Notes on Operations

Cataloging E-Books and Vendor Records

A Case Study at the University of Illinois at Chicago

By Kristin E. Martin and Kavita Mundle

E-books have become a substantial part of many academic library collections. Catalog records for each e-book title enhance discovery by library users, but cataloging individual books may be impossible when large packages are purchased. Increasingly, libraries are relying on outside sources for their e-book catalog records, which may come from vendors or third-party record services and are frequently included in the price of a subscription. Rather than handling individual items, catalogers find themselves managing and manipulating large sets of catalog records. While dealing with the records in batch is the only practical way to provide access to the large sets, batch processing does bring about a new set of challenges. This paper will explore the challenges of managing Machine-Readable Cataloging (MARC) records for the Springer e-book collection at the University of Illinois at Chicago University Library. It discusses tools and methods to improve record quality while working in a consortial setting. It provides lessons learned, continuing challenges of working with vendor records, and some steps that might help other libraries expedite the process of getting vendor records into the catalog.

Kristin E. Martin (kmarti@uic.edu) is Metadata Librarian and **Kavita Mundle** (kavita@uic.edu) is Assistant Catalog Librarian, both at the University of Illinois at Chicago Library.

Libraries continue to increase acquisition of e-books. A recent Ebrary survey of 552 libraries of all types around the world found that 88 percent of respondents owned or subscribed to e-books, with 45 percent further stating they provided access to more than ten thousand e-books.[1] Integrating records for the e-books into the catalog has been an important part of facilitating retrieval and access for this growing set of resources. Ever since vendors began offering MARC records for e-book collections as part of their subscription contracts, libraries have been adding them to their catalogs, typically via batch loads. This seemingly straightforward process has brought new challenges for catalogers in terms of the load itself, sharing records in a consortial environment, quality and completeness of record content, and access problems. Despite these challenges, batch loading vendor records is an expedient solution for libraries wishing to provide timely catalog access to e-book collections.

The University of Illinois at Chicago (UIC) Library has been batch loading vendor-supplied Machine-Readable Cataloging (MARC) records for various e-book collections into its local catalog, UICCAT, since 2006. When this paper was first presented, the library had loaded more than 250,000 vendor-supplied e-book records in the library's catalog. Loading records for e-book collections has been challenging because of the varying quality of different vendors' record sets. The UIC Library's ongoing communications with vendors has produced better records

but at the price of effort and time on the part of the UIC Library Catalog Department. The following case study describes the challenges of managing MARC records for the Springer e-book collection the UIC Library received via a consortial purchase of Center for Library Initiatives (CLI), a collaborative program of the Committee of Institutional Cooperation (CIC), which is a consortium of the Big Ten universities plus the University of Chicago. This study describes UIC Library's and the CLI's efforts to improve the quality of these records and may be useful to other libraries grappling with how to manage and improve vendor records for e-book collections.

Literature Review

This review explores the literature describing how academic libraries are providing access to their large e-book collections through their catalogs, focusing on issues surrounding the addition of vendor-supplied e-book bibliographic records, the batch loading of large record sets, and new developments affecting e-book cataloging. Finding an effective way to provide access to increasing numbers of e-books is an ongoing challenge for libraries. Many academic libraries make their e-book collections searchable through Web-based catalogs or websites. McCall found that nineteen of twenty-one health science libraries provided title-level access to e-books through their websites, and twenty created bibliographic records in their catalogs.[2] Dinkelman and Stacy-Bates's review of Association of Research Libraries (ARL) member libraries showed that 56 percent of ARL library websites had separate pages devoted to e-books, and only 30 percent of library catalogs allowed limiting the search to e-books in a single step.[3] The study recommended that the library websites should include unambiguous descriptions of the contents of e-books, group

e-books by subject, and alert patrons to search for e-book titles in the catalog. Additionally, to make searching less difficult, the study recommended adding a one-step limit option to search e-books in the catalog. A 2008 analysis by Hutton explored how catalogs and websites provide access to open digital collections through library catalogs and websites.[4] When Hutton searched for ten e-book titles from open digital collections in the online catalogs and websites of ten academic libraries, only three of the e-book titles were found in the library catalogs, and none were found on library webpages.

Research by Dillon, Gibbons, and Langston demonstrated that adding bibliographic records to the catalog of university libraries assists users in discovery of and access to e-books and increases their use.[5] Green's study at the University of Surrey found that adding e-book records and URLs to the catalog makes identifying e-book titles easier for students.[6] Research by Armstrong and Lonsdale, Ramirez and Gyeszly, and Snowhill also determined that adding bibliographic records increases e-book use.[7] Conversely, Cox reported that a survey on e-books at an Irish university consortium found that 60 percent of the surveyed population accessed e-books through library websites and only 19 percent accessed them through the library catalog.[8] However, most of the studies described here reinforce the concept of the catalog as an important source of information for e-book discovery and access.

Although the integration of e-books with library catalogs and websites or with federated discovery tools offers greater visibility and enhanced access to libraries' e-book collections, cataloging techniques and guidelines are still evolving. Martin suggested that librarians have been slow and reluctant to catalog e-books because e-book cataloging guidelines have been underdeveloped.[9] The literature on e-books provides few examples of effective ways to catalog e-books individually,

an exception being Bothmann's article, which examined functional aspects of cataloging an electronic book.[10] Belanger's survey of library catalogs of thirty higher education libraries in the United Kingdom provided a snapshot of e-book cataloging practices.[11] The study noted variations between surveyed libraries in terms of which e-books were selected to be cataloged and how were they cataloged. Twenty-eight of the thirty libraries surveyed provided some kind of title-level access in their catalog for at least some of their subscription-based collections, four also provided records for free e-books, and two had not cataloged any e-books at the time of survey. Twenty-three of the twenty-eight libraries reported using separate records for print and e-books, and five reported providing access on the same record as print titles. A check of their catalogs revealed that many libraries used both methods when cataloging e-books.

The challenges of managing batch loading of vendor records and controlling record quality appear as important themes in library literature. Several authors identified problems in government document record loads, foreign vendor records in WorldCat, and microform sets. These problems included incorrect choice or form of headings that affect authority control, missing call numbers, missing or duplicate records, typographical errors, and MARC coding errors.[12] Mugridge and Edmunds described their experiences of batch loading vendor record sets to improve access to electronic and microform collections at Penn State University Libraries.[13] They noted that the process involves collaborations between various library units for purchasing, reviewing records, assessing record quality to meet local customization needs, and coordinating record loads and their ongoing maintenance. They observed that as more print materials get digitized and the availability of record sets for large collections become widespread, batch

loading of vendor records will become a significant workflow for libraries.

The literature documents that many challenges affecting vendor records and batch loads for traditional resources also exist for e-books. These issues have come to the forefront for e-books, however, because of the emphasis libraries are now placing on electronic collections. Nelson and O'Neil examined vendor records for e-books and reported that although the initial evaluation of sample vendor records showed high quality, record-load problems existed and included generating unwanted printed orders, causing load-program crashes, and failing to load subject headings.[14] Gedeon and Meyer concluded that vendor records lacked quality and authority control, which created considerable delays in loading records.[15] Sanchez, Fatout, and Howser described in detail how a library identified problems in NetLibrary records and the tools the library used to manipulate the records before batch loading them into the catalog.[16] Mundle described the experiences of adding vendor records at the UIC Library into the catalog.[17] She proposed guidelines for accepting vendor records and described how some e-book collections were cataloged when vendor-supplied catalog records were unavailable. Two 2009 papers provide additional examples of how libraries modify and improve record quality when handling batches of records for electronic resources. Beall detailed the problems and limitations contained within the freely available Mbook records and the steps taken after the records were loaded into the catalog to improve record fullness and access points.[18] Finn described a method to speed authority control processing by using MarcEdit and an external vendor to provide authority control prior to loading the records into the catalog.[19] These studies indicate that the batch loading process requires libraries to make decisions that balance the desire to quickly provide access through the catalog and the desire to provide the highest quality of records possible.

New Guidelines and Directions in E-Book Cataloging

During the past few years, organizations have developed guidelines to assist both libraries and vendors in record creation for e-books. In 2006, the Program for Cooperative Cataloging (PCC) developed "The MARC Record Guide for Monograph Aggregator Vendors," which includes recommendations to vendors on how to create MARC records for e-books.[20] The guidelines originally preferred an "electronic reproduction" model, following Library of Congress Rule Interpretation 1.11A, which allows all of the original data of the e-book to remain in the standard MARC fields (such as the 300 note for physical description and 260 field for the publication, distribution, etc.) while information specific to the reproduction is listed separately in a 533 reproduction note.[21] This model works well for preserving information about the book, but led to the creation of multiple records for the same title, each distinguishable only by the reproduction note and URL. Any time a title was offered by more than one vendor, a new record was created for each version, putting duplicate records for the same title in OCLC and in local catalogs. In 2009, to avoid duplication and to bring all equivalent manifestations of the same title under one generic record, the PCC approved the *Provider-Neutral E-Monograph MARC Record Guide*.[22] It brings e-book description guidelines more in line with those for e-journals, which have been described using provider-neutral guidelines since July 2003.[23] At the same time, the PCC revised the *MARC Record Guide for Monograph Aggregator Vendors*.[24] All vendors are encouraged to use this revised guide when creating or modifying records for their titles. The revisions instruct vendors to create a single bibliographic record to cover all equivalent manifestations of an e-monograph. The *MARC Record Guide* emphasizes recording only fields that apply to all electronic manifestations by removing provider-specific information given in notes or as added entries, eliminating a provider-specific reproduction note in MARC field 533, and clarifying how to record print and e-book International Standard Bibliographic Numbers (ISBNs).

In all, these guidelines should provide for a cleaner and less confusing display of e-book records in the library catalog but will radically change how libraries have to manage their e-book records. Following the previous cataloging guidelines that preferred the reproduction model, catalogers could quickly identify records for a particular vendor set through an added entry for the vendor name, the vendor record number, or other identifier consistent across all records. Working at the record level meant that if content changed, whole records could be deleted—typically a simpler and more automated process than trying to identify and delete portions of a record. The new guidelines mean that library systems need to provide or develop facilities to make batch changes to records, as additions and deletions will be based on fields within records or new holdings rather than entire bibliographic records. The new guidelines also will affect record loads. The same content coming from multiple vendors should have records that are identical, with the exception of the URL. This should encourage more sharing of records by allowing multiple vendors to use the same record for each manifestation of an e-book, rather than each having to create their own record. A unique identifier, such as an OCLC number, across records from different vendors can be used as a match

point to merge records with additional vendor URLs and prevent duplicate records for the same title. In January 2010, OCLC began to implement the provider-neutral guidelines by cleaning up existing e-book records to conform to the new guidelines.[25]

Effect on WorldCat Local and E-Book MARC Subscription Services

During the past two years, some libraries have moved toward relying on WorldCat as a discovery tool and local catalog through the WorldCat Local service. This may accelerate with the new WorldCat Local Quick Start service offered by OCLC.[26] Currently, records for e-books received from vendors frequently do not contain OCLC numbers or have the OCLC number for the record describing the print version. If WorldCat becomes a library's catalog, having holdings attached to the correct record will be the way users know their library has access to the content. Even if libraries do not use WorldCat Local, the open worldcat.org on the Web that allows users to find library content through Web searches, having holdings attached to the correct record is important. A service such as the OCLC e-serials holdings service might be needed for e-books. One can hope that, as more e-content is available simultaneously with or even before the print version and as the new provider-neutral record guidelines are implemented, libraries will have easy ways to set proper holdings in WorldCat, regardless of whether the records were obtained directly from OCLC or through batch-loading sets of vendors records.

E-book cataloging has lagged behind e-journal cataloging in the availability of subscription services for MARC records. These services, such as Serials Solutions 360 MARC and ExLibris MARCit! provide a way for libraries to rely on third-party vendors to keep track of their e-journal holdings for them.[27] Relying on holdings

data provided through the library subscriptions, these services provide records for all e-journals to which a library subscribes with a link contained within each bibliographic record that takes the user to the aggregated electronic holdings for that journal. Such services have proven especially beneficial for e-journals, where a library may receive the same title through multiple databases, and where aggregator databases are constantly adjusting content by switching titles and changing dates of coverage. As the number of e-books increases, and as libraries both purchase and subscribe to large sets of e-books, the potential for overlap of titles between databases and the need to keep track of multiple holdings increases. Recently Serials Solutions expanded its services to e-books, offering both holdings coverage and record service for e-books.[28] Using an e-book record service would help libraries both track and maintain access to their e-book collections. In addition, it would simplify the transition to the provider-neutral record model by having only one record for content offered through multiple providers, instead of libraries having to coordinate matches and overlays of records for the same content from different vendors. Using a record service for e-books involves a trade-off: increased simplicity for the library, which now only has to manage records from a single source, versus a potential loss of quality since the most complete records may not be available through the service. As the number of both e-books available and the vendors offering them continue to increase, e-book record services will mature and may become increasingly attractive.

Integration of Springer E-books at the UIC Library

The case study presented in this paper describes how the UIC Library, working with the CIC CLI (hereafter CLI), evaluated and improved vendor-supplied

MARC records for Springer e-books and then worked with the Consortium of Academic and Research Libraries of Illinois (CARLI) to load those records into the UIC Library catalog. UIC is an urban land-grant university with an enrollment of more than twenty-five thousand students. The UIC Library, a member of the ARL, has three locations in Chicago: the Richard J. Daley Library, the Science Library, and the Library for the Health Sciences, which includes three branch libraries in Urbana, Rockford, and Peoria. The combined collection consists of more than 2.2 million volumes, 50,000 e-journal titles, and more than 250,000 e-books. The UIC Library is a member of the CLI and CARLI, which has 153 member libraries in the state of Illinois. Seventy-six CARLI members participate in I-Share, the consortial union catalog. The UIC Library works through these consortia to explore and realize opportunities for collaboration and shared acquisition of electronic resources.

As a part of expanding consortial adoption of e-book content, the CLI, in partnership with Ingram Digital and Springer Science and Business Media, purchased access to all English-language e-books from Springer from 2005 through 2010, providing access to approximately twelve thousand titles in 2008 and ongoing access to new titles as published.[29] The CLI's agreement with Springer and Ingram expanded availability of e-book content to all CLI member libraries and gave affiliated users access to the set of Springer e-books hosted on two platforms. The same titles can be accessed both on the publisher's site, SpringerLink, and on MyiLibrary, Ingram's e-content aggregation platform.

CARLI and I-Share Consortial Environment

Although CARLI was not directly involved in this particular purchase of e-books or the subsequent review of

the MARC records, an understanding of the CARLI environment is important because it influences how the UIC Library manages vendor records. CARLI maintains separate databases for each of the (currently) seventy-six CARLI Institutions, including the UIC Library, which participate in the consortial union catalog, I-Share. The content of these databases is then combined and loaded into I-Share. Both the local catalogs and I-Share use the Voyager integrated library system (ILS), an Ex Libris product. Although CARLI manages the technical infrastructure, member libraries are responsible for the content of the records in their individual databases, including performing any authority control work. All additions, changes, and deletions to each library's records are fed into the I-Share database through a nightly updating process. CARLI uses match points and ranking algorithms to add and delete holdings to existing bibliographic records and to determine if existing bibliographic records should be overlaid or new records should be added to the I-Share database. CARLI's preferred match point is the OCLC number or other unique vendor number.

The central CARLI office is responsible for batch loading bibliographic records into member libraries' catalogs, which are then automatically loaded into the I-Share database through the nightly updating process. CARLI handles cataloging updates from all seventy-six member libraries. As more libraries within I-Share request batch load, the limitations on the throughput of records can create delays, which can range from only a few days for smaller loads to weeks or even months for larger loads. Additionally, the current Voyager software does not offer inherent batch record modification options that would allow for large-scale modifications within Voyager. Thus batches of records that might be extracted and modified with third-party software and then reloaded into

the local database and I-Share become caught in the same cycle of load limitations. In most circumstances, records need to be accessed individually to be modified or updated.

Strategies for Record Review

Following the CLI purchase of the Springer content, Ingram offered to provide record sets for all the titles included under the agreement, which provided titles with imprint years from 2008 through 2010 and backlist titles from 2005 through 2007. Records could be downloaded from Ingram's FTP server and loaded into the CLI libraries' catalogs. Early samples of the records from Ingram indicated potential overlay problems, issues with data quality, incorrect use of MARC fields, broken or missing links, and confusion over content coverage. Some CLI libraries developed alternate methods to obtain and load records into their catalogs. Another group of CLI libraries, including the UIC Library, decided to pool their resources and work together to improve the records.

Performing a systematic review of the Ingram-provided record set was important for three reasons. First, the group of CLI libraries wanted to avoid duplication of effort, which would occur if each library individually cleaned up problems in the records. Second, the libraries were interested in establishing a pattern of working with vendors to encourage them to provide usable records as part of their e-book access service. Third, the libraries wanted to have the records as load-ready as possible—records that would require little additional adjustment or manipulation. Additionally, the UIC Library's throughput issues in the CARLI/Voyager environment encouraged the library to make all changes to the records before they were loaded. Because the records would be loaded into the I-Share consortial catalog, the UIC Library wanted to avoid having improperly formed

or ambiguous identifying numbers in the record because these could cause the records to overlay existing records for different titles, thus creating inappropriate holdings for other libraries in the I-Share catalog.

While all the libraries were examining their options for getting records for the Springer titles, the head of the Catalog Department at the UIC Library tried to coordinate efforts. She led a joint effort of the group of CLI libraries to evaluate small samples of records during the early months of 2008, noting the problems and communicating with Ingram through several iterations. As the process evolved, a staff member from the CLI central office became the primary point of contact to represent the interests of the group of CLI libraries. In October 2008, the CLI central office created an electronic discussion list specifically to coordinate evaluation efforts between the group of CLI libraries, with the intention that this list would continue to be a forum to discuss MARC records for consortial purchases in the future. At this point, the head of the Catalog Department at the UIC Library delegated responsibility for record evaluation and communication of issues to the authors of this paper, who represented the UIC Library on the discussion list. The authors performed a systematic review of the sample records sets made available by Ingram (see "Analysis of the Records" below) and shared the results with the representative of the CLI member libraries. Other libraries tested the records and shared their comments and questions through the discussion list and a conference call coordinated by the CLI representative. Review occurred as new record sets were made available, with new problems reported and addressed as they became apparent. Finally, in December 2008 the CLI received a replacement set of records to load, which was complete with all of the records for the Springer titles that Ingram had created to date. After

working with CARLI, the UIC Library was able to have the records loaded in March 2009.

Challenges and Opportunities in Consortial Record Review

While some aspects of batch loading records are unique to the UIC Library's circumstances, others may be applicable to libraries working together with a vendor to obtain sets of records. The CLI libraries that chose to work together were able to leverage the power of the consortium's unified voice to encourage Ingram to address problems and improve record quality. The CLI discussion list provided a forum for the libraries to share their questions and concerns about the records and streamlined communications with Ingram by providing a single point of contact through the CLI representative. Lacking in the process, however, was a concerted effort to agree on the most important changes or to rank changes in order of importance; problems were simply communicated to Ingram as they were identified.

The consortial review process developed and evolved in response both to the needs and requests of the member libraries and to responses from the vendor. Early in the process, the CLI libraries discussed forming a core working group to channel information to other participating libraries and possibly having one library evaluate, modify, and load the records and then distribute the modified and tested set to the other libraries. The libraries' goals had to be scaled back because of personnel and organizational changes that delayed vendor responses. As a result, the libraries ended up testing and evaluating the records simultaneously and sharing evaluation results with each other through the discussion list. These findings were ultimately combined and shared with Ingram. Because Ingram provided several record sets that had to be reviewed, the timeline for the complete record

set was pushed back. Consequently, libraries that were under pressure to quickly load records did not meet their goals. As time went on, more libraries made independent decisions about the modifications needed and how best to represent the Springer titles in their respective catalogs. Some chose to look for records outside of Ingram; others continued to work together with Ingram to enhance the records.

Analysis of the Records

The authors, both librarians with expertise in e-resources cataloging, performed a systematic review of the records and identified a number of issues using simple measures. MarcEdit (http://people.oregonstate.edu/~reeset/marcedit), developed by Terry Reese at Washington State University and regularly used by the UIC Library, was an effective tool for reviewing and revising batch records. Analyzing the records required an understanding of MARC standards and current national and local e-book cataloging guidelines. The UIC Library's database specialist who manages batch loads worked with CARLI's database management staff to troubleshoot any load problems.

The MarcEdit field-count tool provided a list of all MARC tags in use in the record set. The authors examined the field list for any unusual MARC tags or strange numbers. Once the unusual fields or anomalies in the records were identified, their contents could be exported to a tab-delimited file for further examination in a spreadsheet. By manipulating and sorting the information from the records and viewing it in tabular form, the authors were able to identify several types of problems. Simple sorting identified likely problem records. For example, sorts on the date of publication in the 008 field highlighted records with an unusual publication date. Sorting by title identified cases where the

same record was used for more than one e-book (a problem with individual volumes in multivolume sets). The authors reported all problems found in the Springer record sets to the CLI discussion list.

Types of Problems Found

The problems in the records can be classed into three categories: access issues, load issues, and record-quality issues. Access issues prevented users from accessing the e-book, load issues prevented the records from being loaded into the catalog, and record-quality issues hampered users' ability to locate the records in the catalog or presented confusing information. The access and load issues were the most pressing problems because Ingram needed to address them before the records could be loaded and content could be accessed. Record-quality issues were more enduring and difficult to resolve. Although some problems were specific to the Springer e-book records, many can be applied generally to vendor e-book records and, even more broadly, to vendor records. Appendix A summarizes the problems discovered by the UIC Library and the actions taken by both the library and Ingram to correct them.

The first category of problems, access issues, prevented users from using the catalog record to access the resource. The most obvious access issue was missing or broken links, which were identified in many records by the UIC Library and other CLI libraries. All records were supposed to contain two links: one to the title on the SpringerLink platform and one to the title on the MyiLibrary platform. Review of the 856 MARC field (Electronic Location and Access) revealed that a number of records linked solely to MyiLibrary and others that had an incomplete SpringerLink link. Some records also contained broken MyiLibrary links. Without a good way to systematically check links,

the authors were able to find these only through spot checks. Fortunately, other CLI libraries reviewing the records also identified these problems. Ingram quickly corrected its own link problems and worked with Springer to correct the rest.

Another serious access issue involved German-language titles and demonstrates the challenge of maintaining control over the content of e-book packages. Although the original contract with Ingram was for only the English-language collections, the record sets provided by Ingram contained records for German-language titles as well. The URLs to the MyiLibrary platform provided access to the German-language titles, but links to the SpringerLink platform denied access to the content. Although access to the German-language titles was not supposed to be part of the original contract, Ingram assured the CLI libraries that once the content was available through MyiLibrary, the libraries were entitled to access it through the length of the contract. Access to the content would not be provided on the SpringerLink platform. The UIC Library sorted the records using the language code in the 008 (Fixed-Length Data Elements) field and removed the SpringerLink information from the German-language records, a service that Ingram later provided.

A third important access issue relates to obtaining records for older content. The agreement with Springer provided access to back volumes of book series from 1997 through 2005, but that content was available only on SpringerLink and not MyiLibrary. Ingram did not provide records for those volumes and has no plans to do so.

Load issues affected the UIC Library's ability to have CARLI load the records into UICCAT and I-Share. The most important issue that needed to be addressed was the proper formulation of the 001 and 003 MARC

fields (Control Number and Control Number Identifier, respectively). Within the Voyager system, the 001 and 003 MARC fields combine to create a 035 system number, which is used as a match point for record overlays upon load. An improperly formed or duplicate 035 may create load problems by immediately overlaying a record upon load or by being overlaid in the future with another record, a particular problem in the I-Share union catalog, which maintains holdings for seventy-six libraries. Holdings for some libraries would be attached to the wrong record and other resources would no longer have records in the catalog.

Other match problems hampered loading. Initially, many of the records contained OCLC numbers for print versions of the books, both in the 035 and 019 fields, a potential overlay hazard, particularly in the I-Share catalog. While the OCLC number can serve as an important record identifier, if the records were derived from print or other sources, the OCLC number would not be correct for the records representing electronic resources. Additionally, ISBNs recorded in the records proved confusing and would lead to overlay problems should libraries try to match on ISBN. The print ISBN might be repeated in both the 020 ISBN field and the 776 $z (Additional Physical Form Entry) field, or it might exist in only one of the fields. Ingram also assigned a unique number for the version of the e-book on MyiLibrary, and this was initially formulated as an ISBN and placed in the 020. Ingram moved it to the 024 (Other Standard Identifier) field, but the UIC Library decided to strip out all 024s from the record because this served as a match point for the UIC Library's WorldCat Cataloging Partners record delivery service and also could have created overlay issues. Ingram removed this number from later batches of records. Some confusion regarding print and e-book ISBNs remains, though this is

not unique to the Springer e-books. Finally, problems with character encoding caused selected records to fail to load. Ingram addressed most of these issues, but Voyager still rejected a small handful of records, mostly in German-language titles.

Record-quality issues do not present record-loading problems or directly prevent access, but they do inhibit access by confusing the user or making the records difficult to retrieve in the catalog. Some improvements in record quality were simple, such as removing a number of extraneous and no longer relevant fields and notes in the records that were revealed through the MARC field count tool in MarcEdit. These included the 852 (Location) and 049 (Local Holdings) fields for local notes, 300 subfield e (Physical Description) notes for accompanying CD-ROMs, and 530 (Additional Physical Form) notes with "Also issued online." The UIC Library removed the extraneous and misleading fields from the records and requested that Ingram provide that service in the future.

The authors also discovered more serious quality issues with some records, e.g., the "see from" reference forms were used as access points instead of the authorized forms of headings and sets were treated inconsistently. After loading the initial batch of records, authority reports identified many records using the "see from" reference form of access points. The UIC Library corrected name, series, and subject headings on the basis of the Voyager authority reports but has not yet systematically reviewed unauthorized forms of names, titles, and subjects that were in the records. These changes added to the time that catalogers devoted to clean-up work after the batch load of the Springer records.

The second major issue in record quality hampers user access to multivolume sets. For example, Springer publishes the multivolume set *The Handbook of Environmental Chemistry*, and within each numbered

volume is a set of lettered and individually titled volumes. For example, volume 5, *Water Pollution,* has several individually lettered and titled volumes. Libraries handling this title in print would have the choice of analyzing each volume, attaching the volumes to a set record, or a combination of both. Although the new provider-neutral guidelines recommend a single record for an entire set, handling each volume for multi-volume e-books individually frequently proves to be the most practical method for the vendor. Unfortunately, treatment of individual volumes has been inconsistent both on the e-book platforms and in the records provided by Ingram. For example, one record describing water pollution only provides a link to volume 5F, *Environmental Impact Assessment of Recycled Wastes on Surface and Ground Water*, and another only provides a link to volume 5P, *The Caspian Sea Environment*. These volumes are not described individually. This is confusing to the user, and identifying these problem sets without reviewing each individual record and comparing it to the e-book is difficult. The problem of describing and providing access to individual volumes in multivolume sets is not unique to e-book records. However, the problem is more obvious with e-books because accessing content directly from the 856 field creates a disconnect between the content of the work and its description in the record. Exporting record data to a spreadsheet and then looking for duplicate titles can make it easier to identify potential problem titles, but records still have to be examined manually.

Current Record Status

Despite setbacks, records from the initial (corrected) load of Springer titles and several major updates are now in the UIC Library's catalog and I-Share. Ingram resolved most of the problems and made many improvements to the

records suggested by the group of CLI libraries. Some issues remain and are worth noting because they may apply to records from other vendors. Problems of the use of "see from" references in names and inconsistent treatment of multi-volume sets are not being addressed at the vendor level, so clean-up work has been duplicated at the CLI libraries. Also, the record set provided by Ingram remains incomplete; it does not include records for the backfiles of the Springer book series. The UIC Library has not performed a complete reconciliation of records to titles that are supposed to be available through the two e-book platforms. The library must assume that it has received records for all titles to which it has access and that all records lead to the correct titles. Currently, the library reacts to problems identified but cannot work proactively. Not knowing the complete content of the twelve thousand titles in the e-book package makes performing a detailed comparison difficult.

Ingram is continuing to supply updates to the records to cover titles published since the initial load of records. Although the provider-neutral guidelines for e-books have become available, at present the authors are not aware of any Ingram plans to convert the Springer records to provider-neutral records. This may be in part because the contract between Ingram and Springer is set to expire at the end of 2010, and Springer will no longer offer content through MyiLibrary. At that time, Ingram will discontinue supplying MARC records. The UIC Library continues to load new records from Ingram for the Springer titles as they become available and is investigating other options for obtaining records for the Springer content.

Conclusion and Lessons Learned

The UIC Library's experiences with

vendor e-book records serve both as case study and cautionary tale. The methods used to find problems with records and to identify the types of problems by categories may be of use to other libraries working with vendor-supplied records. The authors believe that working with the CLI and the vendor to improve the records before receiving them was the most productive route to quality data in the catalog. Examination of the process reveals both the power and the difficulty of working consortially with a vendor to improve the records. Without the weight of the CLI, the records might not have been improved as much as they were, but coordinating communication between the different parties added a complication. Additionally, while information sharing between the CLI libraries helped improve the records, the consortial work did not include shared authority clean-up and other record modifications.

Library consortia could pursue additional collaborations and share work to provide quality records for their members. Within the group of CLI libraries working together, more complex options, such as funding one library to send the records to an external authority vendor for processing, were not discussed. Each library was left to make corrections and determine load procedures individually. These more complex options would have been difficult to pursue within the CLI group of libraries because each library maintains its own library catalog and policies regarding electronic resources cataloging. Moreover, the CLI libraries do not have a shared union catalog or even use the same ILS software, which leads to different load requirements for the libraries. Consortia that share an ILS or coordinate both purchases and cataloging policies may have more success in providing these additional services for their members and reducing the workload of individual libraries.

As one CLI member commented,

the "free" records ended up being quite expensive, given the initial work to improve the MARC file and the work done by each library to clean up access points and set records in the local systems. This situation is unlikely to change in the near future. Vendors are attempting to automate record creation as much as possible, and changes at the title-level are improbable. The key for efficiency for both libraries and vendors will be to create a high-quality description of each e-book that can be reused and repurposed by any number of libraries to create quality catalog records.

Although the CLI's role remained limited to facilitating communication, other library consortia may be able to take on an additional role to modify and enhance records centrally for all member libraries and could prevent duplication of work by individual member libraries. As an example, CARLI has purchased the Springer e-book collections and is providing an enhanced record set and customization options to members purchasing the collection. CARLI has created a new task force to update cataloging guidelines for e-resources and to determine how to implement the new provider-neutral record guidelines. Because all libraries that participate in I-Share use the same ILS software and share a union catalog, the opportunities for the consortium to provide work on behalf of its members may be greater than with the CLI libraries. E-book record subscription services also may be of assistance by providing enhanced records and deduplication of titles for libraries that subscribe to the service and serving as a single point of contact for record issues. The companies running these services have a business goal of supplying bibliographic data to libraries and can solicit participation and cooperation of all of the vendors supplying e-content. The UIC Library has subscribed to the e-books record service offered by Serials Solutions and will

be investigating options for Springer records from both CARLI and Serials Solutions in the future.

The Springer records evaluation process may serve as a model for record improvement for future consortial purchases both by the CLI and other library consortia. Based on the UIC Library's experience, the following considerations may expedite the process of record evaluation and access to e-resources through the catalog. Libraries should develop record specifications before negotiations with suppliers begin and include them as a part of the negotiations to acquire e-content. This approach will demonstrate the importance of quality bibliographic data to the content providers. Libraries should request a large sample set or full set of records for evaluation. This will reduce the iterative process of evaluating records. Libraries and the vendor should define a mutually acceptable timeline for record improvements. A clear goal and common purpose will help keep the consortium members and the vendor moving the process forward. Within the consortium, libraries should commit to full participation to the extent possible and work with the consortium to share record improvements across all member libraries. Consortia that play a role in managing bibliographic data in addition to shared purchasing may have more resources to provide record enhancements to all member libraries.

The world of e-resources cataloging and the methods that users employ to find e-resources are evolving so rapidly that libraries may serve their users best by providing the best available access to resources (although it may be initially less than ideal) and then working to improve accuracy, completeness, and discoverability after access has been established. Libraries by necessity may need to cede some control over their data as they find more efficient ways to serve their users. Tools to assist in the evaluation,

clean-up, and enhancement of records in batch processes will become even more important. Collaboration with vendors and other libraries to integrate records for e-resources into library catalogs and newer discovery tools will continue to be valuable in making library services more efficient and effective for users.

References

1. Ebrary, Ebrary's Global eBook Survey (2007), www.ebrary.com/corp/collateral/en/Survey/ebrary_eBook_survey_2007.pdf (accessed Apr. 1, 2010).
2. Steven L. McCall, "Online Medical Books: Their Availability and Assessment of How Health Science Libraries Provide Access on their Public Websites," *Journal of the Medical Library Association* 94, no. 1 (Jan. 2006): 75–80.
3. Andrea Dinkelman and Kristine Stacy-Bates, "Accessing E-Books through Academic Library Web Sites," *College & Research Libraries* 68, no. 1 (Jan. 2007): 45–58.
4. Jane Hutton, "Academic Libraries as Digital Gateways: Linking Students to the Burgeoning Wealth of Open Online Collections," *Journal of Library Administration* 48, no. 3 (2008): 495–507.
5. Dennis Dillon, "E-Books: The University of Texas Experience: Part 2," *Library Hi-Tech*, 19, no. 4 (2001): 350–62; Susan Gibbons, "NetLibrary Ebook Usage at the University of Rochester Libraries. Version 2," (2001), www.lib.rochester.edu/main/ebooks/analysis.pdf (accessed Apr. 1, 2010); Marc Langston, "The California State University E-Book Pilot Project: Implications for the Co-Operative Collection Development," *Library Collections, Acquisitions & Technical Services* 27, no. 1 (2003): 19–32.
6. Kirsty Green, "Introducing E-Books at the University of Surrey," *SCONUL Newsletter* 29 (Summer/Autumn 2003): 54–56, www.sconul.ac.uk/publications/newsletter/29/14.PDF (accessed Apr. 1, 2010).
7. Chris Armstrong and Ray Lonsdale,

"Challenges in Managing E-Books Collections in UK Academic Libraries," *Library Collections, Acquisitions & Technical Services* 29, no. 1 (2005): 33–50; Diana Ramirez and Suzanne Gyeszly, "NetLibrary: A New Direction in Collection Development," *Collection Building* 20, no. 4 (2001): 154–64; Lucia Snowhill, "E-Books and Their Future in Academic Libraries: An Overview," *D-Lib Magazine* 7, no. 7/8 (Jul./Aug. 2001), www.dlib.org/dlib/july01/snowhill/07snowhill.html (accessed Apr. 1, 2010).

8. John Cox, "E-Books: Challenges and Opportunities," *D-Lib Magazine* 10, no. 10 (Oct. 2004), www.dlib.org/dlib/october04/cox/10cox.html (accessed Apr. 1, 2010).

9. Kristin E. Martin, "ATG Special Report—Cataloging Ebooks: An Overview of Issues and Challenges," *Against the Grain* 19, no. 1 (2007): 45–47.

10. Robert Bothmann, "Cataloging Electronic Books," *Library Resources & Technical Services* 48, no. 1 (Jan. 2004): 12–19.

11. Jacqueline Belanger, "Cataloging E-Books in UK Higher Education Libraries: Report of a Survey," *Program* 41, no. 3 (2007): 203–16.

12. Deborah E. Kulczak and Cathy Reineka, "Marcive GPO Records and Authority Control: An Evaluation of Name and Subject Headings at the University of Arkansas Libraries, *Cataloging & Classification Quarterly* 38, no. 1 (2004): 87–103; Laura D. Shendenhelm and Bartley D. Burke, "Book Vendor Records in the OCLC Database: Boon or Bane?" *Library Resources & Technical Services* 45, no. 1 (2001): 10–19; Kyle Banerjee, "Taking Advantage of Outsourcing Options: Using Purchased Record Sets to Maximize Cataloging Effectiveness," *Cataloging & Classification Quarterly* 32, no. 1 (2001): 55–64.

13. Rebecca L. Mugridge and Jeff Edmunds, "Using Batchloading to Improve Access to Electronic and Microform Collections," *Library Resources & Technical Services* 53, no. 1 (2009): 53–61.

14. Lesley Nelson and Frances O'Neil, "Electronic Monographs in the Academic Library: An Implementation Story," *LASIE* 32, no. 2/3 (Aug./Dec. 2001): 13–20.

15. Randle Gedeon and Bettina Meyer, "eBooks at Western Michigan University: A Case Study," *Against the Grain* 17, no. 1 (Feb. 2005): 52–54.

16. Elaine Sanchez, Leslie Fatout, and Aleene Howser, "Cleanup of NetLibrary Cataloging Records: A Methodical Front-End Process," *Technical Services Quarterly* 23, no. 4 (2006): 51–71.

17. Kavita Mundle, "Integration of Electronic Books into Library Catalogs: The UIC Library Experience," *Acquisitions Librarian* 19, no. 3/4 (2007): 389–407.

18. 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.

19. Mary Finn, "Batch Load Authority Control Clean-Up Using MarcEdit and LTI," *Technical Services Quarterly* 26, no. 1 (2009): 44–50.

20. Program for Cooperative Cataloging, "MARC Record Guide for Monograph Aggregator Vendors" (2006), www.loc.gov/catdir/pcc/vendorguiderevised.pdf (accessed Apr. 1, 2010).

21. Library of Congress, *Library of Congress Rule Interpretations*, 2nd ed., (Washington, D.C.: Library of Congress, 1989, updated 2002): Rule 1.11A.

22. Becky Culbertson, Yael Madelstam, and George Prager, *Provider-Neutral E-Monograph MARC Record Guide* (Washington, D.C.: Program for Cooperative Cataloging, 2009), www.loc.gov/catdir/pcc/bibco/PN-Guide.pdf (accessed Apr. 1, 2010).

23. *CONSER Cataloging Manual, Module 31*, 2002 ed. (Washington, D.C.: CONSER, 2002), www.loc.gov/acq/conser/guidelines.pdf (accessed Apr. 1 2010).

24. Program for Cooperative Cataloging, *MARC Record Guide for Monograph Aggregator Vendors*, 2nd ed., prepared by Becky Culbertson et al. (Washington, D.C.: Program for Cooperative Cataloging, 2009), www.loc.gov/catdir/pcc/sca/FinalVendorGuide.pdf (accessed Apr. 1, 2010).

25. Robert Bremer, "Provider-Neutral E-Monographs" (presented online for OCLC, December 10, 2009): slide 24, www.oclc.org/us/en/multimedia/2009/Provider_Neutral_Webinar.htm (accessed May 31, 2010).

26. OCLC, WorldCat Local "quick start," www.oclc.org/worldcatlocal/quickstart/default.htm (accessed Apr. 1, 2010).

27. Rebecca Kemp, "MARC Record Services: A Comparative Study of Library Practices and Perceptions," *Serials Librarian* 55, no. 3 (2008): 379–410.

28. SerialsSolutions, 360 MARC Updates, www.serialssolutions.com/360-marc-updates (accessed Apr. 1, 2010).

29. Committee on Institutional Cooperation, Consortial Licensing, www.cic.net/Home/Projects/Library/Licensing/Introduction.aspx (accessed Apr. 1, 2010).

Appendix. Challenges Presented by the Records

Access

Issues	Actions by the UIC Library	Actions by Ingram
Broken/missing links	Discovered through examination of Microsoft Excel spreadsheet and reported to Ingram	Fixed broken MyiLibrary links; added missing SpringerLink links supplied by Springer
German-language titles included in batch of records (access only available through MyiLibrary)	Divided file in MarcEdit based on 008 language code; removed SpringerLink link in German file; reported to Ingram	Divided full batch of records into German-language and non-German-language titles
Back issues of 1997–2005 book series available on SpringerLink, but not on MyiLibrary	Investigated alternative methods for obtaining catalog records	No records will be provided through Ingram

Load

Issues	Actions by the UIC Library	Actions by Ingram
Incorrectly formatted 001/003 could lead to accidental overlays	Reported to Ingram	Problem corrected with vendor code in 003 and 001
Incorrect OCLC numbers in 001 and 019	Removed from records and reported to Ingram	Removed OCLC numbers from 001 and 019; OCLC number in 035 may exist for future records
Multiple ISBNs in record: for e-book, print book, and Ingram European Article Number appearing in ISBN format	Reported to Ingram; removed any numbers in 024 field to prevent accidental overlay	Moved publisher-assigned ISBN to 024 and later removed altogether; some confusion over print/electronic ISBN remains
Improperly coded diacritics	Identified and corrected with MarcEdit; reported to Ingram	Most problems addressed; some problems remain with common Russian-language diacritics

Record Quality

Issues	Actions by the UIC Library	Actions by Ingram
Extraneous fields in records, including 049, misleading 530 notes, 300 fields with $e for CD-ROM, etc.	Identified through MARC field count in MarcEdit and removed; reported to Ingram	Unwanted fields removed from future batches of records
Invalid forms of names in records	Identified and corrected through authority programs in Voyager; reported to Ingram	No action taken
Inconsistent treatment of multivolume sets (record for set may be used multiple times to represent different volumes)	Problem reported to Ingram; also planning on identification of records through Excel spreadsheet and correction in catalog as needed	No action taken yet