Notes on Operations

Using Batchloading to Improve Access to Electronic and Microform Collections

By Rebecca L. Mugridge and Jeff Edmunds

Batchloading bibliographic records into the catalog, as a rapid and cost-effective means of providing access to electronic and microform collections, has become in recent years a significant workflow for many libraries. Thanks to batchloading, previously hidden collections, some costing hundreds of thousands of dollars, are made visible, and library holdings are more accurately reflected by the online catalog. Subject specialists report significant increases in the use of electronic resources and microforms within days (and sometimes only hours) of loading record sets into the online catalog. Managing batchloading projects requires collaboration across many library units, including collection development, acquisitions, cataloging, systems, and public services. The authors believe that their experiences will be instructive to other libraries and that Penn State's processes will assist them in making their own batchloading policies and procedures more efficient.

In the age of Google, when digital natives expect everything—or almost everything—to be discoverable online, libraries face the ever more daunting task of providing title-level access to online resources in their catalogs. Providing access to large microform and digitized collections for which no or only limited (i.e., collection-level) access in the public catalog exists is similarly challenging. Batchloading bibliographic records into the catalog is a rapid and cost-effective means of meeting these challenges.

Given its cost-effectiveness and the wide availability of record sets describing large collections, batchloading has become a significant workflow for many libraries. As more print resources are digitized, more born-digital projects created, and metadata becomes easier to convert and repurpose for bibliographic description, Machine-Readable Cataloging (MARC) records for more collections are likely to become available. Such record sets can be expensive, but given the immense improvement in access to collections they provide compared to a single collection-level record, they are often worth the price.

Some vendors supply MARC records as part of the packages they sell, realizing that libraries may be more likely to purchase or license a resource when they know that bibliographic records will ensure that individual titles in the collection are discoverable in the catalog. In fact, some institutions, individually or in concert, may find that lobbying vendors to make records available for every resource they sell is advantageous. Use of electronic resources is inextricably linked to discoverability, and evidence suggests that title-level records in a library's catalog increase use. At Penn State University Libraries, subject specialists report significant increases in use of electronic resources and microforms within days (and sometimes within hours) of loading record sets. With each batchloading of records, previously hidden collections are made visible, and the vast richness of the libraries' holdings is more accurately reflected by the catalog.

Managing the process of batchloading requires collaboration across several library units. Acquisitions staff work with subject specialists and budget officers to

Rebecca L. Mugridge (rlm31@psu.edu) is Head, Cataloging and Metadata Services and **Jeff Edmunds** (jhe2@psu.edu) is Cataloging and Metadata Specialist at Pennsylvania State University Libraries, University Park.

negotiate with vendors and purchase resources. Collection development librarians decide which files to purchase and set priorities for the order in which to load files. Public services staff review records to ensure their constituents' needs are being met. Cataloging staff assess record quality, customize record sets to meet local needs, and coordinate loads. Systems staff load records and manage the extraction of records for vended authority control.

Penn State University Libraries have devoted substantial financial and staff resources in transforming batchloading (originally a small-scale, project-based activity) into a standardized, institution-wide workflow. We believe that our experiences will be instructive to other libraries and that Penn State's documentation will assist others in making their own batchloading policies and procedures more efficient. This paper discusses the management of ad hoc batchloading; ongoing regular MARC record loads, such as PromptCat or Marcive, which at Penn State occur on a biweekly or monthly basis and are largely automated, fall outside the scope of the present discussion.

Survey of Literature on Batchloading Bibliographic Records into the Online Catalog

The OCLC began working with libraries and other vendors in the 1980s to promote the shared cataloging of microform collections and to provide sets of bibliographic records for batchloading purposes.[1] Benefits to cataloging libraries would be free searching and setting of holdings symbols and complete sets of the bibliographic records that they create or enhance. Benefits to other libraries would be the ability to acquire entire sets of records for discrete collections of microform library resources.

Several projects to catalog collections for the OCLC Major Microforms effort have been documented. Myers described the University of Southern Mississippi's project to create records for the Slavery Pamphlets Collection and indicated that a major consideration in support of the project was the anticipated high use of the collection after title-level access would be available in the catalog.[2] Toombs addressed the St. Louis University project to catalog the Nineteenth-Century Legal Treatises Microfiche Collection, noting that the project added many unique titles to the OCLC catalog.[3] Participation by St. Louis University in cooperative cataloging programs such as the Library of Congress Name Authority Cooperative Program (NACO) and OCLC Enhance has benefited all other libraries who use the records subsequently.

Jones described the development of microforms cataloging projects to create record sets to provide to libraries as well as efforts at Florida State University to batchload records for OCLC Major Microforms sets into their NOTIS online catalog.[4] He reported that OCLC provided record customization options for record sets, including the addition of a call number; however, that feature could be improved by increasing the detail added to the call number. Nevertheless, he found that the addition of records to the online catalog greatly increased the use of microform resources. Dodd described Virginia Tech University's experiences with batchloading record sets for microform collections into the Virginia Tech Library System.[5] She described the need for flexibility and discussion and highlighted the need for cooperation between the cataloging unit and the automation department. Banerjee reported on Oregon State University's experiences batchloading records for two major microforms sets into their online catalog.[6] He stressed the need to analyze record quality before loading and suggested limited criteria for record review and analysis.

He also recommended allowing time for problem resolution and clean-up after the records are loaded.

Martin described the challenges associated with the cataloging of eBooks, including the source of cataloging records, the potential for batchloading, the question of whether holdings for print and electronic should be on the same record, edits that might be needed before record loading, ongoing maintenance, and adding holdings for eBooks to OCLC.[7] She also addressed the increased use associated with eBooks records' availability in online catalogs, citing a number of other studies that indicate that the cataloging of eBooks increases use dramatically, in one case as much as 755 percent. Many of the issues identified and concerns expressed in these articles still exist for libraries today, whether loading records for microform or electronic resources.

Background of Batchloading at Penn State

In 2001, in response to a large number of requests from subject specialists that bibliographic record sets be loaded into the online catalog (the CAT), Penn State's assistant dean for technical and access services convened a working group charged with overseeing the batchloading process (see appendix for the change to this group). The Bibload Working Group (Penn State's integrated library system, SirsiDynix's Unicorn, requires the use of a report called "bibload" for batchloading bibliographic records into the catalog) meets monthly and includes representatives from Cataloging and Metadata Services, Public Services, the Commonwealth Campus Libraries (representing twenty-two Penn State campuses located throughout the state), and the Department for Information Technologies. Originally chaired by the assistant dean for Technical Services, the Bibload Group

was subsequently chaired by the head of Cataloging and Metadata Services, and is now led by the cataloging and metadata specialist, whose position description was rewritten in 2005 to include primary responsibility for managing the batchloading workflow. The responsibilities of the group's members and chair have been documented and are made available to potential members before they agree to serve so that they have a clear understanding of what work and time commitment is expected of them (four hours per week for members, up to thirty-two hours per week for the chair). Managing the batchloading process requires a solid grounding not only in traditional cataloging and the fundamentals of bibliographic description, but also in the technical aspects of data management and systems analysis. Also essential is a grasp of how users search for and discover resources in an online and increasingly networked environment.

Since 2001, the group has overseen the loading of more than half a million records into the CAT. Given that Technical Services at Penn State manually adds between fifty thousand and sixty thousand records to the catalog in an average year, batchloading, measured in terms of quantity, has doubled the productivity of the Technical Services Division. Fourteen percent of the records in the online catalog were batchloaded since 2001.

Policy Issues

In the development of any new workflow, libraries encounter issues that may require extensive discussion resulting in policy decisions. Those decisions that affect access, the quality of the database, or workflow that crosses organizational boundaries require broad input and are best made with consensus. The batchloading workflow has been no exception, and a number of questions have arisen during the development of this workflow at

Penn State. They include issues such as record quality versus access; single versus multiple records for materials held in print, microform, or electronic formats; what protocols or standards will be established to record decisions; which level of staff can do what work; whether the records should be purchased or simply downloaded from OCLC; and who will make these and related decisions.

Record Quality versus Access

Balancing record quality and improvement to access remains one of the biggest challenges in the batchloading process. Ideally, all records loaded into the catalog should conform fully to national and local standards. In practice, this is impossible. Few records sets are perfect and, in cases where the records are felt to be substandard in ways that might seriously affect the library's services or workflows, a decision must be reached about whether to load the files and, if so, how much record modification should occur prior to loading.

Also in question is the completeness of some record sets. Banerjee noted in 2001 that a record set purchased from the OCLC appeared to be missing "as many as 500 records—over eight percent of the entire collection" and Penn State recently encountered a similar situation.[8] Such experiences demonstrate that loading large record sets cannot ensure accurate coverage of collections to the same extent that on-site, title-by-title cataloging can. In some cases missing records likely go unnoticed for years, meaning that collections thought to be fully described in the catalog are not. Without committing resources to painstaking and time-consuming post–load quality checks, avoiding such oversights is nearly impossible.

Penn State's policy is to favor access over record quality. If the "greater good" is served by loading the records into the online catalog, then they are

loaded. However, as will be described later, much effort goes into improving the records through the use of MarcEdit software. Penn State's policy is to consult subject specialists during the decision to load the records and during the record enhancement stage.

Format Duplication, Multiple versus Single Records

The practice of maintaining a single bibliographic record for multiple versions of a given resource is common, even though such practice has, at various times, conflicted with national cataloging standards. Under such a policy, often grounded in a library's belief that users prefer to see holdings in multiple formats on the same record, a single catalog record might describe not only a printed book, but the microform reproduction and a digital version available online.

Both batchloading and the availability of many e-resources from multiple sources have made this policy increasingly difficult to justify or maintain. While standard numerical fields in bibliographic records such as the ISBN, ISSN, or Library of Congress classification number allow a certain degree of record matching, in the absence of unique and universally recognized record identifiers, most integrated library systems are simply unable to prevent duplication with 100 percent efficiency. Because effective de-duplication is not feasible, loading multiple records for different versions of a resource and sometimes for the same resource supplied by different vendors becomes necessary. In addition, the relatively recent availability of e-journal link resolver services such as ExLibris's SFX, many of which require the monthly loading of records that duplicate records already in a library's catalog, has made record duplication commonplace.

On a positive note, keeping each load separate facilitates the batch removal of items should the library

cease subscription to a given collection. It also makes possible setting better and more accurate holdings in the OCLC, thus facilitating the interlibrary loan process and potentially setting the stage for network-level resource discovery services, such as WorldCat Local.

Record Keeping and Documentation of Practices

The batchloading process is inherently complex, involving staff from throughout the organization and sizable amounts of technical data. Detailed record keeping is essential, both as a means of keeping stakeholders informed and of documenting practices so that complex procedures and solutions need not be devised and reformulated repeatedly. Such record keeping will improve the chances for success of a process that is so heavily distributed throughout the organization. The Bibload Group's website (www.libraries.psu.edu/tas/cataloging/dept/bibloads/bibload.htm) describes the group's charge, lists group members, and provides links to documentation. Detailed minutes of monthly Bibload Group meetings are taken by the chair, circulated for comment and correction, and then posted to the page. Technical details about each load, such as file size, are included, as are text versions of each file as well as the raw MARC files. Comprehensive records of report load specifications and load reports generated by the system (which include error logs) for all test and production loads accompany each file. Finally, Microsoft Word documents outlining the analysis of each loaded file along with changes made to the files prior to load are archived on the same page.

Staffing Levels

Experience at Penn State quickly demonstrated that management of the batchloading workflow was best done by a central group, with one person responsible for coordinating the many pieces of the puzzle. Excellent project management skills, the ability to follow through, and a high level of diplomacy are necessary to coordinate a fairly complicated workflow that has many stakeholders with competing priorities. Because this activity has become such a large and ongoing responsibility and includes providing direction to both librarians and staff throughout the libraries, a high-level professional staff position was created from an already existing position and given the responsibility for managing and coordinating the entire workflow.

Batchloading has also resulted in a significant amount of post–load work, including the correction of records that did not load appropriately, cataloging of titles that were missing from the files or simply did not load, and authorities cleanup. Much of this work can be assigned to a lower-level staff member in Cataloging and Metadata Services, but since the problems resulting from different batchloading projects can vary from one project to another, they generally require some direction from the Bibload manager. As each load is completed, the cleanup required is identified by the manager, who drafts procedures to help the staff member assigned to make the corrections. Cataloging knowledge is useful for resolving many of the problems encountered, so post–load projects are usually assigned to an experienced copy cataloger.

Purchasing Record Sets versus Downloading from the OCLC

In some cases the question of whether to purchase records as a set from a vendor or to download on a title-by-title basis from the OCLC is a simple one. If the records are provided as a proprietary service from a vendor, they may not be available in the OCLC; in such cases, the only way to provide access to those materials is to acquire the records from the vendor. If the set of records is so large as to be unwieldy or impossible to handle on a title-by-title basis, the decision to purchase as a set is similarly obvious. At Penn State, this cutoff point is set at one hundred records. If a collection has more than one hundred titles and records available, we will purchase the records as long as funds are available to do so. We have found that batchloading projects involving fewer than one hundred titles—which, like larger loads, still require group input, test loads, and systems office resources— are not worth pursuing through the normal batchloading process. In these cases, assuming records are available in the OCLC, we have chosen to catalog titles individually rather than batchloading the records.

Making Decisions and Getting Input from the Right People

Because anyone who consults a library's catalog is potentially affected by batchloading, identifying and communicating with stakeholders is critical. At Penn State, the Bibload Group includes two members from public service units, but they cannot, nor are they expected to, speak for all of their colleagues. Large records sets have been loaded for materials in many different disciplines, including engineering, social sciences, statistical data, history, literature, medicine, and law. Interested parties in the libraries are invited to review records and to provide input at each step of the process for any given load. In especially significant loads, Penn State's Collection Development Council, charged with coordinating acquisition of materials for the libraries, may be consulted. Batchloading cannot meet everyone's needs perfectly, but broadening the pool from which feedback is solicited both lessens the possibility of errors and heightens awareness of the importance of batchloading throughout the organization. It is the Bibload Group's

policy to seek and consider input from all stakeholders; this policy is codified in procedural documents that the group follows for each batchloading project.

Workflow

The batchloading workflow can vary from project to project. This section describes the typical workflow of a batchloading project, providing examples from Penn State's experiences.

Identification of Available Files

While the OCLC has, for many years, offered MARC records for electronic and microform sets through its WorldCat Collection Sets service (www.oclc.org/worldcatsets/default .htm), an increasing number of vendors of electronic and microform collections are making MARC record sets available for the collections they sell. Records are also available from commercial cataloging firms such as Cassidy Cataloguing Services, based in Rockaway, New Jersey, which sells packages of Westlaw, Lexis, and HeinOnline records targeted at law libraries. A fundamental challenge of batchloading records therefore is keeping abreast of record availability. Subject selectors may not be in the habit of querying vendors about record sets, and records may become available for collections acquired many years earlier. The Bibload Group at Penn State has taken an increasingly proactive role in researching record availability both by encouraging selectors to consider record availability as an important aspect of any new purchase and by researching record availability for sets the libraries already own or license.

A batchloading project begins when either the Bibload Group or a subject specialist becomes aware of the availability of records for a collection that either has already been purchased

or for which purchase is pending. Before the advent of online databases, most such sets acquired described microform collections that the libraries already owned but for which only a single collection-level record was available in the catalog. More recently, most of the sets acquired describe the titles constituting electronic aggregate resources.

Acquisition of Files

Some files are made freely available on a vendor's website. Other files, while free, must be requested, and the vendor may make them available via either a website or FTP, or send them as e-mail attachments.

Purchasing sets of bibliographic records can be more complex, and Penn State has adopted two different models for the process. In some cases, Cataloging and Metadata Services allocate funds for the purchase, are invoiced directly, and must submit a purchase order through the libraries' Business Office. (Depending on the cost of the file, approval for the purchase from a single source may have to be secured from the university's Department of Purchases, a step that may delay the project and must be taken into account during the planning phase.) In other cases, record sets are purchased with the collections fund; such purchases are initiated by staff in the Serials and Acquisitions Department exactly like purchases of items for the collection.

Some vendors offer to modify records to suit local needs. For example, the American Antiquarian Society, which provides records for Early American Imprints, First Series, allows purchasers to select records for a particular version (microopaque, positive microfiche, or negative microfiche), select which MARC field to use for the call number (090, 099, or other), and indicate what the base call number should be. The OCLC provides a number of options for modifying

record sets for both electronic and microform collections, including editing 856 fields (used for access information for electronic resources), deleting fields on the basis of their MARC tag, adding call number fields, customizing call numbers by pulling information from more than one source (such as a series number), adding fields, and more. With the advent of the MarcEdit software (discussed later), Penn State performs all customizations on site rather than asking vendors to modify records prior to purchase.

Acquisition of files has implications for workflow, staffing, server storage space, and network security. File naming conventions must be adopted. Server space must be designated and permissions assigned to appropriate staff. Copies of files must be routinely created and stored in a location accessible to staff charged with manipulating and loading files.

Record Review and Evaluation

Whether purchased from the OCLC, supplied by a vendor, or acquired from a third-party source, bibliographic records intended for batchloading must be reviewed for quality. A preliminary check by the batchloading process manager determines whether the correct number of records has been delivered, whether the records describe the correct set of resources, and whether the records are in the format agreed upon (usually USMARC 21 using either MARC-8 or UTF-8 encoding). Discrepancies are reported promptly to the supplier and arrangements made for a new file to be provided.

Software can be useful to determine quickly whether a file meets validation rules, but human review by experienced catalogers and systems staff is considered essential. To facilitate such review, a file is converted from MARC to text format and made available to members of the Bibload Group and other stakeholders. All

group members are expected to review a given number of records (at Penn State, twenty-five) within an agreed-upon time frame (e.g., five working days) to determine whether the records meet local needs. After records are deemed acceptable by cataloging and systems staff, subject specialists may identify modifications intended to improve their usefulness to patrons, such as notes, links to online guides, or series fields. Using input from subject specialists and members of the group, the records are edited and prepared for load using a freeware software program called MarcEdit (http://oregon state.edu/~reeset/marcedit/html/index .php) developed by Terry Reese.

Record Modification

All record sets require some modification before being loaded into the catalog. For the Unicorn integrated library system at Penn State at least a 949 field (containing the call number, classification scheme, purchasing library, home location, item type, and flags to indicate circulation and permanence) must be added to each record. These elements are required by the CAT; if not supplied during batchloading, the information would have to be manually added to each record after the load.

Many sets require additional modification. Local notes are added to records for online resources to inform patrons that access to the resource is restricted to Penn State users. The address of the libraries' proxy server is pre-pended to URLs so that off-campus users can authenticate to reach licensed products. Additional series statements may be added to assist in the retrieval of records using a single search. Links to guides available online may be added. In some cases, substandard record quality may necessitate corrections or modifications, such as converting 650 fields with indicators 14 (subject headings drawn from a local, usually nonstandard, thesaurus instead of from the Library of Congress Subject Headings) to 653 uncontrolled keyword fields or batch correcting typographical errors. The Program for Cooperative Cataloging (PCC) Standing Committee on Automation has created a guide for use by vendors when creating sets of bibliographic records to accompany monograph aggregations.[9] In theory, this guide should help vendors and publishers create future products that are tailored to meet the needs of libraries. While our discussion with one vendor indicates some interest in conforming to national cataloging standards, our experience suggests that vendors may be slow to adopt practices that fully conform to current library standards for quality.

Modifying Records Using MarcEdit

MarcEdit has revolutionized the ways libraries can manage their MARC records. Until recently, libraries were dependent on local programmers or systems staff to modify large record sets. MarcEdit empowers library staff to do the work themselves quickly and effectively by providing a wide array of tools for manipulating files of MARC records: Fields may be added or deleted, global edits made, and data swapped from one field to another.[10] In addition, MarcEdit's implementation of regular expressions—known in the computing world as regexes, a concise and flexible means for identifying strings of text of interest, such as particular characters, words, or patterns of characters—allows more sophisticated manipulation of data, such as building call numbers from data in multiple fields or selectively removing fields when certain data elements are present. Editing files locally is generally more flexible and more cost effective than requesting record customization from vendors.

Developing Load Specifications

The SirsiDynix Unicorn integrated library system allows several options regarding the batchloading of bibliographic records. Of primary importance is specifying how the unique record-specific identifier (title control number) is to be built during the load: from a numerical field in each record (e.g., 001, 020, 035) or simply system-generated. The presence of unique record-specific identifiers is essential in allowing subsequent updating or overwriting of records. Also configurable is the load rule, which determines how new and duplicate records are handled. Finally, several parameters are set to specify how call numbers and copy information is generated during the load.

Test Loads and Evaluation

Before being loaded into the production catalog, each file is first loaded onto the libraries' test server for review. Experience has shown that subject specialists and public services librarians are more comfortable reviewing records in the CAT than as simple text files and that potential problems not readily apparent based on inspection of the MARC records in isolation often become obvious in the context of the catalog. Furthermore, a test load is crucial for verifying that call number, library, location, and circulation status data has been configured and loaded correctly. Finally, a test load also serves to determine how many, if any, records will be returned as duplicates and to evaluate what action should be taken to address such duplication.

After the file is loaded into the test server, an e-mail message is sent to the Bibload Group and other stakeholders informing them of the availability of the records for review in the test CAT. The message includes information about the size of the file, the number of error records (i.e., records returned as having failed to load), and instructions for retrieving the records in the catalog. Bibload

Group members and other interested parties are requested to review the records within five working days and to send comments or questions to the group.

Production Load

If, following the test load, stakeholders voice concerns that require modifications to the records, a second test load may be undertaken to address the concerns raised. After approval of the final test load, files are loaded into production using the same report specifications as the test load.

An e-mail message is sent to the Bibload Group and other stakeholders informing them of the availability of the records for review in the CAT. Although in principle the production load should have results identical to the approved final test load, this review of the production load is undertaken by the Bibload Group and stakeholders in the interest of quality control to ensure that no unanticipated effects have occurred.

Off-Campus Access

Access to purchased electronic resources is almost always limited to users affiliated with the purchasing institution. Many vendors use IP filtering to manage access, so, for example, authorized Penn State users attempting to access content from off campus (i.e., from non–Penn State IP addresses) find themselves blocked. To ensure access to all authorized Penn State users regardless of their physical location, the Bibload Working Group began modifying vendor-supplied URLs by pre-pending the address for the libraries' proxy server. On-campus users who click on the link are taken seamlessly to the resource, while off-campus users, if they have not already authenticated as PSU users, are required to log in with their Penn State access accounts, and are then passed through to the resource.

Promotion

Making the libraries' community aware of the newly loaded records is seen as a critical step in the batch-loading process. When the Bibload Working Group was first formed, little or no promotion was undertaken. The subject specialist most closely interested in the load was informed that the records were available in the CAT, but no formal announcement was made to the libraries or the campus as a whole. Subject specialists were expected to make their constituents aware of the newly loaded records.

In an effort to educate colleagues about the progress made in providing access to hitherto hidden collections and to promote the work of the Bibload Group, global e-mail announcements are now sent to the entire Penn State Libraries community following each significant load. The announcements, drafted by the chair of the group in collaboration with the subject specialist, include a brief description of the collection's scope and importance as well as instructions for retrieving the records in the CAT. Such announcements not only provide information that allows the libraries' staff to provide better service to users, they also heighten awareness of the importance of batchloading and give credit to the members of the Bibload Working Group.

Vendor-Supplied Authority Control

Like many large academic libraries, Penn State sends records to an external vendor for authority control on a monthly basis. Large batchloading projects, especially those likely to create a sizable number of unmatched headings, are reported to the authorities librarian before the load takes place. In cases where series headings are added to files for the purpose of retrieval, series authority records are established in the Library of Congress Authority File (LCAF) prior to the

production load of the file to ensure that records containing the new series are not returned as part of the unmatched headings report.

Managing Catalog Extracts

Many large record sets purchased from vendors may not, because of contractual obligations, be supplied to the OCLC as part of the libraries' monthly holdings load. As a result, any ineligible records must be removed from the file before it is supplied to the OCLC. A file of unique record identifiers is generated and archived for every file that is batchloaded at Penn State. These files are used by systems staff to remove ineligible records prior to sending extract files to the OCLC and can also serve as a means for batch deleting large record sets in cases where the libraries cancel access to e-resources and must therefore remove records from the catalog. At Penn State the need to batch delete a batchloaded file has not yet arisen, but a similar procedure is used monthly to remove and then reload updated versions of Ex Libris's SFX records.

Post–Load Cleanup

Although one or more test loads can minimize errors, given the size and scope of most batchloading projects, which often involve tens of thousands of records, some post–load manual cleanup is inevitable. Records may fail to load, call numbers may load incorrectly, and the bibliographic records may have problems that are difficult or impossible to correct using MarcEdit. During the test phase the Bibload Group, in consultation with stakeholders, may decide that a certain percentage of errors is acceptable if correcting them after the load is easier or quicker than repeatedly modifying load specifications. When such a decision is made, a document is drafted by the Bibload Group chair outlining the nature and extent of the anticipated cleanup

required. Depending on the resources required, one or more staff may be assigned to work on the project.

Exposure to Risk and URL Management

Unlike physical collections, e-resources are often hosted remotely on vendor or third-party servers over which libraries have no control. When these servers fail or when URLs change, large numbers of e-resources suddenly may become inaccessible. The presence of title-level records in the online catalog heightens the effect of such technological glitches. Two approaches for managing such risk are routinely checking URLs and creating backup copies of remotely hosted resources. Link-checking software, while useful for systematically verifying that URLs in the library catalog are functioning properly, usually generates reports that library staff must review and process manually—a time-consuming procedure. Some vendors, such as Gale/Cengage Learning, supply archival copies in XML format of digital content to libraries so that, in the event that the vendor's server becomes inaccessible, client libraries will be able to ensure access to the content from their own servers. Although this approach is sound in theory, it requires libraries to create and maintain a server infrastructure capable of providing seamless access to e-resources normally hosted off site. For many libraries, such a strategy may be impractical. Penn State has begun preliminary discussions for managing archival content on local servers but has not yet implemented any policies or procedures for doing so.

Managing Ongoing Loads

Some batchloaded files must be supplemented by updates. NetLibrary, for example, regularly adds titles to its collection, as does the American Council of Learned Societies (ACLS)

Humanities E-Book Project. In other cases, vendors do not supply update files but instead provide new releases of entire record sets. In either scenario, provisions must be made for regularly acquiring and loading files and for ensuring that duplication is avoided. Managing ongoing loads can be especially challenging when vendors release updates irregularly, when updates are so small as to render the batchloading process less than ideally efficient, and when record quality is inconsistent, as was recently the case for the ACLS Humanities E-Book Project. Early batches of records treated the project name (History E-Book Project) as a series statement, while subsequent installments treated the project name as a corporate body (History E-Book Project, which later became the ACLS Humanities E-Book (Organization)). Files had to be edited to remove the inconsistency.

What the Future Holds

The biggest challenges of managing batchloading projects are technological and organizational. Validating large record sets, de-duplicating files to prevent duplicate records in the catalog, verifying that URLs function as intended, and ensuring seamless access to remotely hosted content in the event of server outages or other technological failures depend on software and hardware that continuously must be updated and maintained. MarcEdit, perhaps the most powerful software tool in the batchload toolkit, is in continuous development. Future users of the software may have access to even more powerful tools for validating, editing, and converting bibliographic data.

What effect the implementation of the entity-relationship model of metadata recommended in IFLA's *Functional Requirements for Bibliographic Records* and its application through Resource Description

and Access (the successor to the Anglo-American Cataloguing Rules) will have on catalog records and on the structure of the catalogs themselves remains to be seen.[11] Batchloading, which is largely based on the single flat record concept underlying current cataloging standards, will necessarily evolve as bibliographic databases are reconceptualized and restructured to better reflect the current landscape of information discovery and retrieval.

Because batchloading requires expertise in a broad array of library areas (acquisitions, cataloging, systems administration, public service), staff skills must evolve to meet this challenge. Cross-training, efficient models of communication, and up-to-date, concise, accessible documentation of policies and procedures will all be essential elements of the batchloading workflow of the future.

Conclusions

Batchloading is a complex process, both technologically and organizationally, requiring the coordination of resources from throughout a library. The experiences and processes developed at Penn State can help other institutions make more informed decisions and devise policies and procedures most likely to ensure a successful batchloading workflow.

Given the number of variables and the rapidly changing technological landscape, no single batchloading project fully exemplifies the process. Each load is different, requiring that all stakeholders be responsive to new opportunities and new challenges. Large gains in efficiency can be achieved by standardizing workflows and by carefully documenting procedures, but the process must be flexible enough to accommodate variations in the parameters, such as the size and quality of record sets, their cost, the likelihood that access to resources will become available through channels

other than the library catalog, and rapidly changing user expectations.

The goal of batchloading is improved access to the libraries' collections. Every item or resource to which the libraries provide access should be represented in the catalog. Loading large bibliographic files is an especially effective means of working toward this goal, and is much more efficient than traditional piece-by-piece cataloging.

Batchloading also allows improving the granularity of the catalog. Traditionally, online catalogs have described a library's holdings at the item level (for books and monograph-like items in other formats) or at the collection level (for large microform collections, electronic resource aggregator databases, serial publications, and archives and manuscript collections). As user expectations change and full-text databases become increasingly common, batchloading allows for greater granularity—providing title-level access for collections for which only collection-level access was available previously and providing analytical access to items for which only title-level access was available. Batchloading improves what

might be called the resolution of the catalog. Once a magnifying glass that allowed users to see a certain level of detail of the collections, the catalog can be transformed over time into a powerful microscope allowing a more magnified and therefore more detailed examination of an institution's rich collections.

References

1. Jan Nelson Lucas, "OCLC's Major Microforms Project," *Microform Review* 13, no. 4 (1984): 232–33.
2. Florence Myers, "Cataloging the Slavery Pamphlets Collection: An OCLC Major Microforms Project," *Microform & Imaging Review* 27, no. 2 (1998): 43–45.
3. William W. Toombs, "Nineteenth-Century Legal Treatises Microfiche Collection: A Major Microforms Cataloging Project," *Show-Me Libraries* 39 (1988): 26–28.
4. James F. Jones, "Online Catalog Access to the Titles in Major Microform Sets," *Advances in Library Automation and Networking* 3 (1989): 123–44.
5. Janet Dodd, "Integrated Endeavors: Cooperative Efforts in Selection and Implementation of Tape Loads for Major Microforms Set," *Microform Review* 24, no. 2 (1995): 58–60.
6. Kyle Banerjee, "Taking Advantage of Outsourcing Options: Using Purchased Record Sets to Maximize Cataloging Effectiveness," *Cataloging & Classification Quarterly* 32, no. 1 (2001): 55–64.
7. Kristin E. Martin, "Cataloging eBooks: an Overview of Issues and Challenges," *Against the Grain* 19, no. 1 (2007): 45–47.
8. Banerjee, "Taking Advantage of Outsourcing Options," 56.
9. Program for Cooperative Cataloging, "MARC Record Guide for Monograph Aggregator Vendors" (Washington, D.C.: Program for Cooperative Cataloging, 2006), www.loc.gov/catdir/pcc/sca/FinalVendorGuide.pdf (accessed Aug. 8, 2008).
10. Terry Reese. "Information Professionals Stay Free in the MarcEdit Metadata Suite," *Computers in Libraries* 24, no. 8 (2004): 24–28.
11. IFLA Study Group on the Functional Requirements for Bibliographic Records, *Functional Requirements for Bibliographic Records: Final Report*, UBCIM Publication—new Series v. 19 (Munich: G. K. Sauer, 1998), www.ifla.org/VII/s13/frbr/frbr.pdf (accessed Aug. 30, 2008).

Appendix. Bibload Working Group Charge

To manage the purchase, testing, and loading of sets of bibliographic records. Tasks will include:

- Confirm funding source.
- Complete record profile and deliver order to acquisitions staff or Business Office, as appropriate.
- Upon delivery, review record quality.
- Seek input from subject specialists regarding call number or other desirable edits to the bibliographic records.
- Customize records to suit subject specialists' needs.
- Prepare load specifications, consulting with subject specialists or library heads as appropriate.
- Run bibload report in test/development catalog, repeating as necessary.
- Work with Digital Library Technologies staff to run bibload report in production catalog.
- Inform the library community about availability of the records in the CAT.