow familiar to many libraries. The acquisitions department handled ordering (including downloading a brief bibliographic record for each title), receiving, and paying invoices, and passed resources to the cataloging department to search for a bibliographic record in OCLC and subsequent cataloging. A cataloger edited the record if one was found or created an original record if no record was found, then downloaded the record to the local system, attaching an item record and completing physical processing of the book. Since 2010, OSUL has received shelf-ready books from its major vendor as well as a file of bibliographic records from OCLC for each set of orders. This operation is primarily handled by acquisitions staff. Occasional titles needing call numbers or subject headings, or requiring original cataloging, are passed along to the cataloging unit for completion.

The workflow for e-books is quite different:

- For each publisher or distributor, license negotiations and contracts are often required before ordering can begin. Staff within acquisitions and collection development perform these roles, which will not be discussed here.
- Notification of the purchase of one or more e-books needs to be communicated to cataloging staff because no physical piece is passed along to monographic catalogers. This is necessary for both individual, firm-ordered titles and e-book packages. For the former, acquisitions staff notify the cataloging unit at the time of order because titles are immediately available and no order record is needed in the library catalog. For packages, the collection development department alerts the cataloging unit of the impending purchase so that catalogers can determine the best method for acquiring bibliographic records to load into the catalog.
- Order and receipt may happen almost simultaneously, with access provided immediately on payment of the order. Workflows for print books assume a lapse in time between ordering and receiving; therefore these needed to be modified for e-books. With immediate access to electronic titles following purchase, library users do not need an order record in the catalog to be aware of an upcoming purchase.
- Each publisher or vendor has its own way of supplying bibliographic records. Bibliographic records may be available directly from vendors, downloaded title-by-title or in batches from OCLC, or sold in publisher-specific sets from OCLC.
- Bibliographic records may not exist for the electronic version but can be derived from the print version record.

With the exception of PDA titles (i.e., those unpurchased), the e-books discussed below were permanent acquisitions. As such, OSUL wanted to have these titles in the public catalog as well as the OCA union catalog, Summit. For them to display in the latter, all of the titles needed to have holdings set in OCLC. Some vendors' e-books were available for interlibrary loan, so having OSUL's holdings in OCLC is a significant service to users.

OSUL has purchased e-books from many vendors and publishers, including E-Book Library (EBL), Springer Science+Business Media (Springer), and Morgan and Claypool. The library also circulates Kindles; all of the e-books downloaded to these devices are cataloged. Workflows for the Kindle titles have been detailed in another article.[18] In addition to these purchases, OSUL has loaded

more than 400 records made available by the Colorado Library Consortium for freely available Project Gutenberg e-books.

OSUL has been involved in two PDA projects, both using EBL as the vendor. The first project involved a contract between OSUL and EBL; the later project, begun in July 2011, was a similar contract between OCA and EBL. All titles appeared in the OPAC and involved monthly or weekly loading of a file of bibliographic records. Once a title had been viewed a set number of times, the library automatically purchased it.

To illustrate how much variation exists in each vendor's e-book program, the details for four vendors appear below: EBL purchased titles, EBL PDA projects, Springer, and Morgan and Claypool.

EBL

In September 2008, OSUL signed a contract with EBL and Blackwell North America (BNA) to purchase e-books from a variety of publishers, primarily science and technology titles. These were to be ordered individually by subject librarians.

EBL offered free MARC records for these purchases, which could be downloaded into a local catalog. Since OSU's existing workflow involved downloading a bibliographic record at the time of ordering, the initial decision was to use the vendor's records at the point of ordering. These could later be overlaid with an OCLC record. The vendor records had the following characteristics:

1. Call numbers with "eb" appended to the date. For example, "QA612.7 .N452010eb" was used for the title *Algebraic Methods in Unstable Homotopy Theory*. This method simplifies the identification of e-books by users when scanning a list of call numbers.
2. MARC 650 subject fields had a first indicator blank and second indicator "4." This indicates that the source of the subject heading is not specified and is likely not from *Library of Congress Subject Headings*.
3. Each record included the corporate body added entry (MARC 710 field) "E-books Corporation Pty. Ltd.," which is not an authorized heading.
4. None of the records included OCLC control numbers.

At that time, OSUL also contracted with OCLC Cataloging Partners to provide quality bibliographic records for all of the EBL orders as well as tagging records in WorldCat with OSUL's holdings symbol. This three-way arrangement involved EBL, with whom orders were placed; BNA, OSUL's primary book vendor at that time and the entity responsible for billing; and OCLC, to provide the bibliographic records. The purchase of an e-book title from EBL triggered a process that resulted in OCLC sending a bibliographic record to OSUL approximately eight weeks after the order was placed. Unfortunately, only one bibliographic record was actually delivered during the first year of ordering although over two hundred titles were purchased. For this reason, OSUL ultimately cancelled the contract with OCLC Cataloging Partners and chose to catalog the EBL titles one-at-a-time, following each order. The process of downloading vendor records into the catalog was discontinued. To ensure timely cataloging, the head of cataloging trained a library technician in electronic resource cataloging using the Program for Cooperative Cataloging (PCC) provider-neutral e-monograph record guidelines. The library technician was able to search for and adapt cataloging copy from OCLC or create original records when needed. The backlog of EBL titles created during the first year of ordering was eliminated in two months.

The current workflow for e-books ordered from EBL follows:

1. Acquisitions staff place an order for each title through EBL's website. EBL emails a confirmation of the order and a notification of the title's availability to acquisitions staff.
2. Acquisitions staff notify a library technician in the cataloging unit of the title once it has been ordered.
3. The cataloging unit library technician proceeds as with any print title, including searching OCLC, downloading a bibliographic record if one exists or creating a record if none is found. The technician then attaches an item record in the local system indicating that the title is available to the university community via the Internet. Cataloging a title usually happens within twenty-four hours of it being ordered.
4. Library users who find the title need to enter their university identification number and password to read the e-book.

EBL Patron-Driven Acquisitions (PDA)

In June 2010, OSUL began participating in another EBL service, a patron-driven acquisitions program (PDA). EBL provides bibliographic records for titles of potential interest to OSUL users. The records include links to the e-books at EBL's website. After users view a title an agreed upon number of times, OSUL automatically purchases the e-book.

At the beginning of the program, EBL created a list of titles to match a profile based on subject and publishers. To avoid duplication of titles already owned, OSUL sent EBL a file of the ISBNs for all post–2006 imprints. These titles were then deleted from any sets of e-book records sent to OSUL by EBL.

Each month, EBL emailed OSUL a file of MARC records to load into the catalog. The size of the file has changed over time, starting at about 4,500 titles and currently numbering 12,000. On receipt of the file, the previous month's records are deleted from the catalog before the new records are loaded. Millennium's Data Exchange module is used for record loading. The records are minimally edited, primarily adding proxy data to the URL and changing the material type byte to the symbol "@." The latter indicates that the record represents an e-book and enables searchers to limit queries to e-book records. While the loading and editing of these monthly bibliographic record files was originally done by the head of the cataloging unit, a library technician now handles the task.

The EBL bibliographic records lack the quality usually desired in a university catalog. Name headings may not conform to the form found in the National Authority File, subject headings are often overly broad, and titles are sometimes provided in their prepublication form. For example, the title *The Role of Internal Audit in Corporate Governance in Europe* has the corporate heading "(ECIIA), European Confederation of Institutes of Internal Auditing." The correct heading lacks the acronym in parentheses. In addition, the heading is tagged as if it were a conference name. Similarly, the editor's name does not appear in the descriptive fields and the form of the name as an added entry lacks the date of birth that appears in its authorized form (i.e., "Schartmann, Bernd" instead of "Schartmann, Bernd, 1962-"). Finally, although "Auditing, Internal" and "Corporate governance" are valid subject headings, the only subject heading provided by EBL is "Business"—with no geographic subdivision for Europe.

The task of revising subject headings on thousands of temporary bibliographic records is too time-consuming. The effort would also need to be repeated monthly with each new record load; leaving the records as they are results in a much simpler process.

Once an order has been triggered, an acquisitions staff member is notified and the title is cataloged like any other e-book purchased from EBL. When the full record is downloaded from OCLC, it overlays the existing EBL-provided record. When the next month's file is downloaded, the purchased title is no longer included in the file.

Springer

OSUL contracted with Springer Verlag to purchase subject-related e-book sets. These subject sets comprise hundreds of titles, resulting in the need to find simple solutions to catalog a large number of titles quickly and efficiently. Approximately 8,500 e-books have been purchased from Springer from February 2009 through March 2012.

Initially, Springer provided two methods to acquire bibliographic records for its titles. First, vendor-created records could be downloaded free of charge from the Springer website. As with the EBL records described above, these were of poor quality. Typically, the ISBN was recorded in the MARC 001 field, the title was recorded in title case, the extent of item was given simply as "v. : digital," and the subject headings were overly broad. To illustrate this last point, *Multichain Immune Recognition Receptor Signaling: From Spatiotemporal Organization to Human Disease* has the subject heading "Medicine." In addition, and contrary to current practice, the series was traced only in a series statement (MARC field 440) rather than using the combination of series statement and uniform title added entry (MARC fields 490 and 830). These vendor-supplied records often list all authors in the statement of responsibility and trace all authors as added entries. While this is permissible in RDA, most libraries at the time of this writing are still using AACR2 as their catalog code and recording all authors is not an acceptable practice.

Springer's bibliographic records also lack OCLC control numbers. This number is useful in retrieving a bibliographic record in OCLC, usually of better quality, so that OSUL can set its holdings. Tagging records with OSUL's symbol is critical for showing holdings in WorldCat as well as in the Summit union catalog. Searching for OCLC records one at a time would require considerable staff effort and time, a prohibitively inefficient way of processing large numbers of titles.

These records have an advantage in that they are available shortly after publication. They could be overlaid later with a fuller bibliographic record matching on ISBN, although ISBN-matching presents a problem itself. Because many libraries use a single record approach, a record could already be present in the library's catalog for a print version that includes the ISBN for the electronic version. If a library uses a separate record approach to electronic versions but the record used for a print version includes the ISBN for the electronic, then the print will be retrieved with the e-ISBN, causing confusion as to which version the library actually owns.

As a second avenue for acquiring bibliographic records, Springer originally maintained a spreadsheet that included OCLC control numbers for its e-books; the spreadsheet was freely available on its website. The spreadsheet listed the ISBNs (both for the electronic and the print versions) for each title, plus series, OCLC control number, copyright year, Springer subject collection and the "added to WorldCat until" date. This date was when the entry was made in the spreadsheet. The spreadsheet was updated several times a year on an irregular schedule.

The head of cataloging periodically downloaded this spreadsheet and edited it to exclude collections not

purchased and titles for which records had already been downloaded. The remaining titles were those that OSUL had purchased. OCLC control numbers from the spreadsheet were then used to batch download bibliographic records to the local ILS, attaching OSUL's holdings symbol to the OCLC record at the same time. Once the records were downloaded, some minor editing was done, including adding proxy data to the URL in the Electronic Location and Access field (MARC field 856), changing the material type to "@" (for e-resources—a local practice), and removing 856 fields for publisher's descriptions.

Although this process appears to be straightforward, it was not a foolproof way of retrieving bibliographic records for all of the library's purchases. Because Springer did not create bibliographic records in OCLC, the publisher had to rely on library catalogers elsewhere to create records for Springer titles. Since there can be a time lag between when an e-book is published and when it is cataloged, relying on Springer's spreadsheet would only retrieve records for some of the purchased titles. For each set of purchased titles, some had OCLC records available immediately while others did not. OSUL found that for some titles, the lag time between publication of a title and the appearance of an OCLC record (and therefore inclusion in the spreadsheet) could be as much as a year or more. Fortunately, this situation applied only to a small number of titles for each set.

OSUL chose to use the spreadsheet, loading bibliographic records for purchased titles as they became available. Titles without a record in the local ILS would continue to be accessible through the Springer website, although users would need to know to go to the website to find them, a situation that was highly unlikely. The periodic examination of Springer's spreadsheet was done approximately quarterly by the head of the cataloging unit.

To assess how many of the purchased titles from Springer were not yet represented in the ILS, staff loaded Springer-provided MARC records and set the system to reject any ISBN that matched an existing record. About thirty titles not previously cataloged were added to the ILS. The process was quite time-consuming and imperfect, as e-ISBNs also appeared on print records, producing inaccurate matches. Fortunately, the percentage of titles purchased that lacked records in the system was very small.

In 2011, Springer changed how it provides records for its e-books. The publisher arranged with OCLC's Collection Sets service to provide bibliographic records from WorldCat at no cost to libraries. Through this service, a "subscription" can be set up to provide notifications when a new set of bibliographic records became available. Although the time-lag issue described above has persisted, the process of downloading records as they become available has been greatly simplified. The head of cataloging continues to perform this task because it involves restricted permissions. However, the task could conceivably be passed on to a library assistant following training and adjusting permissions in the ILS.

Morgan and Claypool

In 2010, OSUL purchased Morgan and Claypool's Synthesis I and II collections of e-books in computer science and engineering. Morgan and Claypool provides bibliographic records for these collections that are freely available for downloading from their website. Optionally, libraries can also purchase a "collection set" of bibliographic records from OCLC. These MARC records have been created by OCLC member libraries for the titles in the Synthesis collections.

Unlike other publisher-provided MARC records, Morgan and Claypool's bibliographic records are good quality and follow prevailing cataloging standards. Morgan and Claypool contracted their cataloging to a company with skilled and knowledgeable catalogers, Special Libraries Cataloguing.[19] As with the other vendor-supplied records, however, these records do not include OCLC control numbers and so no mechanism is provided by which the publisher is able to tag the OCLC record with a library's holdings symbol.

OSUL opted to purchase the collection set from OCLC since setting holdings in WorldCat is a critical step for resource discovery. The process for downloading the records is straightforward. Staff navigate to the OCLC website, place the order for the record set, and after processing the charges, the set is ready to transfer by ftp to the OSUL catalog using Millennium's Data Exchange module. Initially, the head of cataloging performed this work to deal with any problems, but this task was readily passed along to a library technician following minimal training.

Discussion

The four workflows described above are summarized in table 1.

Since embarking on an aggressive program of purchasing e-books, OSUL has learned much about designing workflows that best integrate the cataloging of these resources into the unit's daily routines. The factors having the greatest effect on designing cataloging workflows for e-books are

- whether bibliographic records are ordered and downloaded in bulk or individually;
- the method for downloading bibliographic records;
- the quality of the bibliographic records, including description, subject analysis and the presence or absence of OCLC control numbers; and
- training personnel.

Table 1. Summary of Characteristics for E-book Workflows at OSUL.

Vendor	Source of Bibliographic Records	Method of Downloading	OSU Personnel Performing Task	Use PCC Vendor Neutral Policy Guidelines?
EBL	WorldCat via Connexion	One by one	LT3	Yes
EBL Patron-Driven Purchase Program	EBL	Batch file download	LT3	No
Springer	WorldCat via Connexion*	Batch search and download	Head of cataloging	Yes
Morgan & Claypool	WorldCat Cataloging Partners	Purchased set	Head of cataloging	No

*Records for Springer titles are now available through WorldCat Cataloging Partners.

The source of bibliographic records for e-books often determines much of the workflow. Records downloaded directly from a vendor allow for a simple process. Some vendors periodically notify the library staff when new records are available, as was the case with OSUL's PDA program with EBL. With minimal editing, OSUL was able to coordinate these monthly downloads, although needing to accept less-than-ideal record quality. The source of records for most of OSUL's purchased e-books was OCLC, although three different methods of acquiring them were used: downloading record sets through the WorldCat Selection Sets program, the WorldCat Cataloging Partners program, and downloading individual records via OCLC's Connexion client. For some e-book sets, such as University of California Press titles (not described above) or Springer books, batch downloading using Connexion was quick and easy. The source of records usually correlated with their quality, as vendor records often required much editing, while OCLC records required little if any editing.

Vendor records are appealing to use because they are often available free of charge. For libraries that are not OCLC members or have limited staff, these records provide a quick way of providing some kind of catalog access to e-books. The quality of that access, however, may be poor. Many vendor records do not follow current descriptive standards, lacking statements of responsibility and subject headings, having titles entirely capitalized, or using unauthorized forms of name headings. Such records could be used as place-holders in the catalog until better records are found to overlay them, but the maintenance and deletion of vendor records requires considerable staff time and effort—a cost that library decision makers must consider when vendors offer "free" bibliographic records along with their e-books. If dealing with a small number of titles, the best practice is to avoid loading vendor records and, instead, download records on a title-by-title basis, as is often done with printed works in conventional cataloging workflows.

OSUL is an OCLC member library, and quality bibliographic records for e-books are considered critical for discovery. That quality includes the presence of an OCLC control number. Interlibrary loan depends on these identifiers, as does the OCA Summit catalog. If a vendor's records include the OCLC control number, the process of tagging the matching OCLC records with a library's holdings symbol is greatly simplified. Unfortunately, in many instances, vendors do not have access to control numbers and cannot supply them in bibliographic records. When evaluating vendor records, the presence of the OCLC control number should be a significant factor in decision-making.

Staff training is an important consideration when integrating e-book cataloging into the department's daily work. For staff accustomed to cataloging print materials, the process of downloading vendor records may seem strange since no physical pieces exist. Copy catalogers may need to download the book to view its title page and compare with the OCLC record when cataloging title-by-title. A thorough understanding of the standards in place for cataloging electronic resources and especially the PCC's policy on vendor-neutral records can go a long way toward alleviating anxieties about cataloging e-books. Staff may need to learn to download records from vendor's websites or open up zipped files of bibliographic records.

Initially, OSUL's head of cataloging and metadata services took responsibility for designing the workflow for each e-book vendor and handled the downloading and cataloging for a sample batch of the first few titles. Once procedures were established and documented, whenever possible, he trained a staff member to handle this work.

Staff also need to be familiar with the e-book cataloging workflows to know how to retrieve and edit records. Because nearly every vendor's records have their own workflow, written procedures that are readily accessible to

catalogers should be distributed to staff, archived and kept up-to-date. These procedures are present on the unit's wiki along with other unit policies and procedures.[20] Adding the library's proxy data and a public note to the MARC 856 fields rounded out the topics to be covered when training staff to catalog e-books.

Communication and planning are essential to ensuring that workflows are designed in advance of e-book acquisitions. This can help avoid the situation where the collection development department is expecting titles in an e-book package to be cataloged, but the cataloging unit has not yet been notified that the e-books have been purchased. The timely communication of such information requires that those deciding on the purchase of e-book packages notify the cataloging unit as early as possible in the process. The cataloging manager can then request information from the publisher or vendor about the availability of bibliographic records and begin planning from whom those records will be acquired (i.e., the vendor, OCLC, or some other source), how they will be downloaded into the catalog, how much post–download editing will be required, and which personnel will be responsible for this process. When the e-books then become available, the cataloging unit is already prepared to catalog them. Without the input of catalogers, and with many collection development librarians lacking the expertise to evaluate the quality of a vendor's MARC records, vendor promises of free MARC records and easy catalog access may sound very attractive. Understanding the reach of WorldCat records makes for an easy decision to forgo using vendor records.

One problem that OSUL continues to struggle with is the number of duplicate e-book purchases. With so many incoming streams of e-books, the libraries sometimes purchase a title without realizing that it is a duplicate. For example, a Springer title may be purchased individually via EBL, but the same title has also been purchased as part of a subject-defined set. Such errors in purchasing were initially discounted, and the records sometimes merged when they were discovered. However, the problem appears to be increasing as more e-book records are loaded into the catalog. At present, we have not yet found an adequate solution to this problem and are relying on chance discoveries to take care of the duplicate records.

Lastly, another step in the cataloging process needs to be noted. OSUL uses Innovative Interfaces' Millennium ILS. Millennium uses load tables that translate incoming MARC records into Millennium's proprietary record format. Load tables also can be used to delete or add specific fields from incoming records. For some of the previously mentioned workflows, new load tables were created to download MARC records with the least amount of post–download editing, such as adding the OSUL proxy into the uniform resource locator (URL). While creating a new load table can be a straightforward process, libraries using Millennium may need load table training and troubleshooting assistance when new e-book packages are purchased. Other library systems may require similar steps to ensure that the downloading of e-book records is performed as efficiently as possible.

Conclusion

OSUL has found that each new e-book vendor requires the design of a workflow tailored to the quality and method of downloading bibliographic records for its publications. With publishers increasingly moving to electronic publication and the rapidly growing number of available e-book titles, libraries need to be expert designers of workflow. The present study highlights several aspects of establishing workflows of which library managers should be aware.

First, staff need training in multiple methods of downloading and manipulating bibliographic records. Facility with using bibliographic software such as MarcEdit and understanding the capabilities of the local ILS should enable cataloging staff to readily participate in the design of new workflows. An agile staff eager to learn new ways of doing their tasks can help ensure that new sources of e-books and bibliographic records are quickly integrated into the daily work of the cataloging unit.

Second, the transition from a traditional workflow to newly developed ones should be a matter of concern beyond the cataloging unit. Communication between collection development and technical services is critical to linking the purchase of e-books with their timely cataloging. Both departments need to be aware of how e-book workflows differ from those for tangible resources so that resources are processed as soon as they are available from the vendor.

Third, staff need to document e-book procedures so that all staff are aware of the differences between vendors. Making such documentation widely available ensures that the absence or departure of personnel does not result in workflow failure.

Last, feedback needs to be provided to vendors about record quality and the importance of linkage with WorldCat. Many libraries rely on WorldCat, at least in part, for their presence on the web as well as for interlibrary loan and other services. The provision of poor-quality records without any connection to that shared bibliographic system is an obstacle for libraries to share their holdings with other institutions.

The problem of purchasing duplicate titles from multiple vendors is one that has not yet been fully addressed at OSUL. Further research needs to be done on the extent of this problem and on ways to mitigate it.

References and Notes

1. Annette Lamb and Larry Johnson, "Nurturing a New Breed of Reader," *Teacher Librarian* 39, no. 1 (October 2011): 56–63.
2. Jeffrey R. Young, "This Could Be the Year of E-textbooks, If Students Accept Them," *Chronicle of Higher Education*, September 7, 2009, accessed March 21, 2013, chronicle.com/article/The-Year-of-E-Textbooks-/48305.
3. Rebecca Miller, "Dramatic Growth: *LJ*'s Second Annual Ebook Survey," *Library Journal*, October 12, 2011, accessed February 11, 2013, www.thedigitalshift.com/2011/10/ebooks/dramatic-growth-ljs-second-annual-ebook-survey.
4. Marc Langston, "The California State University e-Book Pilot Project: Implications for Cooperative Collection Development," *Library Collections, Acquisitions & Technical Services* 27, no. 1 (2003): 19–32.
5. "Electronic Resources Workflows," Oregon State University Libraries, Center for Digital Scholarship and Services, accessed August 20, 2013, wiki.library.oregonstate.edu/confluence/display/TechServ/Electronic+Resources+Workflows.
6. Annie Wu and Anne M. Mitchell, "Mass Management of E-Book Catalog Records: Approaches, Challenges, and Solutions," *Library Resources & Technical Services* 54, no. 3 (July 2010): 164–74.
7. Becky Culbertson, Yael Mandelstam, and George Prager, *Provider-Neutral E-Monograph MARC Record Guide: Includes Revisions to September 2011* (Washington, DC: Program for Cooperative Cataloging, 2009), accessed March 21, 2013, www.loc.gov/aba/pcc/bibco/documents/PN-Guide.pdf.
8. Kristin E. Martin and Kavita Mundle, "Notes on Operations: Cataloging E-Books and Vendor Records: A Case Study at the University of Illinois at Chicago," *Library Resources & Technical Services* 54, no. 4 (October 2010): 227–37.
9. Kavita Mundle, "Integration of Electronic Books into Library Catalogs: The UIC Library Experience," *Acquisitions Librarian* 19, no. 3/4 (2008): 389–407.
10. MarcEdit, accessed September 20, 2013, people.oregonstate.edu/~reeset/marcedit/html/index.php
11. See Kristin E. Martin, "Cataloging e-Books: An Overview of Issues and Challenges," *Against the Grain* 19, no. 1 (February 2007): 45–47; Elaine Sanchez et al., "Cleanup of NetLibrary Cataloging Records: A Methodical Front-End Process," *Technical Services Quarterly* 23, no. 4 (2006): 51–71.
12. Lesley Nelson and Frances O'Neil, "Electronic Monographs in the Academic Library: An Implementation Story," *LASIE: Library Automated Systems Information Exchange* 32, no. 2/3 (August/December 2001): 13–20.
13. Randle Gedeon and Bettina Meyer, "E-books at Western Michigan University: A Case Study," *Against the Grain* 17, no. 1 (February 2005): 52–54.
14. Jeffrey Beall, "Free Books: Loading Brief MARC Records for Open-Access Books in an Academic Library Catalog," *Cataloging & Classification Quarterly* 47, no. 5 (2009): 452–63.
15. Mary Finn, "Batch-Load Authority Control Clean-Up Using MarcEdit and LTI," *Technical Services Quarterly* 26, no. 1 (2008): 44–50.
16. Carrie A. Preston, "Cooperative e-Book Cataloging in the OhioLINK Library Consortium," *Cataloging & Classification Quarterly* 49, no. 4 (2011): 257–76.
17. Culbertson, *Provider-Neutral E-Monograph MARC Record Guide*.
18. Richard Sapon-White, "Notes on Operations: Kindles and Kindle E-Books in an Academic Library: Cataloging and Workflow Challenges," *Library Resources & Technical Services* 56, no. 1 (January 2012): 45–52.
19. "Morgan & Claypool Publishers MARC Records," Morgan & Claypool, accessed August 2, 2013, www.morganclaypool.com/page/marc.
20. "Electronic Resources Workflows," accessed August 2, 2013, wiki.library.oregonstate.edu/confluence/display/TechServ/Electronic+Resources+Workflows.