Supporting Name Authority Control in XML Metadata: A Practical Approach at the University of Tennessee

Marielle Veve, University of North Florida

Available at: https://works.bepress.com/marielle_veve/7/
From Innovation to Transformation
Patrick L. Carr

Criticism of Cataloging Code Reform, as Seen in the Pages of Library Resources and Technical Services (1957–66)
Steven A. Knowlton

Comparing Catalogs
Stephen Hearn

Supporting Name Authority Control in XML Metadata
Marielle Veve

Using Batchloading to Improve Access to Electronic and Microform Collections
Rebecca L. Mugridge and Jeff Edmunds
Your gateway to online cataloging, classification, and metadata tools.

Cataloger’s Desktop
An integrated, online documentation system with the most important cataloging and metadata resources—available 24-7.

- Look up the AACR2 rule and quickly link to the rule’s LC Rule Interpretation (LCRI).
- Cataloging rules link to the right MARC 21 format instructions.
- Consult more than 250 cataloging and metadata resources.
- NEW—OCLC Bibliographic Formats and Standards now integrated into Cataloger’s Desktop.
- Find the rule quickly with Cataloger’s Desktop’s simple interface.

Free trial accounts & annual subscription prices:
Visit www.loc.gov/cds/desktop
For free trial, complete the order form at www.loc.gov/cds/desktop/OrderForm.html

Classification Web
Full-text display of all LC classification schedule & subject headings. Updated daily.

- Find LC/Dewey correlations—Match LC classification and subject headings to Dewey® classification numbers as found in LC cataloging records. Use in conjunction with OCLC’s WebDewey® service.
- Search and navigate across all LC classes or the complete LC subject headings.

Free trial accounts & annual subscription prices:
Visit www.loc.gov/cds/classweb
For free trial, complete the order form at www.loc.gov/cds/classweb/application.html

Dewey and WebDewey are registered trademarks of OCLC, Inc.

Free 30-Day Trials for Both Products!
Library Resources & Technical Services (ISSN 0024-2527) is published quarterly by the American Library Association, 50 E. Huron St., Chicago, IL 60611. It is the official publication of the Association for Library Collections & Technical Services, a division of the American Library Association. Subscription price: to members of the Association for Library Collections & Technical Services, $27.50 per year, included in the membership dues; to nonmembers, $85 per year in U.S., Canada, and Mexico, and $95 per year in other foreign countries. Single copies, $25. Periodical postage paid at Chicago, IL, and at additional mailing offices. POSTMASTER: Send address changes to Library Resources & Technical Services, 50 E. Huron St., Chicago, IL 60611. Business Manager: Charles Wilt, Executive Director, Association for Library Collections & Technical Services, a division of the American Library Association. Send manuscripts to the Editorial Office: Peggy Johnson, Editor, Library Resources & Technical Services, University of Minnesota Libraries, 490 Wilson Library, 309 19th Ave. So., Minneapolis, MN 55455. (612) 624-2312; fax: (612) 626-9353; e-mail: m-johnson@umn.edu. Advertising: ALCTS, 50 E. Huron St., Chicago, IL 60611; 312-280-5038; fax: 312-280-5032. ALA Production Services: Troy D. Linker, Chris Keech, Tim Clifford, and Justine Wells. Members: Address changes and inquiries should be sent to Membership Department—Library Resources & Technical Services, 50 E. Huron St., Chicago, IL 60611. Nonmember subscribers: Subscriptions, orders, changes of address, and inquiries should be sent to Library Resources & Technical Services, Subscription Department, American Library Association, 50 E. Huron St., Chicago, IL 60611. 1-800-545-2433; fax: (312) 944-2641; subscription@ala.org.

Library Resources & Technical Services is indexed in Library Literature, Library & Information Science Abstracts, Current Index to Journals in Education, Science Citation Index, and Information Science Abstracts. Contents are listed in CALL (Current American—Library Literature). Its reviews are included in Book Review Digest, Book Review Index, and Review of Reviews.

Instructions for authors appear on the Library Resources & Technical Services webpage at www.ala.org/alcts/lrts. Copies of books for review should be addressed to Edward Swanson, Book Review Editor, Library Resources & Technical Services, 1065 Portland Ave., Saint Paul, MN 55104; e-mail: swans152@umn.edu.

©2009 American Library Association

All materials in this journal subject to copyright by the American Library Association may be photocopied for the noncommercial purpose of scientific or educational advancement granted by Sections 107 and 108 of the Copyright Revision Act of 1976. For other reprinting, photocopying, or translating, address requests to the ALA Office of Rights and Permissions, 50 E. Huron St., Chicago, IL 60611.


Publication in Library Resources & Technical Services does not imply official endorsement by the Association for Library Collections & Technical Services nor by ALA, and the assumption of editorial responsibility is not to be construed as endorsement of the opinions expressed by the editor or individual contributors.

Library Resources & Technical Services
ISSN 0024-2527 January 2009 Volume 53, No. 1

Editorial

ARTICLES

From Innovation to Transformation
A Review of the 2006–7 Serials Literature
Patrick L. Carr

Criticism of Cataloging Code Reform, as Seen in the Pages of Library Resources and Technical Services (1957–66)
Steven A. Knowlton

Comparing Catalogs
Currency and Consistency of Controlled Headings
Stephen Hearn

NOTES ON OPERATIONS

Supporting Name Authority Control in XML Metadata: A Practical Approach at the University of Tennessee
Marielle Veve

Using Batchloading to Improve Access to Electronic and Microform Collections
Rebecca L. Mugridge and Jeff Edmunds

Book Review

Index to Advertisers

Cover art reprinted from Papillons: vingt planches en phototypie coloriées au patron donnant 81 papillons et 16 compositions décoratives, by E.A. Seguy. (192?), Plate 11, held at the Special Collections Research Center, North Carolina State University Libraries, Raleigh.

Association for Library Collections & Technical Services
Visit LRTS online at www.ala.org/alcts/lrts.
For current news and reports on ALCTS activities, see the ALCTS Newsletter Online at www.ala.org/alcts/alcts_news.
This is the first issue of volume 53 and the start of another exciting year for Library Resources and Technical Services (LRTS). Please join me in welcoming our new editorial board members (Allyson Carlyle, Lewis Brian Day, October Ivins, Edgar Jones, Randy Roeder, Carlen Ruschoff, and Sarah Simpson) and thanking the board members who completed their terms at the end of the 2008 ALA Annual Conference in Anaheim (Tscheri Harkness Connell, Karla L. Hahn, Sara C. Heitshu, Judy Jeng, Bonnie MacEwan, Carolynne Myall, Pat Riva, and Diane Vizine-Goetz). Editorial board members help set the direction of the journal and serve an essential role as paper referees in the double-blind review process. The quality of LRTS depends on their dedication and diligence. I’m honored that the Association for Library Collections and Technical Services (ALCTS) has reappointed me to serve an additional four years as LRTS editor. No one will deny that serving as editor of a peer-reviewed journal is challenging and hard work, but most editors will agree with me in also saying that it is interesting, informative, and (most of the time) fun. I am delighted that Edward Swanson has accepted reappointment as the LRTS book review editor. Do contact him directly (eswanson@eswanson.org) if you are interested in reviewing titles for LRTS.

This issue presents papers that cover the range of responsibilities that define the mission of ALCTS and its nearly five thousand members. Patrick L. Carr provides another installment in the familiar LRTS literature review series as he explores the themes and important works in the 2006–7 literature about serials librarianship. Steven A. Knowlton looks back at the history of cataloging codes and the often heated debates that characterized code reform in the 1950s and 1960s. His premise is that reviewing the debates of the past can prove useful as we engage in another spirited conversation about reforming the current cataloging code. Stephen Hearn suggests an alternative approach to gathering and analyzing catalog data, intended to serve as one possible measure of a technical services unit’s success in attaining its goals. Do spend some time studying the figures that accompany this article. They offer a new way to represent changes in headings over time. The final two papers in this issue are “Notes on Operations.” LRTS publishes papers in this section with the intent to offer innovative approaches to challenges faced in many libraries. Marielle Veve reports on a new solution developed at the University of Tennessee Libraries to support name authority control in Extensible Markup Language (XML) for digitized collections. Rebecca L. Mugridge and Jeff Edmunds share insights from the Penn State Libraries’ experience in developing processes to facilitate batchloading records into the online catalog. Much more than “how we did it good” stories, these papers present approaches that can inform practice in other libraries.

The success of LRTS depends on the quality of the papers published, and these papers are written by you! Consider the issues you have been pondering, the challenges you have been addressing, and the future of libraries and the profession of librarianship. Reflecting thoughtfully on these topics is the first step in writing a paper. Writing a paper enhances your knowledge and expertise. Why not write a paper and submit it to LRTS?
From Innovation to Transformation
A Review of the 2006–7 Serials Literature

By Patrick L. Carr

This paper reviews the leading trends in and contributions to the peer-reviewed and professional literature of serials librarianship published in 2006 and 2007. The review shows that a central topic in the literature is the nature and effect of libraries’ ongoing transition from acquiring serials in print to providing access electronically. Propelled forward by user preferences, this transition is reflected in publications that reconceptualize collections and describe innovative initiatives and strategies for acquisition, access, and management. Throughout the literature, the review traces a prevailing sentiment that libraries are advancing well beyond the confines of print-centered models and are assuming new roles, imagining new possibilities, and developing new solutions.

The literature of serials librarianship published in 2006 and 2007 reveals a field in rapid transition. The changes occurring range from the shifting nature of serial collections to evolving models, initiatives, and management strategies used to acquire and administer access to these collections. According to Plutchak, serials librarianship and scholarly communication as a whole are currently in a period of innovation in which emerging technologies are ceasing their emulation of the past and revealing extraordinary new possibilities.1 Plutchak believes that this period will culminate in the transformation of scholarly communication so that technology “overturns the capabilities that were previously thought to be the pinnacle, and brand new ways of doing things become possible.”2 From this perspective, the 2006–7 serials literature might be said to offer a first, nascent glimpse of the landscape stretching before libraries as they pioneer their way from a period of innovation to one of transformation. Indeed, there is a prevailing sentiment in the literature that libraries have advanced well beyond the confines of print-centered models in their strategies for acquiring and administering serial access. The literature shows libraries assuming new roles, imagining new possibilities, and developing new solutions.

This paper, the latest entry in LRTS’ ongoing series reviewing the serials literature, starts where Genereux’s review of the 2004–5 literature left off.3 It examines the peer-reviewed and professional literature of serials librarianship published in 2006 and 2007. The primary resource for identifying publications to include in the review was Library Literature and Information Science. In addition, citations in publication reference lists, postings on electronic discussion lists, and serendipitous discovery all contributed to forming the body of literature that was examined. Within this body of literature, the criteria for selecting publications to review was based on the author’s judgment of which publications most fully exemplify the leading trends in and contributions to serials librarianship’s literature.
The first section of the review, “Collections and Concepts,” takes a broad perspective, surveying the forces that the literature indicates are reshaping the nature of serials in libraries. Specifically, it reviews changes in the use, formats, and cost of serials and analyzes the effect of these changes on how serials are defined. The next section, “Acquisition,” considers the literature’s discussion of the evolving means through which serial access is acquired. In addition to assessing the current state of publisher packages, it gives particular attention to the effect of the open access (OA) movement and acquisition models that shift emphasis from ownership to access. The third section, “Access,” examines publications describing libraries’ three primary serial access points: online catalogs, link resolvers, and metasearch engines. The fourth section, “Management,” reviews the literature’s discussion of how the managers of serial collections are responding to new challenges and opportunities. It focuses on how these managers can successfully communicate, achieve change, and improve workflows and organizational structures. The final section of the review, “Initiatives,” describes what the literature indicates to be the leading efforts to develop initiatives resulting in the enhanced acquisition, administration, evaluation, and archiving of serials.

Given the far-reaching scope of the serials literature, this review cannot be comprehensive. Among the excluded topics are citation analyses, publishing costs, marketing, the storage and retention of print serials, institutional repositories, and the OA movement’s effect on the publishing industry and scholarly communication. In addition, this review is restricted to literature written in English and places an emphasis on publications geared toward librarians in the United States, Canada, and the United Kingdom.

Collections and Concepts

A central topic in the 2006–7 literature is the nature and effect of libraries’ ongoing transition from acquiring serials in print to providing access electronically. This transition is being propelled forward by user preferences and is manifesting itself in evolving collection formats, costs, and concepts of seriality.

Use Studies

As Johnson and Luther conclude from their interviews with twenty-four librarians and publishers, user preferences are among the primary forces reshaping serial collections.4 Use studies published in 2006 and 2007 show preferences for e-serials among a variety of communities. A representative study is Brady, McCord, and Galbraith’s analysis of the 2003 print and e-serial use of researchers at Washington State University’s Owen Science and Engineering Library.5 Comparing the results of their analysis with a previous study conducted at the same site, the authors discovered that use of the library’s serial collection in electronic formats increased from 71 percent of total use in 2001 to 94 percent of total use in 2003. The authors believe their findings show a “cultural shift” in user preferences.6 Rowlands’s review of e-serial use studies published in the professional literature offers further evidence for users’ preferences for accessing serials electronically.7 One of the author’s key findings is, “Where implemented, electronic versions of journals have displaced print use dramatically and at a much faster rate than many anticipated.”8

Voorbij and Ongering discuss reasons for users’ preferences for e-serials in their survey of Danish faculty conducted in 2003 and 2004.9 The authors found that the most cited reasons for using e-serials over their print counterparts are e-serials’ enhanced functionalities (e.g., the ability to perform full-text searches and use hyperlinks within articles) and increased accessibility. In their survey of the academic staff within the Consortium of Academic Libraries of Catalonia, Borrego and colleagues provide a picture of e-serial use as it relates to users’ discipline and age.10 Use was highest among researchers in biomedicine, engineering, and the exact and natural sciences, who use e-serials either primarily or entirely, and lowest among researchers in the social sciences and humanities, who primarily use print serials. The authors also learned that e-serial use is prevalent among researchers under the age of forty, while most researchers over the age of fifty-one persist in accessing serials in print.

Format

Libraries have responded to users’ preferences by transitioning to e-serials. Prabha documents this in an analysis of the formats in which members of the Association of Research Libraries (ARL) subscribe to a sample of 515 serials.11 From 2002 to 2006, ARL libraries’ print subscriptions to the sample serials dwindled by 32 percent while electronic subscriptions grew by 34 percent. Prabha’s research also shows that the period from 2005 to 2006 was a watershed in which, for the first time, electronic subscriptions to the sample serials surpassed print subscriptions. Hahn gives further evidence for the shift to e-serials in a 2005 survey assessing the participation of eighty-nine ARL libraries in serial packages offered by five large publishers: Blackwell, Elsevier, Springer, Taylor and Francis, and Wiley.12 Of the packages that respondents indicated they were participating in for 2006, 58 percent involved the cancellation of print versions of the serials within the packages. This fact leads Hahn to conclude that libraries are swiftly moving to electronic-only formats for serials within publisher packages. Drawing
on their interviews with librarians and publishers, Johnson and Luther predict that this trend will continue: “Although the pace and likely ultimate extent of the transition differs from institution to institution, all are moving along a continuum from print-only to dual-media to e-only journals.”13 In the near future, they speculate, it is possible that all but 5 percent of many libraries’ serial collections will only be accessible electronically.

Redefining Serials

Changes in the formats of serial collections have introduced deeper questions regarding the nature of seriality. In Soule’s review of the evolving definitions that libraries have applied to serials over the past fifty years, the author comments that a challenge libraries will face in their future efforts to define a serial is the “increasing fragmentation of information” in a digital world.14 Soules contemplates whether this fragmentation might someday manifest itself in a decision by publishers to abandon efforts to organize serials into units such as volumes and issues and instead make articles accessible electronically as they are ready for publication. Van Orsdel foresees a similar disaggregation, commenting that libraries are experiencing “a seeming shift of interest to the piece rather than the container, the article rather than the journal, the definition rather than the dictionary.”15 In Plutchak’s view, the outcome of this shift is that “the serial as defined by the librarian is an anachronism in the digital age, and will not survive for long.”16 The author argues that, in the current period of transition, the attempt to clearly define a serial is futile. While acknowledging that, at present, the article remains prevalent, Plutchak anticipates that data sets and social networking tools have a revolutionary potential.

Cost

The evolving nature of serials has resulted in complex changes in the size and average unit cost of library collections. An ARL report shows that, following fifteen years of stagnation, the number of serials purchased by member libraries skyrocketed by approximately 64 percent from 2001 to 2005.17 The report further indicates that the average unit cost of a subscription has decreased by approximately 23 percent from 2000 to 2005. Explaining the factors behind these trends, Kyrillidou points to libraries’ dual-format subscriptions (e.g., a print plus online subscription), which, according to ARL guidelines, should be counted twice.18 Other contributing factors cited by the author include consortial arrangements and libraries’ transitions to online-only subscriptions, which are sometimes less costly than subscriptions in other formats.

Libraries’ expenditures further reflect the transition to e-serials. ARL statistics suggest that, for the period from 1994–95 to 2004–5, member libraries’ e-serials expenditures have ballooned by over 1,600 percent.19 Libraries’ overall serials expenditures have also experienced rapid increases. Since 1986, for example, ARL libraries’ serials expenditures have increased by 302 percent, a rate of growth that significantly exceeds increases in the annual consumer price index over the same period.

Rising subscription costs is one of the primary factors affecting these complex changes in collection sizes, average unit costs, and expenditures. Reviewing the costs of serials listed in three databases produced by the Institute for Scientific Information as well as EBSCO’s Academic Search Premier database, Van Orsdel and Born estimate that academic libraries in the United States experienced 2007 subscription cost increases of 9 percent for domestic serials and 7.3 percent for foreign serials.20 The authors predict that 2008 subscription costs will increase by an additional 7–9 percent. White and Creaser provide added documentation of the inflating costs of subscriptions.21 Examining data that Swets Information Services provided for the subscription costs of eight commercial publishers and three university presses, the authors calculate overall price inflation of approximately 39 percent between 2000 and 2006. Moghaddan further contributes to the literature’s discussion of pricing through a comparison of the 2003 subscription costs of serials from five commercial publishers and five nonprofit publishers.22 Among the author’s findings is that the average subscription cost of the commercial publishers’ serials exceeded the average subscription cost of the nonprofit publishers’ serials by approximately 280 percent.

Acquisition

As a result of rising subscription costs, predictions regarding the sustainability of established acquisition models can be dire. Van Orsdel, for example, warns that “library budgets are, and will continue to be, no match for journal price inflation or for the cost of new journals as they appear.”23 The author suggests that a key component to overcoming this crisis is developments in the marketplace that foster competition and elasticity. The 2006–7 literature discusses both established acquisition models and their alternatives.

Publisher Packages

The literature shows that the bundling of serials into publisher packages continues to be a prevalent acquisition model. Huhn documents this prevalence in a 2005 survey assessing the participation of eighty-nine ARL libraries in serial packages offered by five large publishers: Blackwell, Elsevier, Springer, Taylor and Francis, and Wiley.24 Most
respondents (93 percent) subscribed to at least one of the publishers’ packages, and, on average, respondents subscribed to packages offered by three of the publishers. The two most cited reasons for participation in packages were that “content and access offered were a good return on the investment” and “alternative non-bundled forms of access to the content were prohibitively expensive.” Together, these responses lead Hahn to speculate that libraries’ participation in packages indicates that they “may be making the best of a bad situation.” The survey further shows that fifty respondents have had one or more cancellation projects for the subscription years 2004–6, and 66 percent of these fifty respondents have protected packages from cancellation. Hahn notes that the implication of this is that any parts of the respondents’ collections have suffered more significant cuts. Ultimately, the author argues that the survey’s results demonstrate that publishers should offer packages with terms and pricing structures that are more accommodating to the needs of libraries.

The OA Movement

The OA movement, which aims to make research freely available online, constitutes a central effort to transform scholarly communication. Although the body of literature discussing and debating the OA movement extends outside the boundaries of serials librarianship, several noteworthy publications examine a topic directly affecting libraries’ serial acquisitions: the correlation between the growth ofOA and library subscriptions.

From the results of a survey of 340 librarians, Ware concludes that, for the time being, libraries do not generally consider the availability of OA content to warrant the cancellation of subscriptions. Among the factors leading to this conclusion are that librarians do not see OA content as an acceptable or reliable substitute for a subscription. Likewise, librarians possess neither an awareness of nor plans to analyze the overlap between subscribed and OA content. However, Ware also found that 81 percent of respondents believe the availability of OA content would be “very important” or “important” in forming cancellation decisions. Moreover, while 32 percent of respondents assured publishers that they should not be worried about cancellations, 54 percent felt that it was too soon to make such a determination. Beckett and Inger’s subsequent survey of 424 librarians portrays the OA movement as a greater threat to the continuation of libraries’ subscriptions. Approximately 40 percent of the survey’s respondents indicated that they feel it is wasteful for a library to subscribe to serials with content that is freely accessible online. Citing findings such as these, Beckett and Inger conclude that “a significant number of librarians are likely to substitute OA materials for subscribed resources, given certain levels of reliability, peer review, and currency of the information available.”

In an editorial appearing in Learned Publishing, Anderson echoes the sentiments expressed in the findings of Beckett and Inger. He comments that “it is highly likely that rational individuals and libraries will cancel subscriptions to journals whose content is immediately, freely, easily, and reliably available at no charge.” Some commentators, however, foresee the coexistence of subscriptions and the availability of OA content. Pinfield, for example, examines four possible scenarios for the future of scholarly communication and concludes that subscriptions and the OA movement can be viewed as complimentary models rather than competitors. For coexistence to occur, Pinfield believes that a number of major changes need to be instituted by both OA repository administrators and publishers. These changes include widespread deployment of repository infrastructure, development of version identification standards, development of value-added features, new business models, [and] new approaches to quality control and adoption of digital preservation as a repository function.

Acquisition and Ownership

The OA movement is not the only threat to established acquisition models. As Anderson states, “The arguments for traditional collection development are losing their strength with every passing day.” Competing with these traditional arguments are models focused on acquisition of access without ownership. Carroll and Brink describe a project at the University of New Hampshire (UNH) Library that exemplifies this trend. Beginning in August 2003, UNH opted to meet users’ growing access needs through a document delivery service rather than the initiation of new subscriptions. The authors deem the project a successful strategy for reducing expenditures and comment that UNH hopes to cancel little-used and high-cost subscriptions and instead provide access to these serials through a document delivery service.

Offering further evidence of libraries’ exploration of nontraditional acquisition models are articles that have been written to assess the full-text access that aggregated databases provide to serials in specific disciplines. Together, these articles suggest a growing interest in leasing content through aggregated databases (which typically do not ensure perpetual access) rather than owning the content through a subscription with perpetual access provisions. Stemper and Barribeau document the trend toward acquiring access without ownership in an article that received the 2007 Best of LRTS Award. The authors’ literature review and informal survey suggests that more than 80 percent of
research libraries will enter into an agreement regardless of whether the agreement ensures that the access acquired is perpetual.

In an article that received the 2007 Blackwell Scholarship Award, Atkinson asserts that this willingness to acquire access without ownership represents "the greatest single failure of research libraries in the past decade." Several publications advocating that libraries secure perpetual access rights reflect this perspective. In their analysis of fifty serial and aggregator license agreements entered into by the University of Minnesota, Stemper and Barribeau found that a majority of these agreements (64 percent) include provisions for perpetual access. Although these provisions often included loopholes, vague wording, and specifications of additional fees, the authors nevertheless deem their findings heartening. However, they temper their optimism by emphasizing that publishers' willingness to grant perpetual access rights is only of value if libraries pursue these rights.

Kenney and colleagues further stress the importance of securing perpetual access. Drawing on interviews in which they assess archiving concerns voiced by fifteen library directors, the authors analyze twelve archiving programs. The conclusions derived from this analysis convey a sense of urgency. Kenny and colleagues state that current license agreements are inadequate to protect a library's long-term interest in electronic journals, that individual libraries cannot address the preservation needs of e-journals on their own, that much scholarly e-literature is not covered by archiving arrangements, and that while e-journal archiving programs are becoming available, no comprehensive solution has emerged and large parts of the e-literature go unprotected.

In light of this finding, they recommend that libraries, publishers, and archiving programs strive to enhance communication, coordinate efforts, advocate change, and make meaningful commitments to participating in initiatives. Publications describing these initiatives are reviewed in the "Initiatives" section of this paper.

Access

Issues related to access were a focal point in the 2006–7 serials literature. Perhaps the broadest contribution on this topic is O'Hara's analysis of the results of a 2005 survey assessing how 145 academic libraries make their e-serials accessible. The survey's findings suggest that libraries are generally relying on three access points: online catalogs, link resolvers (included Web-based lists powered by link resolvers), and metasearch engines.
contention that the greatest promise for a short-term resolution of the MulVer problem rests in enhancements that ILS developers can make to user interfaces. While the realities of current bibliographic control dictate that catalogers continue “to store and exchange data as cohesive manifestation-level description,” Allgood asserts that librarians should advocate the development of interfaces addressing the MulVer problem through enhanced capabilities for record indexing and display. 

Collins and colleagues offer an example of an effort to address the MulVer problem through an enhanced online catalog interface. They discuss a project in which North Carolina State University (NCSU) Libraries and Endeca Technologies collaborated to develop and implement Endeca as the user interface of the libraries’ online catalog. Collins and colleagues explain that the Endeca interface has the potential to automatically “connect or ‘FRBRize on the front end’” different manifestations of the same serial expression. They add, however, that, while the interface could show connections between records, the absence of an identifier in the MARC record for a work prevents the interface from “display[ing] a hierarchical view of the serial work.”

An additional barrier to effective serial access within the online catalog is discussed in a special section of the Serials Librarian featuring four articles examining the relative advantages and disadvantages of latest and successive entry cataloging. These articles discuss whether cataloging codes should retain the convention of cataloging serials according to latest entry, which can force users to search through several records to find the one that is needed. As with the MulVer problem, these articles look to FRBR and enhanced interfaces as possible resolutions.

**Link Resolvers**

O’Hara’s 2005 survey of 145 academic libraries revealed that link resolvers were used as an e-serial access point by 74 percent of respondents. This finding leads O’Hara to conclude that the technology, which can be used to generate Web-based serial lists, is “becoming a second library catalogue for serials.” Apps and MacIntyre discuss how a link resolver works, explaining that the technology supports context-sensitive linking by enabling a library’s authenticated users to seamlessly link from a citation in a database to options that the library offers for accessing the cited content. Beyond this core function, articles have explored additional roles that a link resolver can play. These additional roles include providing data for analyzing users’ search patterns and generating links from citations in the online catalog and free online resources (e.g., Google Scholar, Windows Live Academic, and Open WorldCat, now WorldCat.org).

The widespread implementation of link resolvers has resulted in articles that compare and assess specific products. For example, Livingston, Sanford, and Bretthauer describe a project to determine the best link resolver for the University of Connecticut Libraries (UCL) through an investigation of other libraries’ experiences using link resolvers. Drawing on the results of a literature review, surveys, and on-site visits, the authors were able to make in-depth comparisons between three products: Ex Libris SFX, Endeavor LinkFinderPlus, and Serials Solutions Article Linker. SFX was ultimately selected as being the best fit for the needs of UCL. Among the factors leading to this decision were SFX’s accuracy, flexibility, low maintenance requirements, large market share, and detailed reports and use statistics.

Wakimoto, Walker, and Dabbour assess users’ and librarians’ experiences with the SFX link resolver. Working in the San Marcos and Northridge campuses of the California State University System, the authors conducted online surveys of users, focus groups of librarians, analyses of use statistics, and test searches. In the case of users’ experiences, they found that, by a small margin, expectations regarding SFX exceeded users’ level of satisfaction. Librarians were generally satisfied but expressed unease with inaccurate information that SFX sometimes provided concerning accessible content. The authors note that, in general, complaints were not due to deficiencies of SFX itself but instead involved the databases that SFX links to and from.

The enhancement of link resolvers is the subject of a report by Culling, who recommends means of improving coordination and communication of information in the knowledge bases powering link resolvers. Drawing primarily on the results of interviews with representatives of the various parties involved in managing link resolver knowledge bases, the author describes the nature of the knowledge base supply chain and the relationship of the various stakeholders in this chain. Culling finds misunderstandings and poor coordination throughout the chain and recommends the development of an organization that “would seek to bring stakeholders together to define a visible code of practice for effective participation in the knowledge base supply chain.” Furthermore, the author advocates that stakeholders increase their partnerships with subscription agents while taking a proactive stance in applying tools for the automated exchange of knowledge base information.

**Metasearch Engines**

In O’Hara’s 2005 survey of 145 academic libraries, 30 percent of respondents reported that they make e-serials accessible through a metasearch engine, which enables a user to search multiple databases simultaneously. The nature and effect of metasearch engines as access points is the subject of a special section of a 2006 issue of *Serials Review.* A central focus of a number of the articles in this...
section are the development and features of specific metasearch engines on the marketplace, including SwetsWise Searcher, Endeavor Discovery: Finder, WebFeat Express, and Muse Metasearch Engine. In addition, this section provides guidelines for the selection and implementation of a metasearch engine. For example, Highsmith and Ponsford discuss Texas A&M University Libraries’ implementation of Ex Libris’ metasearch engine, MetaLib, Tracing a process that extended from fall 2004 through January 2006, Highsmith and Ponsford describe the stages of implementation, including database testing and configuration, interface customization, prerelease user testing, beta testing, and staff and user training.

Lindahl contributes another perspective on the implementation of a metasearch engine. The author contends that most commercial products’ out-of-the-box interfaces make the metasearch process more complex and time-consuming than necessary. Drawing on the University of Rochester River Campus Libraries’ development and enhancement of its metasearch engine, Find Articles, Lindahl offers a case study of how a library can collaborate with stakeholders to customize its metasearching capabilities so that they more effectively meet users’ needs and expectations. Walker adds to the literature’s discussion of innovations to metasearch engines by extending focus from locally implemented enhancements to industrywide standards being developed by the National Information Standards Organization (NISO). The author explains that the goals of the NISO Metasearch Initiative are threefold. These goals are to empower

• metasearch service providers to offer more effective and responsive services;
• content providers to deliver enhanced content and protect their intellectual property; and
• libraries to deliver services that are distinguished from those offered by Google and other free Web services.

Management

As serial collections, acquisitions, and access points are evolving, so too are management strategies. The 2006–7 literature features an abundance of publications describing how the transition to e-serials is leading managers to achieve change by enhancing workflows and communication channels.

Achieving Change

At the core of managers’ efforts at enhancement is an ability to achieve change. White explores this topic in a discussion of the University of Memphis’s implementation of staffing changes at the libraries’ periodicals desk. Following an analysis of different change models, White states that the libraries’ plan included five steps: “defining the changing, creating a common goal, involving the staff, providing an opportunity for feedback, and providing an opportunity to learn and grow.” Ohler contributes an additional perspective. Drawing heavily on the professional literature, she discusses four components to achieving change that any manager must grasp: “(1) The information and serials environment, (2) organizational structure and culture, (3) workflow analysis and staff resources, and (4) the implementation and use of technology.” A key concept emphasized throughout Ohler’s analysis is the importance of cultivating an attitude of openness in adapting to users’ expectations, in fostering communication within an organization, and in implementing the tools and technologies needed to manage e-serials.

Workflow Analysis and Reorganization

Managers cannot apply their knowledge of how to achieve change without first being aware of when change is needed. Yue and Anderson describe how the University of Nevada, Reno Libraries increased their awareness on this account through the development of a flowchart depicting the libraries’ workflows for managing e-serials. They explain that, through its illustration of procedures, the flowchart has enabled the libraries to identify ways to clarify responsibilities, streamline operations, and eliminate inefficiencies. Graves and Arthur give another example of the benefits of analyzing serial workflows. They discuss a project that the Serials Unit of Old Dominion University Libraries conducted to assess workflows and resource allocations during the libraries’ transition from print to e-serials. The most influential outcomes of this analysis were the establishment of a Serials and Electronic Resources Unit and the transformation of the titles and responsibilities of two librarian positions so that these positions can better coordinate e-serial management.

As libraries have updated their workflows to address the challenges of e-serials, the need for traditional, print-centered procedures has been called into question. Anderson argues that libraries should adopt practices that are more representative of users’ preferences for accessing serials electronically. In doing so, Anderson finds four examples of tasks that are not always a prudent allocation of time and resources: claiming, binding, subject authority control, and unessential customization of records. Borchert describes one library’s effort to discontinue a fundamental procedure in print serial management: check-in. During the University of South Florida Tampa Library’s migration to a new ILS, managers opted to stop routine serial check-in. Due to such factors as the arrangement of the library's
collection according to the Library of Congress classification system and the library's commitment to continue binding serials. Borchert reports that the experiment led the library to conclude that check-in is still necessary.

Frost and Woo discuss a similar workflow change, this one consisting of the elimination of binding at Hong Kong Baptist University Library. Low use of print serials combined with increasing subscription and binding costs resulted in the authors' recommendation that the library discontinue binding all currently received serials that are either (1) accessible perpetually online, (2) accessible online (regardless of perpetual access provisions) and used less than five times per year, or (3) unscholarly newsletters. Instead of binding these materials, which constitute over 85 percent of the library's currently received serial collection, the authors advocate that noncurrent issues be stored in boxes.

Communication

E-serials are also changing managers' communication channels. Feather explores these changes in a discussion of Ohio State University Libraries' analysis of e-resource management communications. The analysis aimed to develop an awareness of the nature, structure, and role of the varying types of e-resource communication occurring at the libraries. Feather reports that this awareness enabled the libraries to enhance communication by updating and improving online request forms, reducing the number of individuals involved in certain workflow communications, reducing the number of inappropriate messages sent to an e-resources unit group e-mail account, spreading awareness among other staff about the e-mail clutter caused by notifying too many individuals of a problem, and encouraging library-wide staff viewing of ERMS records.

Other publications shift the focus from internal communications to communications between libraries and their external partners. For example, Robertson reports that Strader, Roth, and Boissy presented at the 2005 North American Serials Interest Group Annual Conference on how libraries can better collaborate with publishers and subscription agents. The presenters proposed a checklist outlining the responsibilities that each party has in ensuring a libraries' e-serial access is activated and retained.

Initiatives

The challenges libraries face in the management of their serial collections have led to the development of innovative partnerships among libraries, publishers, subscription agents, and other stakeholders. The initiatives resulting from these partnerships are a major topic of discussion in the 2006–7 literature.

Acquisition and Administration

With the transition to e-serials, acquisition increasingly necessitates the negotiation of a license agreement, which is a complex task involving a significant investment of time and expertise. Hahn describes one effort to simplify this undertaking: NISO's Shared E-resources Understanding (SERU) Working Group. Through its development of a best practices document that both a library and publisher can honor, the SERU Working Group offers a pragmatic alternative to license negotiations. Hahn explains that by accepting the terms of the document, both parties can forgo negotiations, thereby streamlining the acquisition process.

Beyond license negotiations, acquisition and administration require that libraries, publishers, and subscription agents exchange metadata regarding serial access and availability. Miller and Klemperer discuss how the NISO/EDIEUR Joint Working Party for the Exchange of Serials Subscription Information has enhanced this process through its development of three Online Information Exchange (ONIX) formats: Serials Products and Subscriptions, Serials Online Holdings, and Serials Release Notice. Among the positive outcomes that libraries can achieve through these standards are a reduction in unneeded claims for print issues, the automation of URL changes in a library's access portals, and the reconciliation of holdings in preparation for package deals.

Following the acquisition of an e-serial, a library must effectively record, track, and communicate the business and licensing terms. The central tool that libraries rely on to complete this task is an electronic resource management system (ERMS). While the literature of previous years centered on the introduction of ERMS, the 2006–7 literature places increased focus on efforts to enhance these systems. Fons and Jewell, for example, discuss the second phase of the Digital Library Federation's Electronic Resources Management Initiative (ERMI). The authors characterize the 2004 report resulting from the initial phase of ERMI as a "key document for the development of ERMS" and explain the ways in which the second phase of ERMI will further enhance e-resource management. Among the enhancements they cite are a review and update of the first phase's Data Dictionary and the facilitation of opportunities through which librarians can use this document to map licensing terms to ERMS fields. Other focal points include the integration of ERMS with ILS, link resolvers, and standards for evaluating e-resource use.

While many of the ERMS available to libraries are
commercial products, other systems have been developed by libraries themselves. For example, Meyer describes E-Matrix, an ERMS developed by NCSU Libraries, and Stranack describes CUFTS, an open-source serial management software system developed by Simon Fraser University Library. Discussing the lessons learned from implementing a homegrown ERMS, Meyer advises that libraries opting to take this path will need personnel with significant expertise in both programming and e-resource management.

Evaluation

The literature’s discussion of the evaluation of serial use centers around two initiatives: Counting Online Usage of Networked Electronic Resources (COUNTER) and the Standardized Usage Statistics Harvesting Initiative (SUSHI). Pesch describes COUNTER as a code of practice that e-resource access platforms can voluntarily adopt to consistently record and exchange a library’s e-resource use information. In a separate article, Pesch discusses how SUSHI builds on the COUNTER initiative. He explains that SUSHI is a protocol through which COUNTER-compliant use statistics can be automatically transmitted from e-resource access platforms to a library’s ERMS. In doing so, SUSHI relieves libraries from the tedious process of manually retrieving use statistics.

The implications of initiatives such as COUNTER and SUSHI have been explored from a number of contexts. Analyzing the e-resource use statistics of a large research library over a three-year period, Blecic, Fiscella, and Wiberley consider the effect of both COUNTER and enhancements to users’ ability to search and access e-resources. Among the authors’ key findings is that, while COUNTER has significantly enhanced libraries’ ability to evaluate e-resource use, enhancements in users’ abilities to search e-resources redefine the meaning of use statistics. Accordingly, they caution that enhancements in e-resources’ searchability requires corresponding enhancements in the measures libraries rely on for evaluating use.

In a study sponsored by the United Kingdom Serials Group, Shepard examines another topic related to the success of initiatives such as COUNTER and SUSHI: the viability of developing usage factors (UF). The UF would offer a means for measuring a serial’s quality on the basis of use statistics. Describing the results of a survey of authors, editors, librarians, and publishers, Shepard reports that “there is significant support, even among established publishers whose journals perform well in IF [ISI impact factor] rankings, for the development and implementation of journal UF s.” The findings of Duy and Vaughan offer further insight on the relationship between e-serials’ use and IFs. Assessing the use and citations of chemistry and biochemistry serials at Concordia University Libraries, the authors found that, while there were strong correlations between print and electronic use and between electronic use and local citation data, there was no correlation between IFs and electronic use.

Archiving

The 2006–7 literature’s most far-reaching analysis of e-serial archiving initiatives is a Council on Library and Information Resources report authored by Kenney and colleagues. This report discusses the results of a survey of twelve e-serial archiving initiatives in which representatives of the initiatives were questioned regarding six topics: “organizational issues, stakeholders and designated communities, content, access and triggers, technology, and resources.” Based on the responses, the initiatives were evaluated regarding their ability to meet indicators for success. These indicators concerned each initiative’s mission and mandate, rights and responsibilities, content coverage, minimal services, access rights, organizational viability, and role within a network. Key among the report’s recommendations are that initiatives “should present compelling public evidence that they offer at least the minimal level of service for well-managed collections” and that they clearly indicate the publishers and holdings included. Further recommendations involve securing guarantees that holdings can never be removed; considering the implications of holdings’ entry into the public domain; and forming a network of initiatives in order to provide mutual support, broaden collaboration, and enhance communication.

The archiving initiatives receiving the most attention in the literature are Portico and Lots of Copies Keep Stuff Safe (LOCKSS). Portico is a nonprofit initiative developed with support from JSTOR, Ithaka, the Andrew W. Mellon Foundation, and the Library of Congress. Fenton, the executive director of Portico, describes the initiative’s archiving strategy as the normalization of the source files contributed by participating publishers. This approach aims to facilitate the successful migration of the files as new data formats replace current formats. Portico grants supporting libraries access to archived content following designated “trigger events” or, in some cases, following a supporting library’s cancellation of an archived resource. LOCKSS archives e-serials using a different strategy. As Seadle states, this initiative “offers a community-based rather than a corporate approach.” He expands to explain that LOCKSS constitutes a network of libraries using the same open-source software. This software both archives the source files of participating publishers and maintains the integrity of these files by comparing the contents of each libraries’ LOCKSS archive with the archives of other libraries in the network. In contrast to Portico, LOCKSS does not normalize source files. Due to concerns that normalization may corrupt data.
and alter content, LOCKSS relies on a bitstream approach to archiving that preserves content precisely as it appears to users.

**Conclusion**

The 2006–7 literature shows that serials librarianship is in a period of great innovation. Propelled forward by user preferences, libraries are rapidly transitioning from acquiring serials in print to providing access electronically. Accompanying this transition in the formats of collections are evolving concepts of seriality and increases in subscription costs. Among the outcomes of these changes are new ideas regarding the models through which serials are acquired. Although more established models such as publisher packages continue to pervade, libraries are demonstrating growing interest in alternatives. These alternatives include relying on OA content and acquiring access through arrangements that do not include provisions for perpetual ownership. Countering this latter strategy are voices within the profession that advocate the need for libraries to secure perpetual ownership provisions during the acquisition process.

Innovations are equally apparent in the literature's discussion of serial access, management, and initiatives. Online catalogs, link resolvers, and metasearch engines are emerging as libraries' primary points for providing serial access. For each of these access points, efforts are underway to evaluate and enhance users' abilities to search for and access content. Meanwhile, managers are achieving change by reassessing and restructuring workflows, organizations, and communication channels so that they are focused on the electronic access and administration of serials. Finally, stakeholders throughout the serials landscape are partnering to develop new initiatives. For example, SERU holds promise as a pragmatic alternative to license negotiations; COUNTER and SUSHI are enhancing the evaluation of e-serials; and archiving initiatives such as Portico and LOCKSS are providing mechanisms through which libraries can retain perpetual access to their e-serial collections.

Looking to the future, the literature is sure to reflect further innovations in the movement to transform serials and libraries. With these innovations will come significant challenges to the imaginations of those engaged in serials librarianship. For example, the 2006–7 literature shows a gulf between some of the alternative models being explored for acquiring serial access and the perspectives of commentators advocating the need to secure perpetual access provisions. Publications aiming to both clarify and reconcile these differences between the need to meet users' expectations for expansive e-serial access and research libraries' traditional commitment to retaining ownership of their collections would be welcome additions to the professional literature.

Also of value to the professional literature would be more publications examining the wider effect of the transition to e-serials on libraries' organizational structures and tools for providing and managing e-serial access. Indeed, while the 2006–7 serials literature includes numerous contributions discussing the implementation of specific tools and tasks related to e-serials, the literature includes relatively few publications addressing the large-scale implications that the centrality of e-serials is having on libraries. For example, the literature would be enriched by publications describing how the transition to e-serials has led to larger changes in the organization of departments and workflows and in the overall infrastructure of tools libraries rely on to manage and provide e-serial access. The 2006–7 serials literature's focus on specific tools, projects, and procedures likely will serve as a springboard for future contributions to the literature that explore the broader effect of innovations in serials librarianship.

**References**

2. Ibid., 83.
6. Ibid., 362.
8. Ibid., 390.
13. Johnson and Luther, The E-only Tipping Point for Journals,
24. Hahn, “The State of the Large Publisher Bundle.”
25. Ibid., 2.
26. Ibid.
28. Ibid., 20.
30. Ibid., 1.
32. Ibid., 83.
34. Ibid., 163.
40. Stempner and Barrireau, “Perpetual Access to Electronic Journals.”
42. Ibid., 1.
44. Ibid., 126.
47. Ibid., 87.
49. Ibid., 175.
51. Ibid., 92.
52. Ibid., 93.
56. Ibid., 127.
59. Jill Livingston, Deborah Sanford, and Dave Brethauer, “A Comparison of OpenURL Link Resolvers: The Results of a
University of Connecticut Libraries Environmental Scan,”


Ibid., 5.


69. Ibid., 127.


71. Ibid., 22.


73. Ibid., 40.


77. Carol Ann Borchert, “To Check In or Not to Check In? That Is the Question!” Serials Review 33, no. 4 (2007): 238–43.


80. Ibid., 211.


85. Ibid., 152.


91. Ibid., 122.


94. Ibid., 14–15.

95. Ibid., 2.


Criticism of Cataloging Code Reform, as Seen in the Pages of Library Resources and Technical Services (1957–66)

By Steven A. Knowlton

The history of cataloging rules is often written as a story of continuous improvement toward a more rational and efficient code. Not all catalogers, however, have been in agreement that reform of the cataloging code has been improvement. The debate of the 1950s and 1960s over cataloging code reform, hosted in part by LRTS, is an example of conflicting values in the cataloging community. Seymour Lubetzky’s proposal for a cataloging code based on logical principles eventually became the Anglo-American Cataloguing Rules, but many catalogers of the period felt that other values, such as tradition and the convenience of the user, also deserved consideration in the cataloging code.

The library historian Wiegand has said, “We are all prisoners of our own discourses,” meaning that the stories we tell about ourselves influence our views of our place in culture and society.1 For librarians in the United States, that means that they often consider their institutions “cornerstones of the communities they serve” because “free access to the books, ideas, resources, and information in America’s libraries is imperative for education, employment, enjoyment, and self-government.”2

What librarians tell themselves and each other about their professional values plays an important part in how they perceive their own history. Many librarians view the library as an institution that has been instrumental in moving society toward “modernity, progress, and science.”3 Whether the values of modernity, progress, and science are appropriate values to guide librarianship goes unquestioned by librarians, for the most part.

A similar discourse is evident in discussion of the history of Anglo-American cataloging codes. Wynar and Taylor have stated that the current cataloging code, Anglo-American Cataloguing Rules, is “the result of a progression of ideas about how to approach the cataloging process in order to prepare catalogs that provide the best possible access to a library collection.”4 Chan has written that earlier codes were “pedantic, elaborate and often arbitrary.”5 These ideas were introduced in basic cataloging textbooks in 1985 and 1994, and such thinking dominates historical discussion of the efforts of the 1950s and 1960s to reform the cataloging code.

Inspection of the written record of the cataloging profession, however, indicates that the view of the Anglo-American Cataloguing Rules as an improvement over then-current cataloging codes was not universally shared.

The pages of LRTS abound with debate over the cataloging code, and in celebration of the fiftieth year of LRTS, this paper seeks to demonstrate how the
reform of the cataloging code was accompanied by many divergent voices, whose claims may help reframe the discourse about the history of cataloging. The discussion about cataloging code reform was not only a technical debate over the merits of various methods of entry, it was also a multi-layered debate about the values that should prevail in the cataloging profession. At one level was the question of cost in time and money to revamp the existing catalogs—and in the cost to scholarship of retraining the research community in the use of the catalog. At another level was the question of whether the admittedly important value of logic should prevail completely over other values that had motivated earlier framers of cataloging codes, such as tradition and the convenience of the user. The latter term, as used in defense of retaining the former cataloging code, generally referred to the practice of entering a heading where a reasonable user was presumed to be likely to look for it—"the public's habitual way of looking at things." 

While librarians know today that Seymour Lubetzky's vision of a logical, principled cataloging code did indeed prevail, considerable dissent met the notion that his way was, in fact, the best way to prepare catalogs. As catalogers are today working on yet another round of cataloging code reform, a useful exercise for today's catalogers may be to review the debates of the past. In this way, the debate may travel outside the discourses that have dominated thinking about cataloging since the adoption of the first Anglo-American Cataloguing Rules (AACR) in 1967.

The First Century of Cataloging Codes

Since Antonio Panizzi published the “91 Rules” for compiling the book catalog of the British Museum in 1841, cataloging codes have been in a continuous state of change. Charles Coffin Jewett adapted most of Panizzi's rules for the Smithsonian Institution in 1850, Charles A. Cutter devoted several decades of the second half of the nineteenth century to developing Rules for a Dictionary Catalog, and Melvil Dewey’s Library School Rules (1888) reflected his work directing the Columbia School of Library Economy. In addition, codes by Klaus August Linderfelt and Frederick B. Perkins and a pamphlet of suggestions by the Library Bureau were also in circulation in the late nineteenth century. Contemporaneously, a committee (which included Cutter) of the American Library Association (ALA) prepared a “Condensed Rules for an Author and Title Catalog” in 1883. However, within a couple decades, the rules had not prevented “considerable divergence in the practice even of libraries organized subsequent to 1883.” Between 1901 and 1908, a second committee (again including Cutter) worked to develop a revised cataloging code “to bring about uniformity between its revision of the A.L.A. Rules, the 4th edition of Cutter's Rules for a Dictionary Catalog . . . and the Library School Rules.” The committee also worked with the Library Association in Great Britain to harmonize the cataloging codes used in the United States and in the United Kingdom. The resulting Catalog Rules: Author and Title Entries (American edition) and Cataloguing Rules: Author and Title Entries (British edition) were jointly adopted by the ALA and the British Library Association in 1908.

The necessity for all libraries to adopt a shared set of cataloging rules had become steadily more apparent as early union catalogs were created, and had taken on added urgency in 1901, when the Library of Congress (LC) issued printed catalog cards for titles it had received. As librarians across the country took these printed cards into their catalogs, their locally cataloged materials required entry and description according to the same rules as the titles cataloged at LC. Hence, the adoption of the 1908 rules was achieved after only a short review period, and with near-unanimity between the two largest library associations in the English-speaking world (separate British and American editions were issued, but with only minor differences in the rules). It was the first set of cataloging rules to achieve widespread acceptance in libraries in the United States.

The 1908 code, and each code that followed, was limited to rules for descriptive cataloging. Although some of the earlier codes, including Cutter's, included rules for subject entry, an English-language subject cataloging code for universal application has not yet been developed as of 2008.

After 1908, the LC introduced many changes and additions to the rules on an ad hoc basis, to address cataloging issues not covered by the 1908 Catalog Rules: Author and Title Entries. These changes and additions were issued to libraries that subscribed to the LC’s catalog cards, but “in the absence of any supplementary rules from the American Library Association since 1908, libraries . . . had to formulate their own rules, relying chiefly for guidance on rules issues occasionally by the Library of Congress, added to such deductions relating to practice as could be made from the printed cards as examples.” By 1930, librarians felt a need for a revised code to incorporate the LC’s revisions and reduce local variation in cataloging, so work began on an updated set of cataloging rules. The motivating idea for the revisers was the feeling that the 1908 rules had not been extensive enough, so that the revised rules would cover more circumstances that proved troublesome to catalogers—such as serials, anonymous classics, music, maps, pseudonymous works, and corporate authorship. The coverage of such fine details meant that the 88 pages of the original Catalog Rules: Author and Title Entries became 408 pages in the revision published in 1941. Furthermore, most of the justification for the rules came from prior use, or “precedent,” rather than any logical reasoning; many rules had exceptions;
and inconsistencies in the treatment of different types of material were noted. The structure of the rules, consisting of compound complex sentences with few illustrative examples, made the application of the rules difficult.\textsuperscript{20}

The complexity of the 1941 edition (which applied only in the United States, as British librarians could not participate in the revision due to World War II) lead to dissatisfaction in some quarters. Osborn published a famous article, “The Crisis in Cataloging,” in 1941, in which he lamented the large backlogs in cataloging departments and predicted that an even more complex set of rules would further slow down catalogers—an irony in light of the new rules’ purpose of easing catalogers’ work by providing rules for more types of publication and issues of entry.\textsuperscript{21}

Cutter had attempted to generate his rules according to the objectives of a catalog: namely, to allow the user to find a book, to show what the library has, and to assist in the choice of a book.\textsuperscript{22} However, the 1908 rules and their subsequent revisions had excluded the statement of these objectives. In response to the criticism of the 1941 rules, the ALA commissioned another revision (\textit{A.L.A. Cataloging Rules for Author and Title Entries}) to simplify the rules and arrange the presentation so that the principles behind the rules would be more apparent.\textsuperscript{23} This 1949 revision also eliminated rules for description of the book; only rules for entry were included. Despite these changes, the 1949 revision was also criticized for its complexity and unwieldiness.\textsuperscript{24}

Although the 1949 ALA cataloging rules had omitted rules for description, the LC published its own \textit{Rules for Descriptive Cataloging in the Library of Congress} in 1949; the rules were originally prepared for internal use, but were published for the wider library community in order to provide guidance for librarians using catalog cards printed by the LC.\textsuperscript{25} The 1949 ALA rules for entry and 1949 LC rules for description became known, respectively, as the “Red Book” and the “Green Book” from the colors of their bindings. The \textit{Rules for Descriptive Cataloging at the Library of Congress (RDC)} were considerably simplified from the 1908 and 1941 codes, and were largely praised for this fact.

\textbf{Seymour Lubetzky’s Cataloging Rules and Principles}

In light of the praise for \textit{RDC} and the less positive reception of the ALA rules for entry, the Board on Cataloging Policy and Research determined to approach rules for entry in the same fashion that the LC’s rules for description had been developed: namely, “prepare the simplest code of descriptive rules which could meet the established needs.”\textsuperscript{26} To begin the work of preparing the simplest code, the ALA engaged the services of Seymour Lubetzky, a librarian at the LC who had also worked on the \textit{RDC}. Lubetzky first prepared a critique of the 1949 ALA rules for entry, called \textit{Cataloging Rules and Principles: A Critique of the ALA Rules for Entry and a Proposed Design for the Revision}.\textsuperscript{27} Lubetzky’s critique not only pointed out the flaws in the existing rules for entry, but laid out the need for establishing a set of principles from which an improved code could be built. It included his famous question, “Is this rule necessary?” to which the answer was often “no” because the determination of the form of heading or rule of entry could be discerned from a larger principle, without need for a specific rule.\textsuperscript{28}

According to Tillett, Lubetzky felt that the cataloging rules had become so complex because catalogers had lost sight of the reason for the catalog: to help users identify and distinguish among works that meet their needs.\textsuperscript{29} Cataloging rules that expressed the principles defined by Cutter (and refined by Lubetzky) would of necessity be simpler, and would allow catalogers to create better catalogs.

By 1954, the ALA had decided to prepare a complete revision of both the Red and Green Books, and appointed a Catalog Code Revision Planning Committee to the task of overseeing and advising Lubetzky’s drafting of a revised code.\textsuperscript{30} Over the next decade, many discussions about the revised code were held in symposia and in the pages of journals. Almost all discussion focused on the approach to cataloging presented by Lubetzky, whose work became the \textit{sine qua non} of the new cataloging code. As Dunkin wrote in 1959, “The genius of Seymour Lubetzky now dominates our thinking about the catalog as completely as Cutter once did.”\textsuperscript{31}

\textbf{\textit{LRTS} as a Forum for Debate}

In this environment of serious contemplation of the principles by which works should be cataloged, \textit{LRTS} was launched in 1957. Debate over cataloging code reform was not limited to the pages of \textit{LRTS}, but the pieces presented in that journal form a useful record of the voices for and against reforming the cataloging code along Lubetzky’s plan. Although both pro- and anti-reform articles appeared in \textit{LRTS} between 1957 and 1966, this paper concentrates on articles composed by librarians who had reservations about the Lubetzky code, as they expressed a concern for values that have been considered of less importance than those that motivated AACR. Because the articles discussed Lubetzky’s proposed code on its merits, a variety of perspectives (including some commendation of aspects of the proposed reform, as well as reservations about the changes) can be traced through the pieces under consideration.

\textit{LRTS} in its first decade was not the research-oriented journal it is today. Rather, it was a forum for news and debate over the latest trends in technical services librarianship.\textsuperscript{32}
Articles were frequently fewer than five pages long, and included reports from various ALA committees, opinion pieces, and even humor. Because of its nature as a professional round table, LRTS provided an opportunity for librarians to voice their concerns about developments in the revision of the cataloging rules outside of the formal structure of a research article or literature review. Throughout the 1950s and 1960s, discussion of current cataloging issues was published regularly.

The very first volume of LRTS, published in 1957, features a lengthy review of a symposium held at the University of Chicago in June 1956, called “Toward a Better Cataloging Code: A Review.” An “unusually large” number of attendees (148) testified to the interest in the revision of the cataloging code among librarians, and the concerns voiced by some of the speakers foreshadowed the debate that would follow for the next decade. While a number of speakers expressed enthusiasm for Lubetzky’s proposals to return to the basic principles of cataloging, most were concerned about the cost of recataloging items already entered. Angell delivered a more philosophical demurral. Where Lubetzky wished to do away with all entries other than author and title, Angell preferred to retain form headings (for example, “Laws, statues, etc.” or “Anonymous classics”) as a natural entry (that is, an entry that a user would think to look under using his or her native intelligence). Angell also raised the point that both the Red (ALA rules for entry) and the Green (LC rules for description) Books needed to be revised despite the general acceptance of RDC because the choice of entry influences how the name of the author may be described and because the descriptive rules should include provisions for media other than books. Osborn also urged that the new code seek to achieve harmony with codes of other countries. Henkle (Lubetzky’s former supervisor at the LC) raised the issue of user studies; some librarians felt that data from the observation of nonlibrarian catalog users should influence the code. All these issues would continue to be important topics in LRTS for the next two decades.

Another early article in LRTS supported Angell’s proposal to revise rules for description along with rules for entry. “The Red and the Green” by Waters of Georgetown University used a sample of publisher statements (of the RDC) to demonstrate the difficulty in determining proper description of that field according to RDC. Waters felt that a review of the principles and goals that descriptive rules served should accompany the review of principles for entry, and that both set of rules should be revised simultaneously.

The Draft Code and Its Discontents

By 1958, Lubetzky had prepared a draft of a revised cataloging code, which was discussed by more than 175 librarians at the “Institute on Catalog Code Revision” at Stanford in July 1958. As promised in his earlier works, Lubetzky laid out the objectives of the code as the first statement of the code: “1) To facilitate the location of a particular work; and 2) To relate and bring together the works of an author and the editions of a work.” The similarity to Cutter’s objectives was noted—but Lubetzky had done away with another of Cutter’s principles: serving “the convenience of the public” (in the sense of deferring to the searching practices of users). This ambiguous phrase had led to many of the awkward, contradictory, and unintuitive rules in the 1941 and 1949 codes, such as entering certain types of corporate body under their location and the use of form headings. Instead of “the convenience of the public,” Lubetzky relied on logic in the observation that a simple rule, strictly followed, will become apparent to the catalog user and therefore serve him or her better than a maze of unexplained and inconsistent rules with ad hoc exceptions for particular circumstances. In this way, it was believed that the convenience of the public was served more effectively. To achieve the stated objectives, Lubetzky insisted on main entry under a name or title. No entries under location or form were to be made. Lubetzky’s draft code also addressed the contentious issue of corporate authorship by calling for entry of serial titles and corporate bodies that changed name under their successive names. A number of critics felt that this policy undermined the second objective.

According to a report on the Stanford Institute, which was the first public discussion of the draft code, a number of attendees questioned the value of Lubetzky’s second objective (“to relate and bring together the works of an author and the editions of a work”). Many at the institute felt the draft code promoted excessive cross-entry, requiring more complex rather than simpler rules for entry. Wright questioned whether the code should consider subject entry as well. The issue of the cost of converting the catalog to a new code was raised, along with the necessity of international cooperation on cataloging rules. However, the institute achieved a consensus on the notion of preparing the best code and then finding methods to achieve cost savings or international agreement afterward as the most productive approach. Further issues were raised, but left unresolved. These would continue to occupy the minds of catalogers as revision continued—the problems of corporate author entry (which circumstances require corporate, rather than personal, authorship; under successive or latest name; under subdivisions; under location) and serial title entry (successive versus latest title). During this period, Lubetzky wrote an article for LRTS explaining the process of code revision and his own reasoning behind the principles and rules used in the code. After this, the task of defending the code in the page of LRTS against its critics fell to other writers. Following the Stanford Institute
Concerns about Corporate Authorship in Lubetzky’s 1960 Draft Code

Although some of the issues debated were of a technical or practical nature, the issue of corporate authorship was one in which the values of the Lubetzky code stood in strong contrast to the values of the earlier codes. In particular, logic was pitted against tradition and user convenience, the latter referring to the sense that a catalog should have entries for corporate bodies where a user would look for them.

Draper of the University of California, Berkeley was dismayed that discussions of cataloging code reform (and the 1960 Draft Code) had not sufficiently addressed the problem of determining under which circumstances an entry should be made under a corporate author as opposed to a personal author. He found the rule for entry under corporate body to be “highly vulcanized, i.e., full of rubber which can stretch in any direction at will,” because the wording of the rule allowed for much latitude in interpretation.

Haskins of Harvard University defended the to-be-discarded rules requiring entry of local or civic institutions under place by referring again to the convenience of the user:

From the standpoint of the use of the catalog the most direct approach would appear to be by the place where [the institutions] are located. Also, there would seem to be a real advantage in bringing together the schools, hospitals, churches, museums, etc., that generally may be of slight interest individually, but which play such a large part in the life of a city. If an institution bears a name that has little significance without the place where it is located, whether it be the Free Public Library, the First Church, Unitarian, Saint Paul’s Church, or Saint Luke’s Hospital, is it not logical to record it under the name of the place?

Implementation of the Draft Code in Imagination and Experiment

The Summer 1961 (vol. 5, no. 3) issue of LRTS featured a series of articles on the effects that implementing the 1960 draft code (now called Code of Cataloging Rules: Author and Title Entry, an Unfinished Draft, or CCR) would have on the operation of libraries. Dunkin of Rutgers presented an overview of the changes catalogers would have to make in the switch from the 1949 rules to CCR. He called it “Guesstimates Unlimited,” but only pointed out three major areas that would require significant changes in the form of entry: the use of a uniform title following a personal name main entry (a new idea first proposed in CCR); the elimination of the distinction between “institutions” and “societies” among corporate bodies, and the entry of all corporate bodies under name rather than place; and entry of anonymous works under title, rather than form. Dunkin offered suggestions for adapting the catalog to the new rules, such as using guide cards to provide cross-references from the older form of entry to the CCR form.

Wright of Williams College presented the results of a survey of catalogers who were asked to examine entries currently in their catalogs and determine if CCR would require changes in form of entry. Under the rules 70 percent of headings would remain unchanged, 13 percent would require minor changes, and 17 percent of headings would be different. Most respondents reacted favorably to the new rules as “more explicit, more reasonable, and easier to use,” although some expressed reservations about making such a large number of changes.

Haskins wrote—on behalf of the librarians at Harvard—in defense of many of the old ALA rules, including form headings and entry under place for corporate bodies, “What is to be gained by giving up this type of heading which has been in use over a long period and is generally understood and liked?” She also found much to object to in the imposition of new rules, such as uniform titles combined with author main entry, changes in the form of foreign names, and successive entry for corporate bodies that change names—mainly on the grounds of the need to revise and update thousands of catalog cards, with little gained (in the opinion of Harvard’s librarians). She concluded with several thoughts about the flurry of cataloging rules changes that had come in the 1940s and 1950s:

I am beginning to wonder if we, as librarians generally and as catalogers specifically, know what we really want in the way of a cataloging code. We became dissatisfied with the 1908 code. For one thing it was too general. So a large committee made up of extremely able people worked for many years to revise the rules. The result was a very detailed code. In that respect it should have been the answer to a cataloger’s prayer. Perhaps it was for many. But within a short time, even before the second (1949) edition was published, it was on the carpet and was severely criticized for its complex rules, when the trend was toward simplification, for its lack of organization, its lack of basic principles,
and so on. So once again we set to work. This time we started from scratch. . . . But from there on have we gone far enough or have we gone too far? Are we going to be successful this time? . . . We also started this revision by shutting out the past, closing our eyes to all the water that had gone over the dam. We have now come to the point where we can no longer disregard what has gone before. . . . How much can the large research library afford in order to implement rules that call for so many changes in practice?55

Brown, of the Free Library of Philadelphia, wrote with concerns about the rules requiring uniform titles for works that appear under various names.56 She preferred the entry as it appears in the work, whether it is title, corporate name, or personal name. Although it would create a “mongrel catalog,” her opinion was that users would be better served (particularly in a large public library) by reducing the number of “two-step searches,” which would be caused by the creation of uniform headings (step one was finding the proper heading from cross-references, step two was searching under that heading—a lengthy process when using a large card catalog).57 Further, she found that a rigid application of principles should give way to a consideration of user behavior: "The Nibelungenlied, whether considered from the point of view of bibliographical characteristics or from the point of view of use, differs significantly from a recently published government document on jet propulsion. Consistency is a virtue in developing a catalog, but . . . [i]t need not be interpreted to mean that the same policy must be applied to all material regardless of that material's bibliographical characteristics.58

Hines of Rutgers wrote with concerns about Lubetzky's use of the term "work" instead of "book":

The implication is that the work is to be considered as an intellectual rather than as a physical entity. . . . This distinction between the physical and intellectual cannot be pushed too far. It is clear that . . . Lubetzky does not mean that we should have a single main entry for Nine Plays of Bernard Shaw which would file with editions of Caesar and Cleopatra issued as physically separate bibliographic units. . . . It is here that a qualifying phrase seems to be needed in the draft code. It would appear that the code tacitly accepts the long-existing premise that the cataloger deals with physical bibliographic units, and that he catalogs them as such. . . . This preference for the physical bibliographic unit in cases of conflict [with intellectual units larger or smaller than the physical units] should be explicitly stated in the code.59

Beckman reported the results of an experiment at the University of Waterloo in which CCR was used to catalog new acquisitions.60 Although she found the “revised code a pleasure to work with,” and noted the case with which her catalogers now addressed the names of authors, she did describe some difficulty in applying the rules for works of changing authorship, such as yearbooks and dictionaries.51 “The most difficult problem with this rule is that it is impossible to tell when handling a first edition of a reference work whether or not it will go into successive editions.”62 As well, the rules in this section diverged so far from current LC practice that Waterloo was unable to use, even in modified form, printed cards from the LC for those titles.

The Paris Principles

All such criticisms of CCR would no longer be addressed by Lubetzky; in 1960, he left the employ of the LC and accepted a professorship at the University of California, Los Angeles. The Catalog Code Revision Planning Committee turned over the job of editing CCR to C. Sumner Spalding. Lubetzky made one more important contribution to the revised code in the form of his role in formulating what became known as the Paris Principles.63

As Osborn and others had noted, the American cataloging code revision was taking place during a time when librarians in other countries were also contemplating cataloging code reform. The destruction of many libraries in Europe during World War II made the possibility of revising cataloging codes more feasible because the number of books requiring recataloging was much reduced.64 Although the possibility of international agreement on cataloging rules had been explored at the International Congress of Archivists and Librarians at Brussels shortly after the publication of the 1908 rules, those in attendance determined that differences between Anglo-American and continental (particularly German) rules were too great.65 During the 1950s, a number of library associations—including those of France, Poland, Japan, Spain, Italy, Switzerland, the U.S.S.R., and India—worked on revised cataloging codes. The Library Association (of the United Kingdom) determined that it would work with the ALA so that the revised code being prepared by Lubetzky would be Anglo-American. In light of these developments, the International Federation of Library Associations (IFLA) convened conferences in 1958 and 1959 to discuss the possibility of an international agreement on cataloging principles. The result of these discussions was the IFLA International Conference on Cataloguing Principles (ICCP), held in Paris, October 9–18, 1961.

In Paris, representatives from thirty-four national library associations met and agreed on the Paris Principles,
which served as the basis for future cataloging codes in most countries. Lubetzky's contributions included articulation of the principle of main entry. Another important point of agreement was the principle of corporate authorship, which had previously not been observed in Germany.

The importance of the Paris Principles to catalog code revision was that revision could go forward with an internationally accepted set of principles underlying it and also provide strict guidelines. As Kebabian of the University of Florida commented, "There is no doubt that American librarianship will be under world-wide scrutiny as our new code reaches completion." As with Lubetzky's code, the Paris Principles stirred up some criticism. For example, Scott of the University of Oklahoma found some of the guidelines to be too vague: "Consistent catalog entries for current materials cannot depend on 'best known' or 'most frequently used' form. Such criteria are useful only in retrospect." The question of whether the proposed Anglo-American code would go forward truly based on the Paris Principles was another concern. Kebabian noted,

Though the concept of the corporate author has been finding its way during the past ten or more years into French catalogs and bibliographies, for most national delegations this constituted the most fundamental break with tradition, and there was considerable debate at Paris before its final acceptance. In the form which was approved, moreover, it includes at least two provisions which contradict current United States practice. . . . It is ironical to reflect that these two principles were among those suggested by Seymour Lubetzky in his critique, Cataloging Principles, and that his studies and the preparation of that document stemmed from our desire to seek solutions to the inconsistencies of the "corporate complex" as structured in some seventy rules in the 1949 ALA code. While acceptance abroad at the ICCP was achieved, at home these principles constitute a problem of considerable consequence to achieve their reconciliation with existing entries in our long-establish, monolithic card catalogs; they are the one major source of yet unresolved compromise efforts in the current preparation of our code of cataloging rules. It is thus that the dead hand of history plagues us."

**Progress toward AACR**

With the Paris Principles in place, the Catalog Code Revision Planning Committee continued to revise the cataloging code. An important agreement was settled upon at the 1963 ALA Midwinter Meeting in Miami. After the LC and the Association of Research Libraries “complained that they would be unable to pay the cost of changing the headings on cards already in their catalogs if the Committee followed the IFLA Paris Statement of 1961 which called for the entry of all corporate bodies directly under their names,” the Committee “decided to say plainly that the ‘institutions’ rule is an exception to the Paris Statement name-entry principle." Essentially, the parties agreed that entry for corporate bodies could continue under place. Without that agreement (which Dunkin called the “Miami Compromise”), the LC and large research libraries might not have adopted the new code, as it would have created an immense burden of recataloging. In 1963, the committee decided that rules for description should be revised to encompass all media.

These steps toward completion of the new code may have alleviated for some catalogers the weariness with the lengthy process of code revision. As Symons wrote in 1962, "Any cataloging code must be a compromise between the principles of consistency and convenience (but whose convenience? Surely not the cataloger's). There are bound to be areas of conflict. The exact place where the compromise is made seems to me not to matter very much. Rather than waste several more years of time and emotion and inaction, I suggest we encourage the publication of a Revised Code as soon as possible, so that we can all get on with applying in our libraries (or not applying it, if we really dislike it heartily)."

After the Miami Compromise, the Committee (working closely with British and Canadian representatives), labored feverishly to prepare the final edition of the rules, which was published in early 1967 as AACR. Although the rules conformed mostly to Lubetzky's principles, some exceptions were present, particularly those involving corporate entry under place. The committee recorded its regret “that, because of the great size of many American card catalogs, it was necessary for the Catalog Code Revision Committee to agree to the suggestions of the Association of Research Libraries that certain incompatible American practices be continued in the present rules.” Lubetzky himself was disappointed that AACR omitted a statement of principles, on which he had based his draft codes. For the most part, the catalogers accepted the new code and found its revisions worthwhile and useful.

This does not mean that criticism of the cataloging rules ceased in 1967. Indeed, a paper twice this length could be written about the critiques of AACR that led to its revision in 1978.
Conclusion

Most librarians using AACR today entered the profession after the code was published, and therefore accept it as the fundamental basis for cataloging. Further, many librarians believe that cataloging rules have improved over time so that the current rules most closely approach logical, principled cataloging. Nonetheless, AACR was controversial in its day, not least for the major upheaval it caused to previously created catalogs. The process of superimposition followed by many libraries in order to accommodate AACR attests to the wide-ranging consequences of such a thorough revision of cataloging rules.

The historical view of steady improving cataloging codes also feeds the library community’s own self-image as leaders in “modernity, progress, and science.” However, many thoughtful librarians working during the days when AACR was being developed did not necessarily find the principles espoused by Lubetzky to be an improvement over then current practices. Some librarians felt that the values of tradition and user convenience were being disregarded.

An appraisal of the record will show that LRTS served as an important forum for discussing just how, why, and whether catalog code revision would truly make the catalog a better guide to a library’s collection, and that the library community was far from unanimous in regarding AACR as progress. As the cataloging community moves forward with revision of the current catalog code, it would be well-served to examine the values, both stated and unstated, that motivate such revision.

References

13. Ibid., vi.
18. Ibid.
19. Ibid.


34. Ibid., 211.

35. Ibid., 213.

36. Ibid., 217–18.

37. Ibid., 219–20.


45. Ibid., 127–28.


48. Draper, "How to Roll Back the Corporate Empire."

49. Ibid., 81.


53. Ibid., 188.

54. Haskins, "Is Harvard Bound by the Past?" 190.

55. Ibid.


57. Ibid., 208.

58. Ibid., 211.


61. Ibid., 218.

62. Ibid.


69. Paul S. Dunkin, "1963: The Little Year that Wasn’t There,"
70. Ibid.


Comparing Catalogs

Currency and Consistency of Controlled Headings

By Stephen Hearn

Evaluative and comparative studies of catalog data have tended to focus on methods that are labor intensive, demand expertise, and can examine only a limited number of records. This study explores an alternative approach to gathering and analyzing catalog data, focusing on the currency and consistency of controlled headings. The resulting data provide insight into libraries’ use of changed headings and their success in maintaining currency and consistency, and the systems needed to support the current pace of heading changes.

Much of the work of technical services takes place out of public view. Perhaps this explains in part why measures of technical services’ contribution to the library are relatively lacking in compendiums of library measures. The number of volumes and subscriptions in a collection, the rate at which electronic resources are accessed, circulations and reference interviews—all of these are frequently cited as measures of academic libraries’ performance, but rarely is the work contributed directly by technical services used as a library’s performance measure. For some in technical services, there might seem to be an advantage to being “under the radar” when internal or library-to-library comparisons are done; but the lack of measures can also leave any operation unsure of its own success and of the validity of any local or shared set of norms. Having practicable methods of determining a technical services unit’s success in meeting its goals and of assessing that accomplishment in relation to that of peer institutions can help technical services units build confidence in their goals, identify systemic problems, and contribute to library planning and priority setting. The study presented here seeks to define and test an approach to measuring one of the contributions of technical services: the use of consistent and up-to-date headings in the library catalog.

Methods of Measuring Catalog Data Quality

One obvious component of library service is the product of technical services efforts: the data in the library catalog. The catalog assists users with finding known items in the collection; browsing the collection by subject, author, and title headings; browsing the result sets of keyword searches; examining and selecting items via their surrogate records; and locating the items desired. These basic services are provided through a wide variety of interfaces and displays. Vendors and designers of automated library systems offer a range of interface choices to their customers, and each library tailors its system’s functionality and presentation for its users. Comparative evaluation of the differences between such varied interface options would inevitably be complex and highly subjective. In their review of the literature on quality in cataloging, Myall and Chambers note the difficulty and rarity of high-level evaluation of the catalog:
Quality of the overall catalog appears to be less frequently the subject of study, . . . notwithstanding the fact that both Cutter’s objects and much of FRBR’s approach are focused on the catalog as a whole rather than on individual records. Presumably, the limited extent of study at this level is due to the complexity and multi-faceted nature of the task, which now must include not only content and structure of the database, but also completeness and presentation of data on various screens, search engine execution, presence of context-sensitive help, and other elements in an environment in which users are familiar with many other Web-based information tools.¹

Nevertheless, behind the variable screens of automated systems, the data records that feed catalog indexes and displays are highly standardized. The widespread adoption of a core set of data standards by the U.S. academic cataloging community—the Machine-Readable Cataloging (MARC) 21 formats for mark up; the Anglo-American Cataloguing Rules, 2nd ed., rev., for description and name or title access; the Library of Congress’ (LC) Subject Cataloging Manual: Subject Headings rules for subject access; and the LC’s Name and Subject Authority Files (LCNAF and LCSAF, respectively) for authority-controlled headings—has enabled the sharing of catalog records through union databases, of which OCLC WorldCat is the prime exemplar, and the proliferation of library automated systems, all designed in their myriad ways to exploit the data contained in standard catalog records.² Studies of data quality rather than the qualities of automated system interfaces can reasonably claim to be focusing on a crucial and comparable aspect of overall catalog performance.

Past efforts to evaluate catalog data quality have generally relied on thorough review of individual catalog records. In their recent survey of the literature on quality in cataloging, Myall and Chambers found much disagreement over the definition of quality. The most common model they found for data quality analysis, reported in eight studies, calls for selecting a set of catalog records and examining the different areas of each record—fixed fields, standard numbers, title and statement of responsibility, edition and publication statements, notes, and access points—for errors, inconsistencies, and omissions.³ This kind of evaluation tends to be time and labor intensive, requiring a significant level of expertise and often the retrieval of items from the collection for comparison with the catalog records. It also raises comparability issues. Libraries’ standards for what constitutes an acceptable level of data quality and completeness vary across these several areas of description and access. For example, the extent to which libraries invest in the creation of table of contents or summary notes can vary greatly and are a matter of local policy; the trade-off between the added value of such notes and the added liability they represent as additional opportunities for error makes standard measurement difficult.

An alternative for the study of catalog data quality is to look not at a sample set of records, but at a sample set of searchable data. The “Dirty Database Test” takes this approach, offering libraries a set of typographical errors to search in the catalog.⁴ The number of errors thus found does provide a measure of data quality; however, this measure tends to lack both context and focus. The prescribed typographical error search looks for one or one set (using truncation) of erroneous variants, and ignores the number of times the term in all its variants is spelled correctly in the database. In the absence of these other counts, determining an error rate for comparison purposes with other catalogs is difficult. Typographical errors, some of which may be unlikely objects of a searcher’s query, can occur in any term in the record. Lastly, as noted earlier, libraries that include tables of contents and other notes in their records are likely to increase error counts by this measure without regard to the overall enhancement of access that such notes represent. Searching for and correcting typographical errors is an important part of maintaining access, as demonstrated by a 2007 study by Beall and Kafadar, but less effective as a comparative measure.⁵

### Heading Consistency as a Comparative Measure

The aim of the current study is to explore another alternative for evaluating catalog data quality. Rather than looking at whole records for a broad range of error types or searching for typographical errors, this study focuses on the consistency of selected authority controlled headings. Wolverton reports that a commitment to heading consistency as a goal is widespread among technical services departments.² In describing their use of the whole-record review or “audit technique” at University of Bath, Chapman and Massey observe, “Authority control is a valuable form of quality assurance which the audit technique is weak in evaluating, compared to checking descriptive cataloguing.”⁷ The authors go on to note that their pilot study “was unable to confirm the feasibility of comparing headings to an authority file, which would inevitably increase the time required.”⁸

Departing from the whole-record audit approach to focus instead on currency and consistency of controlled headings has a number of advantages as a comparative quality measure. It highlights heading data, which is of high value for discovery. It is less prone to differences over catalogers’ judgment regarding how a particular resource
should be described. It is able to sample a much wider range of records than the audit technique. Lastly, it offers a point of comparison that most if not all catalog managers would acknowledge is valid, given their widespread commitment to maintaining current and consistent headings.

Consistent headings are necessary to ensure that users can find all the items they seek under one heading. Many heading inconsistencies are obvious to the alert user in a browse search, where two similar headings for the same name may appear adjacent to each other in the index display; in other cases, where the entry term for a name or subject heading changes, the two headings may be widely separated in the browse index.

Inconsistencies can be less obvious, and therefore more of a problem, in other searching contexts. A keyword search may retrieve one of the split heading’s forms in its result set, and miss the other. A redirected search, prompted by clicking a highlighted access point in a catalog record’s display, may find only the records that match that access point exactly and exclude the variants. A “faceted browse” display of the type exemplified by North Carolina State University’s implementation of Endeca and Ex Libris’ Primo analyzes headings and other data present in a result set and presents them under such facet headings as “Authors” and “Subjects” in order of frequency in the result set. Such an analysis may find both current and obsolete heading forms in its result set but display only the more frequently used heading in its truncated list of facet terms. Because libraries’ newest items are the likeliest to carry the first instance of the revised form of a name or subject term, they are more likely to be low-posted and therefore less obvious in a faceted browse display, and possibly omitted altogether. Redirected searches are similarly prone to finding only records carrying the obsolete form of a heading and missing new resources with the current form of the heading if the heading split has not been corrected.

In addition to having a clear effect on the service the catalog provides, heading inconsistencies are relatively easy to recognize. Shared standards for heading forms are already in place. Standard forms of headings are widely held goals and data standards. The criteria of comparison that most if not all catalog managers would acknowledge is valid, given their widespread commitment to using the latest form of an authorized heading, since that provides all contributing libraries with a common standard. For that reason, this study examines both these aspects of heading consistency: the rate at which heading “splits” (headings found in both old and new forms) occur in a single catalog; and the extent to which the new form is found to have replaced the old across a set of catalogs. By focusing on the data in the source records, interface variability can be ignored in favor of measuring adherence to commonly held goals and data standards.

Research Method

Using the two identified sources, the LCSH “Weekly Lists” and the OCLC lists of “Closed Dates in Authority Records” for personal name headings, three sample sets of revised subject and name headings were compiled. The source lists are posted weekly, and the sample sets were drawn.
from three separate starting points within each series, both to broaden the range of heading changes gathered and to reveal any changes over time in the updating of headings in the catalogs under review. The name sample sets were collected from lists spaced approximately six months apart. The subject samples were collected from lists spaced approximately a year apart. In each case, the changed headings in the lists were reviewed from the chosen starting point forward in the list or lists until approximately fifty sample headings had been collected. For subjects, several lists were reviewed for each sample of fifty changed headings, since the number of changed subject headings in each LCSH weekly list is fairly small (approximately seven to nine). For names, each OCLC list examined exceeded the required number of fifty changed headings. With each new list sampled, the alphabetical end point of the previous sample set was used as the starting point for assembling the next set of fifty changed name headings.

Once the lists were assembled, a set of target public catalogs was selected. The home for this study is the University of Minnesota, which belongs to the Committee on Institutional Cooperation (CIC), a group of twelve large research universities. The CIC universities have a history of cooperation, including the maintenance of a federated search of CIC catalogs for a time, and are often used as peer institutions for comparison purposes. The thirteen catalogs of the CIC libraries (counting the library catalogs of the Chicago and Urbana-Champaign campuses of the University of Illinois as separate catalogs) were therefore selected to test the method being explored. Because this study is exploratory, the names of the catalogs studied have been randomly ordered and replaced by A, B, C, and so on, in the results. The LC’s public catalog was also included in the set of target catalogs, since it is the source for many of the changed headings. Because the LC’s results show a significantly greater use of the headings under review and therefore make its identity obvious, the LC’s results have been labeled.

All study data were gathered by searching the public catalogs of the target institutions. None of it depended on privileged access to information. The old and new forms of each of the sample headings were browse searched in each target public catalog, and the number of hits found for the old and new forms of the heading was recorded in a spreadsheet, the primary tool for data gathering and manipulation. Spreadsheet formulas were used to calculate the percentage of new headings found. In the spreadsheet, “0%” indicates that only instances of the old heading were found; “100%” indicates that only instances of the new heading were found; and any percentage in between indicates a split between the old and new forms. Where no use of either form was found, “NA” (not applicable) was substituted for a percentage. A sample of the spreadsheet appears in table 1. The project spreadsheets (without institution names) have been deposited in the University of Minnesota’s institutional repository, where they are available for external review.10

For the subject samples, separate counts were made of unsubdivided and subdivided forms of the old and new heading. For the name samples, separate counts were collected under the old and new form for the name as an author and as a subject. These refinements to the counting were made to assess whether heading consistency in the catalogs studied differs for authorized heading strings (the unsubdivided subject strings) versus unauthorized heading strings (most subdivided subjects) and for personal name headings in author indexes versus personal name headings in subject indexes. In expressing the results of the study for this paper, the counts of unsubdivided and subdivided subject heading forms have been merged to show a single count of old versus new main heading forms for the subject index as a whole. The results of the personal name searches are reported separately for name indexes and subject indexes.

Some deselection of the headings initially found in the sampling process proved advisable. Several types of headings were excluded:

1. Headings with old and new forms that would normalize and file identically, e.g., headings changed to remove a hyphen or to correct diacritics, capitalization, or tagging. These differences would be difficult to discern in index displays and, in any case, appear unlikely to affect access.
2. Main headings appearing multiple times with different subdivisions or with different phrase extensions (e.g., “. . . in art”). A few instances of this were retained in the study to explore whether the presence of an established heading-plus-subdivision string in the authority file accounted for a higher rate of correction. The minimal results from the few cases in the study sample suggest not, but are far from conclusive.
3. Headings with more than two forms, e.g., those that changed again following the “new” form’s establishment, or those that merged two earlier headings. The presence of multiple forms would require exceptional forms of counting. Given the relative rarity of instances like this, the few encountered were generally omitted. (An exception: the more heavily used older form “Hog cholera” was counted and the alternative “Swine fever” was not, though both were merged in the new heading “Classical swine fever.”)
4. Narrower subject headings merged into broader pre-existing ones, e.g., the formerly established “Middle Ages—History” being merged into “Middle Ages.” Counting the number of changed headings under the preferred form would be impossible in most public catalogs.
5. Headings with an identical form in another heading system (e.g., the same term used by LCSH and MeSH) or MARC tag category (e.g., the same term used as a topical 650 and a genre/form 655 subject heading) also proved problematic. Some catalogs sort these differences into different indexes, but not all; if they appear in the same index, counting the instances of the changed bibliographic headings becomes difficult, requiring a record-by-record examination of MARC tag values.

The need for sensitivity to these kinds of problems makes the compiling of the sample sets a task that requires a cataloger’s expertise. However, once the list is compiled, the process of searching each form and counting hits can be learned quickly and requires few judgment calls. The method used here was further simplified by regarding only the established old and new heading forms on the list. Other forms were occasionally observed in the indexes (e.g., headings with typographical errors), which also caused split headings, but these were not counted. The only heading splits reported are between instances of the two forms on the sample list.

The data were gathered between January 2 and February 8, 2008, roughly one year after the date of the most recent heading change list sampled. When this task was complete, summary counts were made for each of the catalogs, showing for each set how many of the sample headings were found to be all old form, all new form, split between the old and new form, or unused in each catalog.

Table 1. Changed LC Subject Headings—Data Sample (Excerpts) for One Catalog

<table>
<thead>
<tr>
<th>Cancelled Term</th>
<th>New Term</th>
<th>List Year and No.</th>
<th>No. of Uses of Old LC Heading</th>
<th>With Subdivision</th>
<th>No. of Uses of New LC Heading</th>
<th>With Subdivision</th>
<th>Percent Using New LC Heading</th>
<th>Percent with New Subdivision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baldwin Hills (Calif.)</td>
<td>Baldwin Hills (Calif.: Mountains)</td>
<td>2005.1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Breast feeding</td>
<td>Breastfeeding</td>
<td>2005.1</td>
<td>0</td>
<td>0</td>
<td>203</td>
<td>323</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Breast feeding—Immunological aspects</td>
<td>Breastfeeding—Immunological aspects</td>
<td>2005.1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Breast feeding—Law and legislation</td>
<td>Breastfeeding—Law and legislation</td>
<td>2005.1</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>7</td>
<td>100</td>
<td>91</td>
</tr>
<tr>
<td>Church of England. Book of common prayer. Psalter</td>
<td>Church of England. Psalter</td>
<td>2005.1</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>North Shore (Mass.)</td>
<td>North Shore (Mass.: Coast)</td>
<td>2005.1</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>1</td>
<td>NA</td>
<td>10</td>
</tr>
<tr>
<td>United States Highway 58</td>
<td>United States Highway 58 (Va. and Tenn.)</td>
<td>2005.1</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Dargari language</td>
<td>Tharrkari language</td>
<td>2005.2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Reparation</td>
<td>Reparation (Criminal justice)</td>
<td>2005.2</td>
<td>0</td>
<td>4</td>
<td>88</td>
<td>703</td>
<td>100</td>
<td>99</td>
</tr>
<tr>
<td>Aranda language</td>
<td>Western Arrernte language</td>
<td>2005.3</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>50</td>
</tr>
<tr>
<td>Black humor (Literature)</td>
<td>Black humor</td>
<td>2005.3</td>
<td>3</td>
<td>0</td>
<td>29</td>
<td>2</td>
<td>91</td>
<td>91</td>
</tr>
</tbody>
</table>
The summary for the sample data in table 1 is illustrated in table 2.

## Results

The summary table data have been expressed as a series of doughnut graphs for quicker comprehension. Each doughnut graph shows proportionally the four states of the sample set headings found in each catalog: all instances in the old form, all instances in the new form, instances split between old and new forms, and no instances in the catalog. The outermost ring represents the sample set of the earliest changed headings, and the innermost ring represents the most recently changed sample (see figures 1–3). This arrangement of the data makes it easier to see patterns over time in the proportions of each institution's sample sets.

Figure 1 represents the states of changed LC subject headings in the target catalogs. Some catalogs (LC, D, F, G, H, I, K) show most of their headings fully converted to the new form, while others (A, B, C, E) are largely unconverted or split for all three sample sets. Changed subjects also account for the largest proportions of heading splits overall when compared with the changed name headings. Catalogs J and L show the least use of the headings studied, while catalogs K and F and the LC's catalog show the greatest use.

Figure 2 represents changed LCNAF personal name headings in author indexes. Changed names show a slightly
higher rate of use across the CIC catalogs than did changed LCSH headings. Catalogs G, I, and K show large proportions of fully converted headings across all three samples, while catalogs C, E, and M are largely unconverted. The mixed results are seen for catalogs A, D, H, J, and L. Each shows one or two of the sample sets represented by each of the rings largely converted, but not all three. Such mixed results are more common for changed names than for subjects, suggesting that different approaches are being taken for these different kinds of maintenance.

Figure 3 represents changed LCNAF personal name headings in subject indexes. Not surprisingly, this set of results shows the lowest use rates, though even here use of less than a quarter of the sample headings is relatively uncommon. Catalogs G, I, and K again stand out as the most fully converted, while catalogs B, C, and E are largely unconverted and A, D, and H show mixed results.

This analysis does not take into account the number of hits found for each heading in each catalog. It looks only at whether the heading is present, whether all matches are on the old or the new form, and whether they are split between the two forms. The raw data, however, do include hit counts for each form of the heading and could support other kinds of analysis. For example, splits can only occur when a heading appears more than once in a catalog. Single appearances ranged from two to ten per catalog in the name samples, averaging five to six, and were always outnumbered by headings appearing multiple times. This indicates that a low count for splits in a particular catalog cannot simply be attributed to single appearance headings that could not be split.

Discussion

An examination of the data gathered prompts three kinds of analysis. The first considers the extent to which the sampled headings were found to be in use in the target catalogs and whether the method chosen to assemble the sample sets of headings was effective. The second responds to the question that prompted this project—can a study of the state of changed headings in library catalogs provide a useful measure of data quality for those catalogs? The last takes up two more general questions—are the true states of headings in
our catalogs sufficiently consistent with common models of data quality; and are there changes to the systems involved in revising headings that could simplify the task of heading maintenance?

**Headings**

The sample headings used in this study are listed in appendices A and B. Each list shows both of the forms searched for each heading, the list from which they were drawn, and the number of catalogs out of the thirteen CIC catalogs surveyed in which the heading was found. The entries in the list have been sorted in order of their frequency of occurrence.

One question being explored by this study was whether the sample-generating method used would yield headings that produced useful data. Many name and subject headings occur only infrequently even in large catalogs, and selecting headings without regard for their narrowness or obscurity might have resulted in very few hits. Looking only at the CIC catalogs, twenty-one out of the 147 subject heading pairs checked were unused. At the same time, thirty-six were found in all thirteen CIC catalogs, including twelve from each of the three sample lists. Subject headings appearing in at least half of the target catalogs account for 50 percent of the full subject sample list. In the case of name headings, fifteen out of 155 name pairs sampled were unused, while fifty were found in all thirteen CIC catalogs, including at least twelve in each sample set. Name pairs used in at least half of the CIC catalogs accounted for 68 percent of the name sample as a whole. The investigator's ability to discern which name headings would have no or few hits was minimal. Subject headings proved a bit more predictable. Headings for geographic features had very low hit counts, as did headings reflecting narrow ethnicities or nationalities and headings for uncommon species.

An alternative approach to selecting the target headings would be to have a threshold count in one or more test catalogs that each candidate heading would have to meet. This would provide a larger volume of data for determining the use of old versus new forms in the target catalogs. Nevertheless, the samples used have generated sufficient evidence to provide useful data for this exploratory study. Furthermore, they demonstrate that any large catalog comparable to those studied is likely to have a number of heading changes to attend to in every sequence of fifty changed headings appearing in these Web-distributed lists. The heading changes being promulgated in these lists are having a constant effect on the currency of headings in library catalogs. In the case of the most widely found headings, the effect is often also significant, with some heading changes affecting hundreds of bibliographic records.

**Catalogs**

The data gathered by the study and represented in the graphs in figures 1–3 indicate each catalog's performance against the measures being explored, heading currency and heading splits. Catalogs G, I, and K show a high proportion of consistently used new heading forms across all the sample sets. Catalogs B, C, E, and M show a predominance of consistent use of older forms. In some cases, the message is mixed; e.g., Catalogs A and D show a high proportion of consistently new headings in the oldest name sample, while consistently old forms still predominate in the more recent sample sets. Larger proportions of split headings were found in those catalogs with larger proportions of old heading forms—Catalogs A, B, C, and E. This indicates that the greatest reductions in split headings are achieved in catalogs that also show the greatest success in updating headings to their newer forms.

The purpose of this exploratory study was primarily to develop a method that can demonstrate significant differences between catalogs and thereby provide a useful measure of performance. The study was not designed to explain these differences. However, a number of factors can be suggested.

Some automated systems provide more efficient functionality than others for automatically updating authorized headings; however, no system in use at more than one CIC institution was found to correlate consistently with more current or more consistent headings. The extent to which any library is able to exploit its system's helpful features can vary depending on the availability of staff time and expertise and the press of other significant priorities. The inclusion in a library's system of current authority records and access to a vendor's authority processing service also might be factors in explaining the differences found between catalogs.

Many catalogs are subject to influxes of older or otherwise problematic catalog records, e.g., when the records for a microfilm set or a retrospective conversion project are batchloaded or when records from a foreign vendor are loaded for acquisitions purposes. The fact that the present survey was carried out over a limited period of time may have meant that some catalogs were reviewed at a “bad time.” Repeating the data gathering exercise for the sample headings used here at a later date to determine how the proportions of new, old, and split headings might have shifted would be an interesting exercise; though once this paper is published, it may itself have an effect on the state of this particular set of sample headings. Repeating the study with a new sample set of changed headings could amplify or correct impressions left by the current study.

In any case, the data from this study do support the notion that maintaining current, authorized, consistent
headings in the library catalog is an achievable goal. None of the catalogs performed perfectly in this regard, but perfection is a Procrustean standard of measurement. The new appearance of old headings and of heading splits is a constant in catalog management, and can never be eliminated. Realistically, the goal should be to keep heading currency high and heading splits within tolerable limits, as determined first by each library’s policies and goals and then by an awareness of what peer institutions are achieving. The study results support this kind of comparative judgment and goal setting by revealing differences between catalogs and illustrating the relative success of some catalogs—here catalogs G, I, and K—in responding to the challenge of maintaining current and consistent headings. Catalog data quality should not be taken for granted. The variation observable across catalogs in a performance measure—heading currency and consistency—which is essential to interoperability and uniform search results, highlights both the need for greater effort and realistic benchmarks for success.

**Systems**

The term “systems” here refers broadly to the complex of rules, technology, and practice that governs the management of catalog headings. It does not refer simply or even primarily to integrated library systems.

Many heading changes do not involve any change in the definition of the entity or concept named. The name headings “Abbey, Edward, 1927–” and “Abbey, Edward, 1927–1989” represent the same entity, with or without a death date. “Breast feeding” and “Breastfeeding” represent the same concept, with or without the space. As long as what the authority record names does not change, changes to the name itself are easily managed, at least in principle. However, rules limiting the types of references permitted on LC authority records have made this situation more complex. LC policy currently does not allow the retention of the older, open-dated personal name heading form as a reference on LCNAF records when a death date is added. Similarly, no reference from an earlier LCSH subject heading form is allowed in some cases when LCSAF headings are updated. In January 2007 the MARC Advisory Committee and ALA’s MARBI (Machine-Readable Bibliographic Information) Committee passed Proposal No. 2007-02, which introduces new coding to enable the inclusion of these kinds of references. The proposal was approved in May 2007 by the LC, Library and Archives Canada, and the British Library, though no implementation plans or dates have been announced (as of October 15, 2008). Implementing the proposed changes could lead to simpler and more standardized online system functions for automated maintenance of changed headings.

In other cases the constancy of the definition of the changed heading is more problematic. Subject headings for open-ended periods in a country’s history may have been used on bibliographic records for events that fall beyond a later-assigned closing date for the period. When the heading change is not one-for-one, each instance of the older form needs to be evaluated to ensure a correct revision. In many of the catalogs studied, including the LC’s, closing historical period headings—e.g., changing “Cuba—History—1959–” to either “Cuba—History—1959–1990” or “Cuba—History—1990–” —lagged behind other types of subject heading maintenance. These kinds of changes will resist automated solutions and account in part for the larger number of split headings found for changed subject headings than for changed name headings. Developing more automated means for managing routine one-for-one heading changes would enable more staff time to be focused on those changes that require intellectual decisions.

Lastly, the inefficiencies inherent in maintaining catalog headings across multiple distributed catalogs could be addressed at the systems level. The more libraries can share a single record for bibliographic access, the fewer the records that will need to be maintained. This potential for increased efficiency is one of the motivations behind the current interest in OCLC’s development of WorldCat Local, a catalog model that filters widely shared and maintained OCLC bibliographic records against each record’s holding institutions to provide distributed access to local collections from a centralized database. If sufficient functionality can be built into this model to make it competitive with more conventional library systems, its advantages in terms of shared data management could be significant.

**Conclusion**

The method tested in this study bears out the hypothesis that examining headings across library catalogs for currency and consistency can produce quantified, comparable results, and can serve as one useful measure of catalog data quality. A study of this kind can indicate how well a library’s catalog is performing in relation to locally established goals and to the catalogs of peer institutions, and it can indicate areas needing greater attention. The results also indicate that heading changes in the LCNAF and LCSAF are having negative as well as positive effects on catalog performance. Maintaining catalog headings is a constant challenge, and not one that is being universally or consistently met.

The proposed method for data gathering and analysis could be improved upon in several ways. Better methods might be devised for selecting sample heading sets to reduce the number of unposted or rarely posted headings.
A more nuanced analytical approach might factor in the number of bibliographic records containing new versus old headings found in each catalog to produce a more balanced accounting of heading currency. Other sources of heading variations—e.g., typographical errors or unexplained variant forms—could be included to give a more complete measure of the occurrence of heading splits in the target catalogs.

More research into explanatory factors could also prove valuable. Are there common elements in the technical services operations or system implementations of those libraries that do well on this measure? Do apparent patterns in the heading currency of particular catalogs reflect changes in policies or procedures and their effect? Would closer attention to the types of changed headings that do and do not receive prompt maintenance attention suggest alternative ways of distributing this work?

Lastly, the fundamental question behind this study remains unanswered—should the library catalog’s data quality be evaluated as an outcome measure of the work of technical services? This study has attempted to demonstrate a practical method for such measurement, but it cannot answer the question of whether such measurement should be undertaken by a library or a group of libraries or included in models of library evaluation. Further discussion of that question would also be enlightening.

References and Notes


8. Ibid.

9. CIC members: University of Chicago, University of Illinois, Indiana University, University of Iowa, University of Michigan, Michigan State University, University of Minnesota, Northwestern University, Ohio State University, Penn State University, Purdue University, University of Wisconsin–Madison. For further information, see www.cic.uiuc.edu.

10. Name headings spreadsheet: conservancy.umn.edu/handle/37367; Subject headings spreadsheet: conservancy.umn.edu/handle/37371.

### Appendix A. Changed LC Subject Headings, Sorted by Frequency of Use in CIC Catalogs

<table>
<thead>
<tr>
<th>Cancelled Term</th>
<th>New Term</th>
<th>List No.</th>
<th>Catalogs Using Headings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angels (Judaism)</td>
<td>Angels—Judaism</td>
<td>2005.3</td>
<td>13</td>
</tr>
<tr>
<td>Black humor (Literature)</td>
<td>Black humor</td>
<td>2005.3</td>
<td>13</td>
</tr>
<tr>
<td>Breast feeding</td>
<td>Breastfeeding</td>
<td>2005.1</td>
<td>13</td>
</tr>
<tr>
<td>Breast feeding—Immunological aspects</td>
<td>Breastfeeding—Immunological aspects</td>
<td>2005.1</td>
<td>13</td>
</tr>
<tr>
<td>Calligraphy, Islamic</td>
<td>Islamic calligraphy</td>
<td>2005.5</td>
<td>13</td>
</tr>
<tr>
<td>Calligraphy, Zen</td>
<td>Zen calligraphy</td>
<td>2005.5</td>
<td>13</td>
</tr>
<tr>
<td>Children’s web sites</td>
<td>Web sites for children</td>
<td>2007.1</td>
<td>13</td>
</tr>
<tr>
<td>Crimes against humanity, German</td>
<td>Crimes against humanity—Germany</td>
<td>2007.4</td>
<td>13</td>
</tr>
<tr>
<td>Cross Florida Barge Canal (Fla.)</td>
<td>Marjorie Harris Carr Cross Florida Greenway (Fla.)</td>
<td>2007.5</td>
<td>13</td>
</tr>
<tr>
<td>Cuba—Foreign relations—1959–</td>
<td>Cuba—Foreign relations—1959–1990</td>
<td>2006.3</td>
<td>13</td>
</tr>
<tr>
<td>Definition (Logic)</td>
<td>Definition (Philosophy)</td>
<td>2007.4</td>
<td>13</td>
</tr>
<tr>
<td>Friendly societies</td>
<td>Fraternal organizations</td>
<td>2006.7</td>
<td>13</td>
</tr>
<tr>
<td>Friendly societies—United States</td>
<td>Fraternal organizations—United States</td>
<td>2006.7</td>
<td>13</td>
</tr>
<tr>
<td>Gutters</td>
<td>Roof gutters</td>
<td>2005.4</td>
<td>13</td>
</tr>
<tr>
<td>Hog cholera—Vaccination</td>
<td>Classical swine fever—Vaccination</td>
<td>2007.9</td>
<td>13</td>
</tr>
<tr>
<td>Hog cholera</td>
<td>Classical swine fever</td>
<td>2007.9</td>
<td>13</td>
</tr>
<tr>
<td>Insurance, Unemployment—Claimants</td>
<td>Unemployment insurance claimants</td>
<td>2007.2</td>
<td>13</td>
</tr>
<tr>
<td>Islam and terrorism</td>
<td>Terrorism—Religious aspects—Islam</td>
<td>2005.2</td>
<td>13</td>
</tr>
<tr>
<td>Kennebec Patent</td>
<td>Kennebec Patent (Me.)</td>
<td>2007.9</td>
<td>13</td>
</tr>
<tr>
<td>Knizhnik-Zamolodchik equations</td>
<td>Knizhnik-Zamolodchik equations</td>
<td>2006.3</td>
<td>13</td>
</tr>
<tr>
<td>Lady and the Unicorn (Tapestries)</td>
<td>Lady and the Unicorn</td>
<td>2005.4</td>
<td>13</td>
</tr>
<tr>
<td>Online data processing—Downloading</td>
<td>Downloading of data</td>
<td>2006.1</td>
<td>13</td>
</tr>
<tr>
<td>Path analysis</td>
<td>Path analysis (Statistics)</td>
<td>2006.4</td>
<td>13</td>
</tr>
<tr>
<td>Reciprocity</td>
<td>Reciprocity (Commerce)</td>
<td>2007.8</td>
<td>13</td>
</tr>
<tr>
<td>Reparation</td>
<td>Reparation (Criminal justice)</td>
<td>2005.2</td>
<td>13</td>
</tr>
<tr>
<td>Student loan funds</td>
<td>Student loans</td>
<td>2006.1</td>
<td>13</td>
</tr>
<tr>
<td>Swastika</td>
<td>Swastikas</td>
<td>2005.2</td>
<td>13</td>
</tr>
<tr>
<td>Wages—Gas industry employees</td>
<td>Wages—Gas industry</td>
<td>2007.1</td>
<td>13</td>
</tr>
<tr>
<td>Weblogs</td>
<td>Blogs</td>
<td>2006.8</td>
<td>13</td>
</tr>
<tr>
<td>Ak Koyunlus (Turkic people)</td>
<td>Ak Koyunlu (Turkic people)</td>
<td>2006.5</td>
<td>12</td>
</tr>
<tr>
<td>Art, Papua New Guinea</td>
<td>Art, Papua New Guinean</td>
<td>2006.8</td>
<td>12</td>
</tr>
<tr>
<td>New Zealand—History—Maori War, 1845–1847</td>
<td>New Zealand—History—New Zealand War, 1843–1847</td>
<td>2005.2</td>
<td>12</td>
</tr>
<tr>
<td>Papua New Guinea literature (English)</td>
<td>Papua New Guinea literature (English)</td>
<td>2006.8</td>
<td>12</td>
</tr>
<tr>
<td>Usury (Islamic law)</td>
<td>Interest (Islamic law)</td>
<td>2005.5</td>
<td>12</td>
</tr>
<tr>
<td>Amendments (Parliamentary practice)</td>
<td>Legislative amendments</td>
<td>2005.8</td>
<td>11</td>
</tr>
<tr>
<td>Friendly societies—Law and legislation</td>
<td>Fraternal organizations—Law and legislation</td>
<td>2006.7</td>
<td>11</td>
</tr>
<tr>
<td>Harmonica and electronic music</td>
<td>Electronic and harmonica music</td>
<td>2005.5</td>
<td>11</td>
</tr>
<tr>
<td>Karen language</td>
<td>Karen languages</td>
<td>2007.9</td>
<td>11</td>
</tr>
<tr>
<td>Lullabies, American</td>
<td>Lullabies, English—United States</td>
<td>2006.1</td>
<td>11</td>
</tr>
<tr>
<td>Pallavas</td>
<td>Pallava dynasty, 4th–9th centuries</td>
<td>2007.9</td>
<td>11</td>
</tr>
<tr>
<td>Cancelled Term</td>
<td>New Term</td>
<td>List No.</td>
<td>Catalogs Using Heads</td>
</tr>
<tr>
<td>------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>----------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Tabla and sitar music</td>
<td>Sitar and tabla music</td>
<td>2007.8</td>
<td>11</td>
</tr>
<tr>
<td>Uttar Khand Region (India)</td>
<td>Uttaranchal (India)</td>
<td>2005.2</td>
<td>11</td>
</tr>
<tr>
<td>Abelam (Papua New Guinea people)</td>
<td>Abelam (Papua New Guinean people)</td>
<td>2006.8</td>
<td>10</td>
</tr>
<tr>
<td>Akademicheskii malii dramaticheskii teatr (Saint Petersburg, Russia)</td>
<td>Akademicheskii malyi dramaticheskii teatr (Saint Petersburg, Russia)</td>
<td>2005.4</td>
<td>10</td>
</tr>
<tr>
<td>Angels (Islam)</td>
<td>Angels—Islam</td>
<td>2005.3</td>
<td>10</td>
</tr>
<tr>
<td>Canons, fugues, etc. (Voice)</td>
<td>Canons, fugues, etc. (Voices)</td>
<td>2006.5</td>
<td>10</td>
</tr>
<tr>
<td>Electronic and harpsichord music</td>
<td>Harpsichord and electronic music</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Online data processing—Uploading</td>
<td>Uploading of data</td>
<td>2006.4</td>
<td>10</td>
</tr>
<tr>
<td>Siddhi (Indic people)</td>
<td>Siddi (Indic people)</td>
<td>2007.4</td>
<td>10</td>
</tr>
<tr>
<td>Yog language</td>
<td>Bouyei language</td>
<td>2007.2</td>
<td>10</td>
</tr>
<tr>
<td>Anostraca</td>
<td>Fairy shrimps</td>
<td>2007.1</td>
<td>9</td>
</tr>
<tr>
<td>Aranda language</td>
<td>Western Arrernte language</td>
<td>2005.3</td>
<td>9</td>
</tr>
<tr>
<td>Banda language</td>
<td>Banda language (Central Africa)</td>
<td>2007.9</td>
<td>9</td>
</tr>
<tr>
<td>Breast feeding—Law and legislation</td>
<td>Breastfeeding—Law and legislation</td>
<td>2005.1</td>
<td>9</td>
</tr>
<tr>
<td>Congridae</td>
<td>Conger eels</td>
<td>2006.4</td>
<td>9</td>
</tr>
<tr>
<td>Creeper lanes</td>
<td>Climbing lanes</td>
<td>2007.1</td>
<td>9</td>
</tr>
<tr>
<td>Midea Site (Greece)</td>
<td>Midea (Extinct city)</td>
<td>2005.4</td>
<td>9</td>
</tr>
<tr>
<td>Dargari language</td>
<td>Tharrkari language</td>
<td>2005.2</td>
<td>8</td>
</tr>
<tr>
<td>Karakoyunlus</td>
<td>Kara Koyunlu (Turkic people)</td>
<td>2006.5</td>
<td>8</td>
</tr>
<tr>
<td>Mesaras Plain (Greece)</td>
<td>Mesara Plain (Greece)</td>
<td>2005.2</td>
<td>8</td>
</tr>
<tr>
<td>Papua New Guinea fiction (English)</td>
<td>Papua New Guinean fiction (English)</td>
<td>2006.8</td>
<td>8</td>
</tr>
<tr>
<td>Saint Martin—Description and travel</td>
<td>Saint Martin (West Indies)—Description and travel</td>
<td>2006.3</td>
<td>8</td>
</tr>
<tr>
<td>Sculpture, Kota</td>
<td>Sculpture, Kota (Africa)</td>
<td>2006.4</td>
<td>8</td>
</tr>
<tr>
<td>Dolgan dialect</td>
<td>Dolgan language</td>
<td>2007.8</td>
<td>7</td>
</tr>
<tr>
<td>Karts (Midget cars)</td>
<td>Karts ( Automobiles)</td>
<td>2006.5</td>
<td>7</td>
</tr>
<tr>
<td>Daba language</td>
<td>Daba language (Cameroon and Nigeria)</td>
<td>2007.4</td>
<td>6</td>
</tr>
<tr>
<td>Little League World Series, Williamsport, Pa.</td>
<td>Little League World Series (Baseball)</td>
<td>2005.4</td>
<td>6</td>
</tr>
<tr>
<td>Marriage (Luo law)</td>
<td>Marriage (Luo (Kenya and Tanzania) law)</td>
<td>2007.2</td>
<td>6</td>
</tr>
<tr>
<td>North Shore (Mass.)</td>
<td>North Shore (Mass. : Coast)</td>
<td>2005.1</td>
<td>6</td>
</tr>
<tr>
<td>Art, Parsic</td>
<td>Parsee art</td>
<td>2007.7</td>
<td>5</td>
</tr>
<tr>
<td>Back River (Nunavut)</td>
<td>Back River (N.W.T. and Nunavut)</td>
<td>2005.5</td>
<td>5</td>
</tr>
<tr>
<td>Back River Valley (Nunavut)</td>
<td>Back River Valley (N.W.T. and Nunavut)</td>
<td>2005.5</td>
<td>5</td>
</tr>
<tr>
<td>Caernarvon Castle (Caernarvon, Wales)</td>
<td>Caernarvon Castle (Caernarvon, Wales)</td>
<td>2006.4</td>
<td>5</td>
</tr>
<tr>
<td>Calligraphy, Buddhist</td>
<td>Buddhist calligraphy</td>
<td>2005.5</td>
<td>4</td>
</tr>
<tr>
<td>Calligraphy, Islamic, in art</td>
<td>Islamic calligraphy in art</td>
<td>2005.5</td>
<td>4</td>
</tr>
<tr>
<td>Calligraphy, Taoist</td>
<td>Taoist calligraphy</td>
<td>2005.5</td>
<td>4</td>
</tr>
<tr>
<td>Castel Roncolo (Bolzano, Italy)</td>
<td>Castel Roncolo (Bolzano, Trentino-Alto Adige, Italy)</td>
<td>2006.8</td>
<td>4</td>
</tr>
<tr>
<td>Cookery, Papua New Guinea</td>
<td>Cookery, Papua New Guinean</td>
<td>2006.8</td>
<td>4</td>
</tr>
<tr>
<td>Hog cholera—Diagnosis</td>
<td>Classical swine fever—Diagnosis</td>
<td>2007.9</td>
<td>4</td>
</tr>
<tr>
<td>Mass media in breast feeding promotion</td>
<td>Mass media in breastfeeding promotion</td>
<td>2005.1</td>
<td>4</td>
</tr>
<tr>
<td>Alfures (New Guinea people)</td>
<td>Alfures (New Guinean people)</td>
<td>2006.8</td>
<td>3</td>
</tr>
<tr>
<td>Emirian periodicals</td>
<td>Emirati periodicals</td>
<td>2007.3</td>
<td>3</td>
</tr>
<tr>
<td>Frake family</td>
<td>Frakes family</td>
<td>2007.7</td>
<td>3</td>
</tr>
<tr>
<td>Papua New Guinea drama (English)</td>
<td>Papua New Guinean drama (English)</td>
<td>2006.8</td>
<td>3</td>
</tr>
<tr>
<td>Pochard</td>
<td>Common pochard</td>
<td>2005.4</td>
<td>3</td>
</tr>
<tr>
<td>Public interest (Islamic law)</td>
<td>Istitilah (Islamic law)</td>
<td>2006.8</td>
<td>3</td>
</tr>
<tr>
<td>Rajbangsi dialect</td>
<td>Rajbangsi language</td>
<td>2007.8</td>
<td>3</td>
</tr>
<tr>
<td>Sigaw Karen dialect</td>
<td>Sigaw Karen language</td>
<td>2007.9</td>
<td>3</td>
</tr>
<tr>
<td>Angels (Buddhism)</td>
<td>Angels—Buddhism</td>
<td>2005.3</td>
<td>2</td>
</tr>
<tr>
<td>Croatia—History—Zrinski-Francopan Conspiracy, 1664–1671</td>
<td>Zrinski-Francopan Conspiracy, Croatia, 1664–1671</td>
<td>2007.7</td>
<td>2</td>
</tr>
<tr>
<td>Cancelled Term</td>
<td>New Term</td>
<td>List No.</td>
<td>Catalogs Using Headings</td>
</tr>
<tr>
<td>---------------</td>
<td>----------</td>
<td>---------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>Gabbard family</td>
<td>Gebhardt family</td>
<td>2007.6</td>
<td>2</td>
</tr>
<tr>
<td>Kayan language</td>
<td>Kayan language (Borneo)</td>
<td>2007.9</td>
<td>2</td>
</tr>
<tr>
<td>Kayu Agung dialect</td>
<td>Kayu Agung language</td>
<td>2007.9</td>
<td>2</td>
</tr>
<tr>
<td>Kittiwake</td>
<td>Kittiwakes</td>
<td>2006.6</td>
<td>2</td>
</tr>
<tr>
<td>Liang Mountains (China)</td>
<td>Liang Mountains (Shandong Sheng, China)</td>
<td>2007.8</td>
<td>2</td>
</tr>
<tr>
<td>Polish American friendly societies</td>
<td>Polish American fraternal organizations</td>
<td>2006.7</td>
<td>2</td>
</tr>
<tr>
<td>Pwo Karen dialect</td>
<td>Pwo Karen language</td>
<td>2007.9</td>
<td>2</td>
</tr>
<tr>
<td>Saint Martin—Antiquities</td>
<td>Saint Martin (West Indies)—Antiquities</td>
<td>2006.3</td>
<td>2</td>
</tr>
<tr>
<td>Victor (Jet planes)</td>
<td>Victor (Jet bomber)</td>
<td>2007.5</td>
<td>2</td>
</tr>
<tr>
<td>Arnica (Drug)</td>
<td>Arnica montana—Therapeutic use</td>
<td>2007.4</td>
<td>1</td>
</tr>
<tr>
<td>Ferto-Hansag Nemzeti Park (Hungary)</td>
<td>Ferto-Hansag Nemzeti Park (Hungary)</td>
<td>2007.9</td>
<td>1</td>
</tr>
<tr>
<td>Flower pots in art</td>
<td>Flowerpots in art</td>
<td>2006.3</td>
<td>1</td>
</tr>
<tr>
<td>Hu Mountain (China)</td>
<td>Hu Mountain (Jiangsu Sheng, China)</td>
<td>2005.5</td>
<td>1</td>
</tr>
<tr>
<td>Hypsiglena ochrorhynchus</td>
<td>Spotted night snake</td>
<td>2006.4</td>
<td>1</td>
</tr>
<tr>
<td>Immoral contracts (Islamic law)</td>
<td>Illegal contracts (Islamic law)</td>
<td>2005.5</td>
<td>1</td>
</tr>
<tr>
<td>Jardins du Prieure de Salagon (Mane, France)</td>
<td>Jardins du Prieure de Salagon (Mane, Provence-Alpes-Cote d’Azur, France)</td>
<td>2006.8</td>
<td>1</td>
</tr>
<tr>
<td>Kornelsen family</td>
<td>Cornelsen family</td>
<td>2007.9</td>
<td>1</td>
</tr>
<tr>
<td>Lewis and Clark Cavern State Park (Mont.)</td>
<td>Lewis and Clark Caverns State Park (Mont.)</td>
<td>2005.6</td>
<td>1</td>
</tr>
<tr>
<td>Mythology, Kota</td>
<td>Mythology, Kota (Africa)</td>
<td>2006.4</td>
<td>1</td>
</tr>
<tr>
<td>Pahute Mesa (Nevada)</td>
<td>Pahute Mesa ( Nev.)</td>
<td>2005.6</td>
<td>1</td>
</tr>
<tr>
<td>Ricinodendron raunenii</td>
<td>Manketti</td>
<td>2005.6</td>
<td>1</td>
</tr>
<tr>
<td>Samo language</td>
<td>Samo language (Western Province, Papua New Guinea)</td>
<td>2007.4</td>
<td>1</td>
</tr>
<tr>
<td>Satluj River (India)</td>
<td>Sutlej River</td>
<td>2007.5</td>
<td>1</td>
</tr>
<tr>
<td>Scottish American friendly societies</td>
<td>Scottish American fraternal organizations</td>
<td>2006.7</td>
<td>1</td>
</tr>
<tr>
<td>Spruce Island (Alaska)</td>
<td>Spruce Island (Kodiak Island Borough, Alaska)</td>
<td>2005.2</td>
<td>1</td>
</tr>
<tr>
<td>Taungthu dialect</td>
<td>Taungthu language</td>
<td>2007.9</td>
<td>1</td>
</tr>
<tr>
<td>Alamblak (Papua New Guinea people)</td>
<td>Alamblak (Papua New Guinean people)</td>
<td>2006.8</td>
<td>0</td>
</tr>
<tr>
<td>Ansitz Rottenbuch (Bolzano, Italy)</td>
<td>Ansitz Rottenbuch (Bolzano, Trentino-Alto Adige, Italy)</td>
<td>2006.8</td>
<td>0</td>
</tr>
<tr>
<td>Atlases, Emirian</td>
<td>Atlases, Emirati</td>
<td>2007.3</td>
<td>0</td>
</tr>
<tr>
<td>Bagore-ki-haveli (Udayapura, India)</td>
<td>Bagore-ki-Haveli (Udaipur, Rajasthan, India)</td>
<td>2005.5</td>
<td>0</td>
</tr>
<tr>
<td>Baldwin Hills (Calif.)</td>
<td>Baldwin Hills (Calif. : Mountains)</td>
<td>2005.1</td>
<td>0</td>
</tr>
<tr>
<td>Bankudu-Balue language</td>
<td>Bakundu-Balue language</td>
<td>2007.3</td>
<td>0</td>
</tr>
<tr>
<td>Congrina</td>
<td>Bathycongrus</td>
<td>2006.4</td>
<td>0</td>
</tr>
<tr>
<td>Economic assistance, Emirian</td>
<td>Economic assistance, Emirati</td>
<td>2007.3</td>
<td>0</td>
</tr>
<tr>
<td>English language—Augment</td>
<td>English language—Augmentatives</td>
<td>2005.2</td>
<td>0</td>
</tr>
<tr>
<td>Hare Island (Ireland)</td>
<td>Hare Island (Cork, Ireland)</td>
<td>2007.6</td>
<td>0</td>
</tr>
<tr>
<td>Lakhra Coal Field (Pakistan)</td>
<td>Lakhra Coalfield (Pakistan)</td>
<td>2006.1</td>
<td>0</td>
</tr>
<tr>
<td>Lullabies, Puerto Rican</td>
<td>Lullabies, Spanish—Puerto Rico</td>
<td>2006.1</td>
<td>0</td>
</tr>
<tr>
<td>Lycopersicon pimpinellifolium</td>
<td>Currant tomato</td>
<td>2007.7</td>
<td>0</td>
</tr>
<tr>
<td>Mist Gas Field (Oregon)</td>
<td>Mist Gas Field (Or.)</td>
<td>2005.6</td>
<td>0</td>
</tr>
<tr>
<td>Pangaimotu Island (Vava’u Group, Tonga)</td>
<td>Pangaimotu Island (Vava’u, Tonga)</td>
<td>2007.4</td>
<td>0</td>
</tr>
<tr>
<td>Peng Chau (China)</td>
<td>Peng Chau (Islands District, China)</td>
<td>2005.3</td>
<td>0</td>
</tr>
<tr>
<td>Proverbs, Yay</td>
<td>Proverbs, Bouyei</td>
<td>2007.2</td>
<td>0</td>
</tr>
<tr>
<td>Sonda Coal Field (Pakistan)</td>
<td>Sonda Coalfield (Pakistan)</td>
<td>2006.1</td>
<td>0</td>
</tr>
<tr>
<td>Songs, Gaviao</td>
<td>Songs, Gaviao (Para, Brazil)</td>
<td>2007.7</td>
<td>0</td>
</tr>
<tr>
<td>Sylarna (Sweden)</td>
<td>Sylarna (Sweden : Mountain)</td>
<td>2005.4</td>
<td>0</td>
</tr>
<tr>
<td>United States Highway 58</td>
<td>United States Highway 58 (Va. and Tenn.)</td>
<td>2005.1</td>
<td>0</td>
</tr>
</tbody>
</table>
### Appendix B. Changed LC Name Headings, Sorted by Frequency of Use in CIC Catalogs

<table>
<thead>
<tr>
<th>Cancelled Name</th>
<th>New Name</th>
<th>OCLC List Date</th>
<th>Catalogs Using Heading</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adams, Norman Isley, 1895–</td>
<td>Adams, Norman Isley, 1895–1985</td>
<td>20060208</td>
<td>13</td>
</tr>
<tr>
<td>Adler, Mortimer Jerome, 1902–</td>
<td>Adler, Mortimer Jerome, 1902–2001</td>
<td>20060208</td>
<td>13</td>
</tr>
<tr>
<td>Alley, Rewi, 1897–</td>
<td>Alley, Rewi, 1897–1987</td>
<td>20060208</td>
<td>13</td>
</tr>
<tr>
<td>Arafat, Yasir, 1929–</td>
<td>Arafat, Yasir, 1929–2004</td>
<td>20060208</td>
<td>13</td>
</tr>
<tr>
<td>Axelrod, Julius, 1912–</td>
<td>Axelrod, Julius, 1912–2004</td>
<td>20060208</td>
<td>13</td>
</tr>
<tr>
<td>Ball, Lucille, 1911–</td>
<td>Ball, Lucille, 1911–1989</td>
<td>20060208</td>
<td>13</td>
</tr>
<tr>
<td>Bancroft, Anne, 1931–</td>
<td>Bancroft, Anne, 1931–1937</td>
<td>20060208</td>
<td>13</td>
</tr>
<tr>
<td>Block, Herbert, 1900–</td>
<td>Block, Herbert, 1900–2001</td>
<td>20060208</td>
<td>13</td>
</tr>
<tr>
<td>Bolitho, Hector, 1898–</td>
<td>Bolitho, Hector, 1898–1974</td>
<td>20060208</td>
<td>13</td>
</tr>
<tr>
<td>Burns, George, 1896–</td>
<td>Burns, George, 1896–1996</td>
<td>20060208</td>
<td>13</td>
</tr>
<tr>
<td>Callaghan, James, 1912–</td>
<td>Callaghan, James, 1912–2005</td>
<td>20060208</td>
<td>13</td>
</tr>
<tr>
<td>Carson, Johny, 1925–</td>
<td>Carson, Johny, 1925–2005</td>
<td>20060208</td>
<td>13</td>
</tr>
<tr>
<td>Cheney, Brainard, 1900–</td>
<td>Cheney, Brainard, 1900–1990</td>
<td>20060208</td>
<td>13</td>
</tr>
<tr>
<td>Chisholm, Shirley, 1924–</td>
<td>Chisholm, Shirley, 1924–2005</td>
<td>20060208</td>
<td>13</td>
</tr>
<tr>
<td>Clark, Kenneth Bancroft, 1914–</td>
<td>Clark, Kenneth Bancroft, 1914–1994</td>
<td>20060208</td>
<td>13</td>
</tr>
<tr>
<td>Hadamowsky, Franz, 1900–</td>
<td>Hadamowsky, Franz, 1900–1995</td>
<td>20070207</td>
<td>13</td>
</tr>
<tr>
<td>Hogben, Lancelot Thomas, 1895–</td>
<td>Hogben, Lancelot Thomas, 1895–1975</td>
<td>20070207</td>
<td>13</td>
</tr>
<tr>
<td>Howes, Raymond F. (Raymond Floyd), 1903–</td>
<td>Howes, Raymond F. (Raymond Floyd), 1903–1986</td>
<td>20070207</td>
<td>13</td>
</tr>
<tr>
<td>La Sale, Antoine de, b. 1388?</td>
<td>La Sale, Antoine de, b. 1385?–1461?</td>
<td>20060201</td>
<td>13</td>
</tr>
<tr>
<td>Landeck, Armin, 1905–</td>
<td>Landeck, Armin, 1905–1984</td>
<td>20070207</td>
<td>13</td>
</tr>
<tr>
<td>Ley, Hermann, 1911–</td>
<td>Ley, Hermann, 1911–1990</td>
<td>20060201</td>
<td>13</td>
</tr>
<tr>
<td>Mantle, Mickey, 1931–</td>
<td>Mantle, Mickey, 1931–1995</td>
<td>20070207</td>
<td>13</td>
</tr>
<tr>
<td>Moore, Barrington, 1913–</td>
<td>Moore, Barrington, 1913–2005</td>
<td>20060816</td>
<td>13</td>
</tr>
<tr>
<td>Onassis, Jacqueline Kennedy, 1929–</td>
<td>Onassis, Jacqueline Kennedy, 1929–1994</td>
<td>20060201</td>
<td>13</td>
</tr>
<tr>
<td>Rufer, Josef, 1893–</td>
<td>Rufer, Josef, 1893–1985</td>
<td>20060816</td>
<td>13</td>
</tr>
<tr>
<td>Salgado, Plinio, 1895–</td>
<td>Salgado, Plinio, 1895–1975</td>
<td>20060816</td>
<td>13</td>
</tr>
<tr>
<td>Salisbury, Harrison Evans, 1908–</td>
<td>Salisbury, Harrison E.(Harrison Evans), 1908–1993</td>
<td>20060816</td>
<td>13</td>
</tr>
<tr>
<td>Tikhonov, Nikolai Semenovich, 1896–</td>
<td>Tikhonov, Nikolai Semenovich, 1896–1979</td>
<td>20060816</td>
<td>13</td>
</tr>
<tr>
<td>Achelis, Elisabeth, 1880–</td>
<td>Achelis, Elisabeth, 1880–1973</td>
<td>20060208</td>
<td>12</td>
</tr>
<tr>
<td>Cancelled Name</td>
<td>New Name</td>
<td>OCLC List Date</td>
<td>Catalogs Using Heading</td>
</tr>
<tr>
<td>------------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>----------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Benenson, Peter, 1921–</td>
<td>Benenson, Peter, 1921–2005</td>
<td>20060208</td>
<td>12</td>
</tr>
<tr>
<td>Bernhard Leopold, Prince, consort of Juliana, 1911–</td>
<td>Bernhard Leopold, Prince, consort of Juliana, 1911–2004</td>
<td>20060208</td>
<td>12</td>
</tr>
<tr>
<td>Bronfenbrenner, Urie, 1917–</td>
<td>Bronfenbrenner, Urie, 1917–2005</td>
<td>20060208</td>
<td>12</td>
</tr>
<tr>
<td>Erskine, Ralph, 1914–</td>
<td>Erskine, Ralph, 1914–2005</td>
<td>20070207</td>
<td>12</td>
</tr>
<tr>
<td>Harley, Robison D. 1911–</td>
<td>Harley, Robison D. 1911–2007</td>
<td>20070207</td>
<td>12</td>
</tr>
<tr>
<td>Hoffmeister, Adolf, 1902–</td>
<td>Hoffmeister, Adolf, 1902–1973</td>
<td>20070207</td>
<td>12</td>
</tr>
<tr>
<td>Hojo, Hideji, 1902–</td>
<td>Hojo, Hideji, 1902–1996</td>
<td>20070207</td>
<td>12</td>
</tr>
<tr>
<td>Luzzati, Emanuele, 1921–</td>
<td>Luzzati, Emanuele, 1921–2007</td>
<td>20070207</td>
<td>12</td>
</tr>
<tr>
<td>Machel, Samara, 1933–</td>
<td>Machel, Samara, 1933–1986</td>
<td>20070207</td>
<td>12</td>
</tr>
<tr>
<td>Simpson, Robert Wilfred Levick, 1921–</td>
<td>Simpson, Robert, 1921–1997</td>
<td>20060816</td>
<td>12</td>
</tr>
<tr>
<td>Edmunds, Murrell, 1898–</td>
<td>Edmunds, Murrell, 1898–1981</td>
<td>20070207</td>
<td>11</td>
</tr>
<tr>
<td>Habermann, Abraham Meir, 1901–</td>
<td>Habermann, Abraham Meir, 1901–1980</td>
<td>20070207</td>
<td>11</td>
</tr>
<tr>
<td>Haksar, P. N. (Parmeshwar Narain), 1913–</td>
<td>Haksar, P. N. (Parmeshwar Narain), 1913–1998</td>
<td>20060201</td>
<td>11</td>
</tr>
<tr>
<td>Lewis, Joseph H., 1907–</td>
<td>Lewis, Joseph H., 1907–2000</td>
<td>20070207</td>
<td>11</td>
</tr>
<tr>
<td>McAvoy, May, 1901–</td>
<td>McAvoy, May, 1901–1984</td>
<td>20070207</td>
<td>11</td>
</tr>
<tr>
<td>Pasternak, Joe, 1901–</td>
<td>Pasternak, Joe, 1901–1991</td>
<td>20060816</td>
<td>11</td>
</tr>
<tr>
<td>Santorsola, Guido, 1904–</td>
<td>Santorsola, Guido, 1904–1994</td>
<td>20060816</td>
<td>11</td>
</tr>
<tr>
<td>Tubb, Ernest, 1914–</td>
<td>Tubb, Ernest, 1914–1984</td>
<td>20060816</td>
<td>11</td>
</tr>
<tr>
<td>Elwood, Muriel, 1902–</td>
<td>Elwood, Muriel, 1902–1976</td>
<td>20070207</td>
<td>10</td>
</tr>
<tr>
<td>Michelin, Bernard, 1918–</td>
<td>Michelin, Bernard, 1915–2003</td>
<td>20060816</td>
<td>10</td>
</tr>
<tr>
<td>Strock, Herbert L., 1918–</td>
<td>Strock, Herbert L., 1918–2005</td>
<td>20060816</td>
<td>10</td>
</tr>
<tr>
<td>Foss, Joe, 1915–</td>
<td>Foss, Joe, 1915–2003</td>
<td>20060201</td>
<td>9</td>
</tr>
<tr>
<td>Kamlleshwar, 1932–</td>
<td>Kamlleshwar, 1932–2007</td>
<td>20070207</td>
<td>9</td>
</tr>
<tr>
<td>Okada, Jo, 1911–</td>
<td>Okada, Jo, 1911–1981</td>
<td>20060816</td>
<td>9</td>
</tr>
<tr>
<td>Rybar, Peter, 1913–</td>
<td>Rybar, Peter, 1913–2002</td>
<td>20060816</td>
<td>9</td>
</tr>
<tr>
<td>Alice, Duchesse of Gloucester, 1901–</td>
<td>Alice, Duchesse of Gloucester, 1901–1904</td>
<td>20060208</td>
<td>8</td>
</tr>
<tr>
<td>Banner, Donald W., 1924–</td>
<td>Banner, Donald W., 1924–2006</td>
<td>20060208</td>
<td>8</td>
</tr>
<tr>
<td>Brown, Clarence, 1924–</td>
<td>Brown, Clarence, 1924–2005</td>
<td>20060208</td>
<td>8</td>
</tr>
<tr>
<td>Dufresne, Don, 1917–</td>
<td>Dufresne, Don, 1917–1993</td>
<td>20070207</td>
<td>8</td>
</tr>
<tr>
<td>Keshet, Yeshurun, 1893–</td>
<td>Keshet, Yeshurun, 1893–1977</td>
<td>20070207</td>
<td>8</td>
</tr>
<tr>
<td>Lishner, Leon, 1913–</td>
<td>Lishner, Leon, 1913–1995</td>
<td>20070207</td>
<td>8</td>
</tr>
<tr>
<td>Moulton, Augustus Freedom, 1848–</td>
<td>Moulton, Augustus Freedom, 1848–1933</td>
<td>20060201</td>
<td>8</td>
</tr>
<tr>
<td>Philips, Frits, 1905–</td>
<td>Philips, Frits, 1905–2005</td>
<td>20060816</td>
<td>8</td>
</tr>
<tr>
<td>Galai, Binyamin, 1921–</td>
<td>Galai, Binyamin, 1921–1995</td>
<td>20070207</td>
<td>7</td>
</tr>
<tr>
<td>McManus, Frederick R. (Frederick Richard), 1923–</td>
<td>McManus, Frederick R. (Frederick Richard), 1923–2005</td>
<td>20060816</td>
<td>7</td>
</tr>
<tr>
<td>Cancelled Name</td>
<td>New Name</td>
<td>OCLC List Date</td>
<td>Catalogs Using Heading</td>
</tr>
<tr>
<td>----------------</td>
<td>----------</td>
<td>----------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Helou, Charles, 1912–</td>
<td>Helou, Charles, 1912–2001</td>
<td>20070207</td>
<td>6</td>
</tr>
<tr>
<td>Hunter, Joe, 1927–</td>
<td>Hunter, Joe, 1927–2001</td>
<td>20070207</td>
<td>6</td>
</tr>
<tr>
<td>Parker, Lester Shepard, b. 1860</td>
<td>Parker, Lester Shepard, 1860–1925</td>
<td>20060201</td>
<td>6</td>
</tr>
<tr>
<td>Silsoe, Malcolm Trustram Eve, Baron, 1894–</td>
<td>Silsoe, Malcolm Trustram Eve, Baron, 1894–1976</td>
<td>20060201</td>
<td>6</td>
</tr>
<tr>
<td>Adler, Friedrich, b. 1878</td>
<td>Adler, Friedrich, 1878–1942</td>
<td>20060208</td>
<td>5</td>
</tr>
<tr>
<td>Bailey, Derek, 1930–</td>
<td>Bailey, Derek, 1930–2005</td>
<td>20060208</td>
<td>5</td>
</tr>
<tr>
<td>Mager, Gus, 1878–</td>
<td>Mager, Gus, 1878–1956</td>
<td>20070207</td>
<td>5</td>
</tr>
<tr>
<td>Moore, Garry, 1915–</td>
<td>Moore, Garry, 1915–1993</td>
<td>20060816</td>
<td>5</td>
</tr>
<tr>
<td>Kagoshima, Juozo, 1898–</td>
<td>Kagoshima, Juozo, 1898–1982</td>
<td>20070207</td>
<td>4</td>
</tr>
<tr>
<td>Radcliffe, Ted, 1902–</td>
<td>Radcliffe, Ted, 1902–2005</td>
<td>20060816</td>
<td>4</td>
</tr>
<tr>
<td>Sakazaki, Shizuka, 1887–</td>
<td>Sakazaki, Shizuka, 1887–1978</td>
<td>20060816</td>
<td>3</td>
</tr>
<tr>
<td>Gos, Francois, b. 1880</td>
<td>Gos, Francois, 1880–1975</td>
<td>20060201</td>
<td>2</td>
</tr>
<tr>
<td>Carpenter, Clinton A. 1921–</td>
<td>Carpenter, Clinton A. 1921–2005</td>
<td>20060208</td>
<td>1</td>
</tr>
<tr>
<td>Carrasquilla L., Juan de Dio, b. 1833</td>
<td>Carrasquilla L., Juan de Dio, 1833–1908</td>
<td>20060208</td>
<td>1</td>
</tr>
<tr>
<td>Daniels, John S., 1906–</td>
<td>Daniels, John S., 1906–1996</td>
<td>20070207</td>
<td>1</td>
</tr>
<tr>
<td>James, Cecil, 1913–</td>
<td>James, Cecil, 1913–1999</td>
<td>20070207</td>
<td>1</td>
</tr>
<tr>
<td>Mock, Richard, 1944–</td>
<td>Mock, Richard, 1944–2006</td>
<td>20060816</td>
<td>1</td>
</tr>
<tr>
<td>Roger, Roger, 1911–</td>
<td>Roger, Roger, 1911–1995</td>
<td>20060816</td>
<td>1</td>
</tr>
<tr>
<td>Sampson, Alistair, 1929–</td>
<td>Sampson, Alistair, 1929–2006</td>
<td>20060816</td>
<td>1</td>
</tr>
<tr>
<td>Anderson, Sigurd, 1904–</td>
<td>Anderson, Sigurd, 1904–1990</td>
<td>20060208</td>
<td>0</td>
</tr>
<tr>
<td>Blitch, Iris, 1912</td>
<td>Blitch, Iris, 1912–1993</td>
<td>20060201</td>
<td>0</td>
</tr>
<tr>
<td>Chabloz, Fritz, b. 1841</td>
<td>Chabloz, Fritz, 1841–1905</td>
<td>20060208</td>
<td>0</td>
</tr>
<tr>
<td>Devine, Bing, 1916–</td>
<td>Devine, Bing, 1916–2007</td>
<td>20070207</td>
<td>0</td>
</tr>
<tr>
<td>Fujimoto, Yoshimichi, 1919–</td>
<td>Fujimoto, Yoshimichi, 1919–1992</td>
<td>20070207</td>
<td>0</td>
</tr>
<tr>
<td>Javierre Ortas, A. M. (Antonio Maria), 1921–</td>
<td>Javierre Ortas, A. M. (Antonio Maria), 1921–2007</td>
<td>20070207</td>
<td>0</td>
</tr>
<tr>
<td>Kavanaugh, Ken, 1916–</td>
<td>Kavanaugh, Ken, 1916–2007</td>
<td>20070207</td>
<td>0</td>
</tr>
<tr>
<td>Nelson, Paul, 1936–</td>
<td>Nelson, Paul, 1936–2006</td>
<td>20060816</td>
<td>0</td>
</tr>
<tr>
<td>Puzo, Mario, 1920–</td>
<td>Puzo, Mario, 1920–1999</td>
<td>20060816</td>
<td>0</td>
</tr>
<tr>
<td>Szabo, Sandor, 1915–</td>
<td>Szabo, Sandor, 1915–1997</td>
<td>20060816</td>
<td>0</td>
</tr>
<tr>
<td>Tichenor, Jerome, 1911–</td>
<td>Tichenor, Jerome, 1911–2006</td>
<td>20060816</td>
<td>0</td>
</tr>
<tr>
<td>Udam, Haljand, 1936–</td>
<td>Udam, Haljand, 1936–2005</td>
<td>20060816</td>
<td>0</td>
</tr>
<tr>
<td>Vidal, Pietro, b. 1867</td>
<td>Vidal, Pietro, 1867–1938</td>
<td>20060201</td>
<td>0</td>
</tr>
</tbody>
</table>
Supporting Name Authority Control in XML Metadata: A Practical Approach at the University of Tennessee

By Marielle Veve

While many different endeavors to support name authority control in Extensible Markup Language (XML) metadata have been explored, none have been accepted as a best practice. For this reason, libraries continue to experiment with the schema, tool, or process that best suits their local authority control needs in XML. This paper discusses current endeavors to support name authority control in XML for digitized collections and demonstrates an innovative manual solution developed and implemented by the University of Tennessee Libraries to achieve this goal. Even though this method for authority control in XML metadata still relies on manual efforts, it effectively reduces time and research work by efficiently setting priorities, identifying critical descriptive areas in the digital transcriptions, and identifying the most appropriate biographical resources to consult. The effectiveness of this approach in improving the rest of the metadata production workflow is evaluated and presented.

Soon after starting digitization projects, many libraries and other institutions often find that keeping track of name access points in the Extensible Markup Language (XML) is a huge challenge, regardless of the XML schema used. This is particularly the case in many types of digitized objects such as manuscripts, music, and other types of special collections where the number of personal names is exponentially more than the number of items digitized; the names are dispersed all over the digitized transcriptions; and information about these names is ambiguous, vague, and incomplete. However, no matter how difficult keeping track of name access points in digitized materials is, it is necessary in order to keep digitized objects retrievable. Access points not only help in the retrieval process of documents, but also help keep materials by the same creators or about the same subjects together.

To keep a successful track of name access points in XML documents, libraries have been experimenting with many different endeavors to find an effective way to achieve this goal. So far the efforts created to support name authority control in XML metadata consist of (1) using XML schemas to encode authority data; (2) endeavors for shared, cooperative, national, and international XML name databases; (3) manual and automated conversion tools from Machine-Readable Cataloging (MARC) to XML; and (4) automated generation of authority control through especially designed systems. The problem with most of these endeavors is that they only address the issue of how to encode name access points utilizing XML authority schema; they do not address the issue of how to extract or harvest these names directly from the XML records and transform them into useful access points. The few endeavors that have tried, such as the systems for automated generation of authority control, have only been successful in extracting names from XML records but not in turning them into
reliable access points. This is because their name matching processes fail. For this reason these endeavors will always depend on human intervention to work properly. In addition to not being completely reliable, many of these endeavors are costly and labor intensive, not to mention that most of them only display newly created access points locally.

The method introduced in this paper to support name authority control in XML metadata addresses the issue of extracting or harvesting names directly from XML documents manually and turning them into useful access points. In contrast to the previously mentioned methods, this method is effective, relatively simple, cost efficient, and has the ability to display the new access points at the national level by still using the richest authority file available—the Library of Congress's (LC) authority file (LCAF, http://authorities.loc.gov). This method consists of a simple manual approach to extract and create name access points that effectively reduces time and research efforts by efficiently setting priorities, identifying critical descriptive areas in the digital transcriptions, and identifying the most appropriate biographical resources to consult. When using this method, libraries will not have to go through the work of encoding authority data into XML schemas or translating authority data from one schema to another. Neither will they have to worry about hiring a programmer to build an XML name repository to store these records nor to create "shareable" XML metadata in order to make local authority records interoperable within national and international cooperative XML authority databases. Finally, this method is a practical alternative for those libraries and institutions that do not plan to build an automated tool to extract names directly from the XML records, which so far has not proven to be a reliable alternative.

The University of Tennessee Libraries' Name Authority Challenge

At the beginning of 2007, the University of Tennessee Libraries (UTL) transferred the creation of descriptive metadata for digitized manuscripts from the Digital Library Center (DLC) to the Technical Services Department. After archives were scanned and digitized in the DLC, digital surrogates of the manuscripts were created using the Text Encoding Initiative (TEI) schema. TEI is a markup language for representing structural and conceptual features of texts. It is used primarily for the encoding of documents in the humanities and social sciences and, in particular, in the representation of primary source materials for research and analysis. Files in TEI were sent to the cataloging department to be transformed into rich descriptive metadata using the Metadata Object Description Schema (MODS), UTL's selected schema for digitized manuscripts.1

As a requirement of using MODS, catalogers have to use controlled vocabularies to assign access points to the records. Soon after receiving their first batch of TEI encoded records, catalogers encountered serious difficulties in assigning personal names to the access points of MODS records. The following were the main problems:

- Difficulty in finding names in TEI records in the LCAF. This problem occurred because either the record did not exist in the LCAF or because names in the TEI records could not be matched with names in the LCAF because of the lack of sufficient biographical information in the TEI records to identify individuals. For example, proving that the individual in the TEI record was the same one listed in the LCAF was difficult because no data other than name were given.

- Inconsistency in the establishment of names not found in the LCAF. When names were not found in the LCAF, different catalogers assigned different headings for the same person, depending on the form of the name given in the manuscript. Entering the same individual's name in many different ways can create a serious problem for future discovery and access.

- Difficulty in differentiating individuals with similar names within the same collection. Many people whose names appeared in the manuscripts in the UTL collection shared the same or similar names with relatives mentioned in the collections. Distinguishing between two or more individuals with similar names became difficult because little, if any, biographical data were provided in the manuscripts. To make matters worse, individuals sometimes were called only by nicknames or had very commonly used names, which were difficult to differentiate from other similar headings. These factors created confusion for catalogers and made the process of differentiation almost impossible.

- Uncertainty about how to handle misspelled names and other typographical errors in the TEI transcriptions. Sometimes errors were made in transcribing names from the digitized image to the TEI files. Catalogers did not know whether to go back and fix the misspelling by editing the TEI record or to create an access point using the form found in the manuscript, even if it were a misspelled form. Different decisions made by various
catalogers brought more inconsistency to the access points.

The problems in assigning personal headings to the access points of MODS records demanded an effective method to handle name authority control in the UTL’ s digitized collections.

**Literature Review**

Libraries and other institutions seeking to support name authority control in XML metadata have tried various approaches. Some have proven more successful than others, but none has consistently been implemented for XML documents. Some commonly mentioned approaches in the library literature include Metadata Authority Description Schema (MADS), MARC Extensible Markup Language (MARCXML), Encoded Archival Context (EAC), OCLC Linked Authority File (OCLC LAF), and the Automated Name Authority Control (ANAC). Most of these authority initiatives for non–MARC metadata are designed to handle authority control at the local level; only a few try to do so at the national level. Some are XML schemas for authority elements created for use in conjunction with particular XML bibliographic schemas. Others are conversion tools that convert MARC into XML records. Some authority initiatives claim to be automated, but usually these are really semiautomated approaches that apply a mixture of manual and automated approaches to generate authority control.

**XML Schemas for Authority Elements in Non–MARC Metadata**

In the early 2000s, the LC created MARCXML, an XML schema that can be used for authority purposes and is based on and very similar to MARC 21. It was first presented by the Information Technology Section at the IFLA conference in Glasgow in 2002. In a recent report of that meeting, McCallum states that "a key characteristic of MARCXML is that it produces an exact equivalent of the MARC 21 record so that roundtrip conversion to and from it is lossless. This schema has been widely used and is the basis for the international standard for an XML version of the MARC structure that Danish standards have proposed." In summary, she concludes that MARC/XML "provides a basis for evolution while maintaining standardization."53

Later in 2005, the LC developed MADS, another schema for authority elements, but this one was created to be used in conjunction with MODS, a particular bibliographic schema. As in the case of MARCXML, MADS also has a strong relationship with the MARC 21 authority format. Guenther describes one advantage of MADS: "Because MADS is derived from the MARC 21 Authority format, which has been used for more than 30 years, its underlying model is well-established [and] a MODS description could link to a MADS description to eliminate redundant information."4 She also mentions disadvantages: "Since MADS has not yet been widely implemented, it could still be considered experimental, and wider experience using it may result in refinements to the schema."55

EAC (www.library.yale.edu/eac) is another schema for authority elements created to be used in conjunction with a bibliographic schema, Encoded Archival Description (EAD). EAC started as an original effort from a group of archivists who met in Toronto in March 2001 to create a model for name authority control in archival materials. The initiative, still in the beta phase, is currently managed by an international group of archivists and Yale University. Thurman explains that EAC allows "archivists to encode information [in XML] about the creators and context of creation of archival materials, and to make that information available to users as an independent resource separate from individual finding aids."6 He notes that EAC "development is not yet complete, and it has so far been implemented only experimentally."7 In the effort to create an XML encoding standard for archival authority control, Pitti concludes that "there are many difficult intellectual, technical, cultural, linguistic, and political challenges to be addressed in order for the effort to be successful. While all of the challenges are significant, the political challenges stand out as particularly difficult."9

The MARC Extensible Markup Language Document Type Definition (MARCXML DTD), which is not the same as the MARCXML schema, is an older schema format for XML created by the LC. It started in the mid-1990s as an SGML DTD that supported the conversion of data from MARC Authority to SGML (and back) without loss of data. In the early 2000s, as technology developed and changed, the MARC SGML DTD became converted to MARCXML DTD. McCallum states in her report that this method "yielded very large DTDs since [XML DTD] is naturally verbose, and the tagging approach mandated a DTD element specification for every MARC subfield or coded character position."9 An entry in Wikipedia summarizes the problems with DTD, noting that it is limited because “it has no support for newer features of XML, most importantly namespaces; uses a custom non–XML syntax, inherited from SGML, to describe the schema; and lacks expressiveness [because] certain formal aspects of an XML document cannot be captured in a DTD.”10 Nonetheless, even through its limitations, MARCXML DTD is still used and is kept available in the MARC 21 website. The reason some keep using it is that “several users have stated that they find it appropriate for certain applications, especially those needing extensive validation of records.”11 Libraries that decide to use any of
these schemas to encode their XML authority data first will have to decide which names they want to extract from the XML records to be used as access points. This process has to be done manually because XML authority schemas do not extract information directly from XML records, but only encode it. The chosen names are then turned into access points, for which research is required. Next, the information is encoded into the desired XML authority schema. After this is done, a local XML name repository will need to be built and sustained to store and retrieve these authority records in XML. The problem with relying on XML name repositories for authority control is that catalogers frequently do not have the technological background to build or maintain the repositories. For these tasks catalogers often will have to rely on the library’s programmers, who have competing responsibilities such as technical support or database and catalog maintenance. Hiring a programmer to work exclusively with the technical services department may be an option, but can be very expensive. For these reasons, the use of schemas for metadata authority control may not be the best solution for some libraries.

Conversion Tools from MARC to XML Authority Schemas

Another option for metadata authority control involves taking names that appear in the LCAF (in MARC format) and converting them into XML schemas using conversion tools between MARC and XML. Some of these tools involve automation, while others do not. An example of an automated conversion tool between MARC and XML is the MARC Tool Kit. This tool provides converters for transforming data from MARC 21 to MARC-XML and back, including character set conversion to and from Unicode. These converters can be downloaded from the MARC website and used by others in their own systems where they can also shape them to their own data and needs.

[This] conversion software was developed by Bas Peters in the Netherlands and made available by him as open source software. It is in part adapted from an extensive set of programs for manipulating MARC 21 data.12

The LC sees these transformations provided from the MARC 21 maintenance agency as “being valuable to the community to help maintain the savings and interoperability built up through use of a common format.”13

Maps and crosswalks between MARC and XML are other types of conversion tools used to translate authority data from one schema to another. These tools use manual approaches and, for this reason, require more effort on the part of the cataloger. The number of this type of conversion tool parallels the number of XML schemas. Some include conversions from MARC to Dublin Core and Dublin Core to MARC, others from MARC to MODS and vice versa. Almost all XML schemas have a crosswalk to convert their schemas into MARC metadata or from MARC to XML. An assessment published in Online Libraries and Microcomputers reveals some of the common challenges faced when using crosswalks and maps.14

This analysis reports that

there is often not a one for one mapping between fields in different metadata schemes. This means that many fields may need to be mapped into fewer fields (or vice versa). There can be a loss of granularity in metadata descriptions that may result in poorer searching. Many specific metadata schemes are targeted to a specific subject or type of material. When converting to another scheme there may be a loss of specificity and granularity. In metadata mapping one may want to parse through free text data to extract relevant data to extract for a more detailed scheme. This is difficult, time consuming and fraught with error because of variations in actual content.

. . . How does one handle subfields and indicators (e.g., MARC) when mapping to systems that do not support the same detail? How should subfields from more complex metadata schemes be delimited in less complex metadata schemes? How does one map and handle local control numbers? Without the transferring of local control numbers there may be later problems in a shared database for updates, deletes and overlapping records.15

The idea of converting MARC authority records into records that use the local XML schema sounds appealing, but this method creates double work for the library. Converting authority records from MARC to another metadata schema requires translation of records plus the construction of an XML name repository to support the records. Many of the manuscript names do not exist in the LCAF, so locally established headings will have to be created for these names following the construction format of headings in the LCAF. Following the same construction format keeps consistency between locally created headings and those exported from the LCAF so that the headings look the same and index the same way.

If many headings have to be locally established in XML schema following the rigorous LCAF standards,
then libraries may find establishing the headings directly in the LCAF more worthwhile because other libraries can benefit from this authority work. This approach can also save the time necessary to convert names to another schema and to build a database to manage them. For these reasons, relying on conversion of authority records from MARC to XML may not always be the best approach to support name authority control in XML metadata.

Endeavors for Cooperative Searchable XML Name Databases

Since the early 2000s, libraries and other institutions have attempted to create a national searchable XML name repository. Shared XML name repositories try to harvest name authority data from different sources distributed throughout the country and make it interoperable between different institutions. One example of such an attempt is the OCLC Linked Authority File Project (http://alcm.s.org/laf), an endeavor between the Open Archives Initiative and the OCLC. The Linked Authority File (LAF) was developed in 2002, hosts a shared server containing LC authority records and potentially authority records supplied by others, and is intended to provide Web-based access to interactive and automated authority records. This national name repository periodically uploads names from the LCAF and presents them in both MARC and MARCXML formats.

Even though the LAF’s original intention was to harvest names from different sources besides the LCAF, this has not yet been fully implemented. When asked if the LAF plans to harvest authority data from other sources besides LCAF, an OCLC Research representative replied that “no further development of the system itself is planned.” No explanation was given on why the LAF only harvests authority data from the LCAF and not from other sources, but this may be due to the difficulty of making authority metadata interoperable between different institutions, a common problem faced by cooperative, inter-institutional databases. Given the lack of promotion, this initiative is fairly unknown and, consequently, has not been widely implemented.

The Linking and Exploring Authority Files project (http://xml.coverpages.org/leaf.html) was an attempt to create a cooperative searchable XML database for authority names for the European community. It was created with the purpose of being accessed by anyone, regardless of affiliation, who might be interested in name authority files from European manuscripts. This three-year project (2001–4) was cofunded by the Information Society Technologies Program of the Fifth Framework of the European Commission.

Linking and Exploring Authority Files (LEAF) sought to develop a system model that uploaded name authorities—distributed through local servers of participating European organizations—to the central LEAF system. Authorities then were converted and stored into EAC schema, with authorities that belonged to the same entity being automatically linked. To have a network where those linked records could be applied, LEAF was integrated into a search engine called Manuscripts and Letters via Integrated Networks in Europe (MALVINE). MALVINE “is a search engine that harvests databases which provide information about letters written by famous persons that are kept in different European institutions.”

After being integrated into the MALVINE search engine, the linking process of LEAF proved not to be reliable. Kaiser and colleagues stated that it is inevitable that in some instances the linking process will produce incorrect results. Records describing two different persons might be automatically linked because they do not contain enough discriminating information. On the other hand, two records representing the same person might not be linked because they do not share an identical name form. Recollecting the main purpose of library authority records—the disambiguation of persons described—it may be argued that those records leading to wrong links are not sufficiently rich in content to serve their original purpose.

The project ran as a funded test for thirty-six months, ending in May 2004. Thereafter it was left as an integrated part of the MALVINE search engine, where it is still used because it is seen as “highly relevant [content] to the cultural heritage of Europe.”

In theory, using a shareable XML name database sounds like a great plan for libraries and institutions that already have built a local XML name repository because records can be uploaded by one entity and shared between different institutions. In reality, experience has shown that this approach does not work because for metadata to be successfully harvested by a national cooperative repository, all locally created authority metadata needs to be “shareable.” Shareable metadata are metadata that need to follow a set of standards to be interoperable between different institutions. The standards needed to create shareable metadata have not yet been established because of the lack of cooperation between different institutions. Pitti states, “As economically and professionally desirable as cooperative, shared authority control, and biographical, historical description is, successful realization will require standards and systems that are collaboratively developed, administered, and maintained. These
Automated Endeavors to Support Name Authority Control in XML

Other projects have sought to solve the problem of addressing authority control in XML records on a local scale by implementing automated processes to extract and detect possible name access points in XML records. For example, in 2003, the Digital Knowledge Center (DKC) at John Hopkins University explored the application of automating metadata generation for name authority control. To achieve this purpose, the DKC created a tool called the Automated Name Authority Control (ANAC). This automated metadata generator applies an established algorithm to identify LC–authorized names for each name in descriptive metadata records. Patton and colleagues stated that “The main reason for undertaking ANAC was to develop a tool that would reduce the costs associated with introducing name authority control to metadata [because] relying exclusively on human catalogers would be substantially more expensive and time consuming.” After evaluating the tool, the authors determined that the automated system was not sufficiently reliable in many cases. They added, “Even though ANAC could be a valuable complement [to authority control], it was never anticipated that it would entirely replace the human effort.” The authors concluded that the most effective and cost-efficient workflow would couple ANAC with human oversight.

Another attempt for automated generation of name authority control in digitized collections was suggested by French, Powell, and Schulman. They introduced the concept of approximate word matching similar to the approximate string matching techniques traditionally used in detecting variant names in databases. This approach detects variable forms of strings in names through clustering algorithms and then groups the strings together under a standard form. The authors observed that, even though this automated clustering approach can reduce human effort by half, a certain amount of human effort will always be required to verify the output, thus this approach is semiautomatic.

Although systems created for the “automated” generation of name authority control claim to be automated, they are not completely so. They are really semi-automated approaches because they will always rely on human intervention for the process to work properly. Because of the need for human intervention and the high cost of creating such an endeavor, systems for the so-called automated generation of name authority control may not always be the best approach to support XML name authority control in many libraries.

Designing a Practical Approach at UTL

After reviewing the library literature and analyzing the advantages and disadvantages of the different approaches to support name authority control in XML metadata, UTL decided that none of these approaches was appropriate for the local situation. Because of time and funding constraints and lack of technological support, UTL decided to design a different approach that would be customized for UTL. Several points needed consideration. First, the new authority control method had to be completely sustainable by the UTL catalogers. Sustainable in this context meant the method had to be cost effective and use a level of technology with which the catalogers were comfortable. Because the cataloging department could not hire a local programmer, they ruled out building a local XML name repository and decided to capitalize on existing staff knowledge instead. Second, the authority control had to be achieved within a reasonable amount of time. Because this work is time consuming, priorities for which names to establish and which ones to leave out had to be set from the beginning. These priorities will be referred to from now on as “establishment criteria.” The reason for using establishment criteria is that searching, verifying, and establishing each name found in the TEI records would be impossible because they number in the thousands. The criteria would determine the cases in which names would be searched and established in the LCAF. Third, the new authority method had to support a level of quality if UTL wanted to keep materials by the same creators together. Given these considerations, UTL decided to use a manual approach that keeps taking advantage of the largest name authority file available—the LCAF.

The method UTL implemented integrated all the previous points. The process is described in detail in the following section. Briefly, authority control is performed as soon as the DLC sends the TEI transcriptions to the cataloging department. Authority control, then, is performed before records are cataloged, and only by one person to avoid future inconsistencies in establishing names. The person chosen to perform this task is one of the catalogers, who had previous experience creating authority records through the Name Authority Cooperative Project (NACO). After authority work is finished, headings are stored in a Microsoft Excel shareable spreadsheet. As soon as the spreadsheet is ready, catalogers are notified and TEI records are sent to them. The catalogers then have the necessary resources to catalog and create rich, descriptive MODS records with the least amount of effort.
The detailed process followed at UTL consists of the following steps: The librarian charged with authority control receives a batch of one hundred to three hundred TEI transcriptions with their digital images from the DLC. The files usually originate from many different collections in the university archives. The authority control librarian devotes approximately two weeks (full time) to authority work for this batch of records. After receiving the TEI files, the authority librarian opens and browses the files using XML the TEI files, the authority librarian this batch of records. After receiving weeks (full time) to authority work for this batch of records. After receiving weeks (full time) to authority work for this batch of records. After receiving weeks (full time) to authority work for this batch of records.

The authority control receives a batch of one hundred to three hundred TEI transcriptions to search for names:

sections to browse in the TEI file to get names and how to properly keep record of them in a list

The detailed process followed at UTL consists of the following steps: The librarian charged with authority control receives a batch of one hundred to three hundred TEI transcriptions with their digital images from the DLC. The files usually originate from many different collections in the university archives. The authority control librarian devotes approximately two weeks (full time) to authority work for this batch of records. After receiving the TEI files, the authority librarian opens and browses the files using XML the TEI files, the authority librarian this batch of records. After receiving weeks (full time) to authority work for this batch of records. After receiving weeks (full time) to authority work for this batch of records. After receiving weeks (full time) to authority work for this batch of records. After receiving weeks (full time) to authority work for this batch of records.

The authority control receives a batch of one hundred to three hundred TEI transcriptions to search for names:

- **Title section**, to retrieve the names of senders and recipients.
- **Body section**, to find names that “pop out” as important access points. If summary sections are provided, the librarian should browse through them as well.
- **Signature section**, to determine the preferred form of heading for the sender.

Names found in these sections become crucial because they will form access points for the MODS record, the equivalent of the fields 1xx (main entry fields), 6xx (subject access), and 7xx (added and linking entry fields) in MARC records. The authority librarian should pay attention to variant forms of headings as well.

The authority librarian makes a list of the names found, as well as of their variant forms, along with the record number of the TEI where the names were found. By recording this number, the librarian can later retrieve the exact location of these names in case more information is needed. In going through the rest of the TEI files, the authority librarian may encounter the same names, as well as other new names or variant forms of names, and keeps a list. This process is illustrated in figure 2.

After browsing the TEI files in one collection, the authority librarian looks at the names gathered so far and...
compares them to see if some names have been mentioned more than once using the same form or a variant one. She also counts the number of times each name is mentioned in different TEI files.

The establishment criteria are then applied to names in that particular collection. These criteria help determine which headings will be searched, verified, or established, and which ones will not. UTL developed establishment criteria that worked well in most situations. Names mentioned in at least three separate TEI files are searched in the LCAF and established if not found. The same process applies to names mentioned in the “Title” section of the TEI files as senders or recipients and to names that have a collection with their name (this can be checked in the “Collection” section of the TEI file). The one exception to the establishment criteria is the handling of names for prominent historical individuals. Because they are also searched. Figure 3 illustrates application of the establishment criteria.

For names that will be established according to the criteria, the authority librarian returns to the TEI files in which they were found. This is done by using the TEI record number that was noted on the list. When retrieving the TEI records, the librarian browses the text around the area where the names were found to get as much information—stated directly or indirectly—about the person as possible. Examples of useful areas to browse in TEI files are the “Title” section, which gives the date and place a letter was sent, and the “Body” section, which may provide information on people’s roles, relationships, and so on. The authority librarian annotates this information, along with the variant forms of the name found. The result might look like this:

Jacob Breck, Jab Breck, Jacob B.; sender of letter from Franklin, TN in 1864 to Philadelphia, Penn; judge; wife Lizzie.

These brief factual data will provide a general idea of who this individual was and when he or she lived. The data found can be expanded later through further research in outside sources.

After all biographical facts available in TEI files have been annotated, the authority librarian then consults various research tools such as finding aids. The University Archives, which own the original texts for the TEI files, have created finding aids, many of which are online, Tennessee state and county archives also may contain related finding aids. These tools may provide information on the person’s time period, family, place of residence, and more details. This information will be used by the authority librarian to place this person in context, see with whom he or she associated, and differentiate the individual from others with similar names when searching the LCAF.

If finding aids do not provide enough information about a person or are not available, the librarian searches other outside sources such as Google Book Search. This tool provides the ability to search sections within long texts of reliable resources that are freely accessible online. Other useful and freely accessible websites for historical biographical research include the Political Graveyard—A Database of Historic Cemeteries (http://politicalgraveyard.com), the state finding aids via the state library or state historical society websites, Genealogybuff.com, the Biographical Directory of the United States Congress (http://bioguide.congress.gov/biosearch/biosearch.asp), and the Civil War Rosters website (www.geocities.com/Area51/Lain/3680/cw/cw.html). The latter can be searched by soldier, regiment, and more. In addition, the authority librarian searches the Tennessee Genealogy and History Web (TnGenWeb.)
org)—other states may have similar sources. Other fee-for-service genealogy databases are also available. When searching historical personal names online, a useful tip for best results is to search using very specific factual data. For example, if the only information available from a TEI transcription about “Lord Cornwallis” indicates that he was alive during 1791 and wrote from Blount County, these facts should be integrated into the search.

As relationships between individuals start becoming clearer, the authority librarian should illustrate the relationships using visual aids in addition to noting the information. Visual aids such as genealogical trees, arrows, and diagrams can prove useful to represent relationships between individuals. Visual aids are important for the authority librarian because she may need to consult these aids to create names, and they are important for the rest of the metadata team, who will assign the names as access points in the MODS records. Understanding relationships among the individuals in the digitized transcriptions is crucial to create useful access points for the records.

After gathering enough data about a particular individual, the librarian searches the name in the LCAF. At this point, she will have enough biographical information to distinguish that individual from others with similar names in the LCAF. If the heading is found in the LCAF, the authority librarian copies and pastes the heading into the local Excel spreadsheet, along with its cross-references and notes. The spreadsheet serves two purposes: It builds a local database of the established names found in UTL’s digitized archives and provides catalogers with a narrower list of established names that appear in the TEI files they will catalog, saving them the time and effort of searching the LCAF. Sometimes, the authority librarian has located extra information not already mentioned in the LCAF about the individual listed. This extra information, such as biographical details or other variant forms of names, can be added optionally to the LCAF record in order to enhance it. This additional information can help differentiate this person from others with similar names in the future.

If the heading is not found in the LCAF, the librarian searches the OCLC Connexion Bibliographic File for records that used this name in any of their access points. To search in these areas of the bibliographic files, the authority librarian performs keyword searches using the index labels “au” (author) and “su” (subject). Searching in the OCLC Bibliographic File is a step required before establishing any heading in the LCAF. This search often leads to records that mention variant forms of this person’s name as well as extra facts not discovered previously.

After searching the OCLC Bibliographic File, the authority librarian establishes headings that were not found in the LCAF using the biographical information gathered to this point. Headings can be established locally or nationally, depending on the institution’s involvement with NACO. Libraries that are NACO members or part of a NACO Funnel Project have the option of making name contributions nationally. A NACO funnel project is a group of libraries who together are authorized to contribute name authority records to the LCAF. On the other hand, libraries that are not NACO members will not have the option of making national name contributions and will have to establish them locally. UTL has the option of making name contributions to the LCAF because it is a member of the Tennessee NACO Funnel Project. When establishing a heading, the authority librarian includes all the cross-references and factual data found previously in the research that may prove useful for the future. After establishing a heading in the LCAF, the authority librarian copies and pastes the heading into the local Excel spreadsheet with all the other LCAF names already found in OCLC.

After the chosen headings from one particular collection have been searched and established, the authority librarian browses the TEI files of the next collection, repeating the steps described above until all collections in the batch of TEI files are completed.

After names that met the establishment criteria have been searched or established, the lists of names that did not meet the establishment criteria remain. These lists are kept by the authority librarian in case any of the names need to be established in the future. Each list contains the TEI record numbers indicating where those names were found and can help retrieve the records if they are needed later.

After the authority librarian completes these steps, the authority work is considered completed. The tools and resources needed for cataloging metadata are then placed in a shareable department server. These include the digitized files in JPEG, transcription files in TEI, visual tools, and the Excel spreadsheet with the authorized name headings. The catalogers are then prepared to start creating MODS descriptive metadata with the least amount of inconvenience.

Implementing this authority control process before the rest of the metadata production starts solved the problems UTL initially faced when trying to assign name access points to MODS records without authority control. This approach to authority control solved both the difficulty in finding TEI names in the LCAF and the inconsistency in establishing names if they were not found there. Now that the authority librarian provides all the necessary authority work, the catalogers will not have to worry about searching these names in the LCAF or establishing them. The catalogers will
find the established forms plus their variants in a local, shared Excel list.

Placing authority control before the rest of the metadata process permitted the catalogers to focus on the rest of the description. It solved the difficulty of differentiating individuals with very similar names within a collection by providing useful biographical information. The use of qualifiers and other attributes in authority control also helped in this purpose. The provision of visual aids such as genealogical tables helped catalogers throughout the process of visualizing family relationships and helped to diminish confusion about similar names.

The problem of misspelled names and other typographical errors that occurred when transcribing names from the original text to the TEI files was also solved with this authority method. By receiving the TEI files with their digitized images as a first step, the authority librarian had the opportunity to catch transcription mistakes and fix them before the catalogers had the chance to discover them.

Assessing the Effectiveness of UTL’s Approach

To assess the effectiveness of this approach, UTL decided to compare the metadata workflow before having authority control with the workflow after implementing authority control. UTL performed an informal assessment through a questionnaire, asking the six catalogers who experienced the first workflow without authority control to compare particular production aspects within both workflows. The questionnaire was distributed three months after the implementation of authority control into the metadata workflow and consisted of ten closed questions and one open question to provide suggestions. The questionnaire is presented in the appendix to this paper.

In the questionnaire, the six catalogers were asked if the speed of producing MODS records improved after the implementation of pre-cataloging authority control. All six agreed that the speed of producing MODS records was higher after the implementation of authority control. When asked to estimate the number of MODS records produced per week before the implementation and the number produced per week after the implementation, they reported a higher number of MODS records produced per week after the implementation. The six catalogers responded that before the implementation, an average of five or less records were produced per week; after the implementation, five catalogers reported an average of ten or more MODS records produced per week and one cataloger reported six to nine records per week.

Catalogers were asked if the provision of authority control, before they began metadata work, freed them to concentrate on other important descriptive metadata tasks such as assigning subject headings, writing summaries, and analyzing the TEI record. To this question five of the six catalogers responded yes, the provision of authority control freed them to perform other important metadata tasks; one cataloger answered that it made no difference. Concerning quality of MODS records produced, all six agreed that the quality of MODS records improved after the implementation of authority control. Reasons for the quality improvement of MODS records were that more controlled access points were available than before the process changed, and that they were more consistent. Five of the six catalogers agreed that MODS records were more difficult to create before having the new-approach authority control. Reasons given to explain this difficulty before having the new approach were that there were inconsistencies in names established, distinguishing different persons with similar names was more difficult, and no visual tools were available to clarify relationships between individuals. Of the six, only one cataloger reported that the difficulty of creating MODS records was the same before and after the implementation of authority control.

Future Plans

While UTL’s informal assessment demonstrated the effectiveness of this authority method in improving the MODS metadata production workflow, it also showed aspects that need improvement and issues that will need to be addressed in the future. In the suggestions at the end of the assessment, two catalogers showed concern about what will happen to the metadata workflow if the authority librarian leaves. To solve this, UTL will eventually need to expand and delegate authority control tasks to other members in the cataloging department so that authority control does not depend on one person’s contribution. Initially, some authority control responsibilities, such as research tasks, can be delegated to members within the cataloging department. Eventually this responsibility can expand, with the catalogers creating personal authority records. They will need training either from the local authority cataloger who has NACO experience or through the closest NACO Funnel Project. Both alternatives would require initial time investment by the staff and institution, but this option could help make the workflow run more smoothly and to cover for the person performing authority work in case he or she leaves.

Another issue identified through the questionnaire was the increasing difficulty of searching names with many cross-references in Excel. As the number of names with cross-references increases, so does the difficulty in handling them effectively by the Excel software. Excel was not designed to
handle information arranged in thesauri format but primarily to handle numerical data. For this reason, commercial software that is better suited to handle cross-references will be needed in order to substitute for Excel. Thesauri software, which is software designed to build and edit thesauri headings, can manage cross-references very well and is cheaper than hiring a programmer to build an XML repository. Thesauri software is available in standalone packages and as database modules, which are integral parts of larger systems and need to run with them. Examples of popular standalone packages are MultiTes, Data Harmony, a.k.a. Classification Software, STRIDE, and Term Tree. Examples of database modules are STAR and TheMa Thesaurus Manager for Oracle. Using thesauri software is an economical and attractive option to store and manage local authority names and one that UTL will begin to explore.

**Conclusion**

As evidenced throughout this paper, many libraries and institutions are looking for ways to turn necessary tasks over to machines, but experience suggests this is not yet possible for name authority control in XML metadata. The efforts created so far to achieve this goal, besides being costly and work intensive, have proved to be ineffective and unreliable. Most do not address the issue of how to extract or harvest names directly from the XML records and transform them into useful access points, but focus on how to encode the access points into XML authority schema. The few endeavors that have tried to harvest names directly from XML records have proved not to be completely reliable in their processes of matching and linking names, making them dependent on human effort.

In addition to not addressing how to select and extract access points from the XML records, most of these endeavors require labor-intensive encoding of authority data into XML schemas and, subsequently, the creation of a local XML name repository to store and manage these records. Building an XML name repository is a task that requires a high level of technological background most catalogers lack. For this reason a programmer will have to be hired to build a name repository, and this is an expensive approach not many libraries can pursue. Furthermore, creating authority data to be stored in a local repository will only benefit the local institution, causing inconsistencies and duplication of efforts between different institutions that try to set up access points for the same individuals. Initiatives that tried to avoid the duplication of efforts in name authority control—by creating a national XML name repository to share authority data and make it interoperable between different institutions—have not been successful because the XML authority data needs to be shareable to be interoperable between the national repository and the other institutions. To date, this has not been successfully achieved.

In contrast to these approaches, UTL’s method to support name authority control in XML metadata is effective, reliable, and cost effective. It addresses the issue of extracting names directly from the XML documents and turning them into useful access points that can be shared nationally through the LCAF, thus avoiding duplication of efforts and benefiting all libraries who may share the same access points.

UTL’s approach is simple and can be used by other libraries and institutions that face similar issues when trying to support name authority control in their XML metadata. Common problems such as inconsistency in the establishment of names, difficulty in differentiating individuals, and deciding which names to turn into access points can be solved by implementing this method before creating any descriptive metadata for digitized transcriptions. Regardless of the local XML schema used, this approach can be applied in the same way to different collections.

**References**

3. Ibid., 6.
5. Ibid.
7. Ibid., 199.
12. Ibid., 5.
13. Ibid.
15. Ibid.
Appendix. Comparison of Metadata Workflow Before and After Implementation of Authority Control

1) Do you think the speed of producing MODS records was higher?
   a) Before the provision of authority control
   b) After the provision of authority control
   c) It was the same before and after

2) Do you think the quality of MODS records produced was better?
   a) Before the provision of authority control
   b) After the provision of authority control
   c) It was the same before and after

3) If you answered "after" to the previous question, why do you think the quality of MODS records was better after the implementation of authority control? Choose all that apply:
   a) Because there were more controlled access points
   b) Because access points were consistent between records
   c) Because records were more accessible to users
   d) None of the above

4) Do you think the production of MODS records was more difficult?
   a) Before the provision of authority control
   b) After the provision of authority control
   c) It was the same before and after

5) If you answered “before” to the previous question, why do you think it was more difficult to produce MODS records before the provision of authority control? Choose all that apply:
   a) Because there were inconsistencies in names established
   b) Because it was harder to distinguish different persons with similar names
   c) Because there were no visual tools available to understand relationships between persons
   d) None of the above

6) Do you think the provision of authority control for names in metadata records frees you to concentrate in other important tasks such as assigning subject headings, writing an abstract, or analyzing the TEI records?
   a) Yes
   b) No
   c) It makes no difference

7) Do you think the provision of authority control for names before MODS are produced improves the metadata workflow in general?
   a) Yes
   b) No
   c) It makes no difference

8) On average how many MODS records did you create per week before the implementation of authority control into the metadata workflow?
   a) More than 10
   b) 6–9
   c) 5 or less

9) On average how many MODS records did you create per week after the implementation of authority control into the metadata workflow?
   a) More than 10
   b) 6–9
   c) 5 or less

10) In which aspects of authority control would you like to see more improvement? Choose all that apply:
    a) Searching names in Excel spreadsheet
    b) Illustration of visual aids
    c) Time for authority control to be ready
    d) Other, please explain: __________
    e) None

11) Do you have any additional comments or insights regarding authority work for the metadata workflow? (For instance, recommendations for workflow, tools improvement, adjustments, and so on?)

Thank you for taking the time to answer the questionnaire!
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...
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 batch-loading (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 batch-loading purposes. 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

---

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.


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. 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](http://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.
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://oregonstate.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 prepended 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. 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. 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 batchloading 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

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.
Book Review
Edward Swanson


Written from a cataloging practitioner’s point of view and set against the backdrop of rapid changes in journals publishing, this book examines the changes in cataloging theory and practice that have ensued from the rapid proliferation of electronic journals and aggregator databases. Heinrich states that the book is intended principally for cataloging managers who have the responsibility for developing the cataloging policies and procedures that will define the resources discovery routes that govern how a library’s users will gain access to electronic journals. It is also aimed at cataloging practitioners.

The opening chapter deals with the effect of the Internet on the work of catalogers, starting with the MARC format familiar to most catalogers and charting the development of the basic concepts within MARC that were originally established to describe physical printed works. From this familiar territory, Heinrich moves on to describe emerging metadata schema, including MARCXML, Metadata Object Description Schema (MODS), and Dublin Core. She helps those new to such concepts to improve their understanding through the use of illustrations, numerous tables, and examples that outline both the data entry elements incorporated in each schema and the public views of different record types taken from online library catalogs and the Internet.

From this basic introduction to the problems that libraries have been facing and the development of new metadata schema, the author goes on to describe the changes made in the Anglo-American Cataloguing Rules, 2nd ed., Machine-Readable Cataloging (MARC), and Cooperative Online Serials (CONSER) rules, focusing on the revisions required in each standard to provide bibliographic control for remote electronic resources. Having completed this useful background information, Heinrich starts to address some of the real issues and decisions facing librarians, including the central problems of whether to adopt a user-friendly single cataloging record approach, which keeps all data on both the print and electronic versions of a journal within one catalog record, or to go for the administratively easier separate-records approach. The role of aggregators or electronic journal package providers is discussed at length, and the effect of the ever-changing journals market, with the consequent virtual impossibility of libraries being able to keep up with in-house catalog record creation and maintenance overheads, will be familiar to many journals cataloging practitioners.

As libraries have struggled to keep up with cataloging and record maintenance tasks, a new broad market and increasing demand for commercial MARC services has developed. In examining a variety of vendors’ MARC products, the author discusses many of the dilemmas encountered by libraries when considering acquisitions of such services. Heinrich describes and analyzes the complexities of incorporating commercially produced catalog records into local library databases and offers some practical solutions to many of the most common questions and issues that libraries would face.

In chapter 4 the author brings together all the theoretical and historical strands of the emergence of the Internet, new cataloging rules, and issues, and tries to put them in a practical perspective by describing how all of these issues have been locally addressed at her library at California State University, Northridge. The university library’s step-by-step implementation of a commercial MARC record service is described in an attempt to “help libraries avoid feeling blindfolded during the course of implementation and post implementation maintenance” (127). The information in this section is also backed up with quotations and references from other libraries across North America and Europe, giving readers easy access to journal articles that have addressed many of the key themes of single versus separate online public access catalog records, e-resource cataloging practice, and the effects of e-journal management tools and services on serials cataloging. The book concludes with a look into the future of cataloging generally, citing the “noticeable shift from ‘deep’ quality cataloguing to ‘light’ cataloging” (195) and the move to the supersizing of cataloging as libraries have graduated from single record downloading to bulk ingesting of files with hundreds or thousands of records. New developments such as metasearching or federated searching, the emergence of the Open URL standard, the use of Digital Object Identifier (DOI)-based linking, and open access initiatives are briefly described, continuing the theme of providing basic introductions and descriptions to key themes and developments.

Although from the outset Heinrich states that the book is intended for cataloging practitioners and cataloging managers, she also acknowledges that
it may be useful to vendors and commercial suppliers that provide online journal services to libraries. To this list could also be added serials librarians, for whom responsibility for cataloging journals may be a new and daunting responsibility with a minefield of acronyms, issues, and standards that need to be safely crossed, or simply as background for understanding the difficulties faced by their cataloging colleagues in describing and facilitating access to the collections they manage and administer.

Although numerous journal articles have been published addressing many of the issues raised and discussed in the book, this work is unique in its attempt to chart the historical context of developments in this field of librarianship and put them in perspective for those facing the challenges of handling and managing electronic resources today. Heinrich’s book is not only very readable as a complete work but can also be used as a quick reference guide for those wanting to look up specific terms and acronyms or read a case study of an actual implementation of an e-journals cataloging service from an external vendor. The work provides a very useful comprehensive overview of all the issues and developments and as acts as a one-stop shop for those wishing to gain a better understanding of the complexities of current-day serials cataloging.—Helen Adey, (helen.adey@ntu.ac.uk), Nottingham Trent University, Nottingham, England.
Statement of Ownership, Management, and Circulation

Library Resources & Technical Services, Publication No. 311-960, is published quarterly by the Association for Library Collections & Technical Services, American Library Association, 50 E. Huron St., Chicago (Cook), Illinois 60611-2795. The editor is Peggy Johnson, University of Minnesota, 499 Wilson Library, 309 19th Ave. South, Minneapolis, MN 55455. Annual subscription price, $85.00. Printed in U.S.A. with periodicals-class postage paid at Chicago, Illinois, and at additional mailing offices. As a nonprofit organization authorized to mail at special rates (DMM Section 424.12 only), the purpose, function, and nonprofit status of this organization and the exempt status for federal income tax purposes have not changed during the preceding twelve months.

(Average figures denote the average number of copies printed each issue during the preceding twelve months; actual figures denote actual number of copies of single issue published neared to filing date: June 2008 issue.) Total number of copies printed: average, 6,074; actual, 5,910. Paid distribution outside the mails including sales through dealers and carriers, street vendors, counter sales, and other paid distribution outside the USPS: average, 461; actual, 316. Total paid distribution: average, 5,535; actual, 5,266. Free or nominal rate copies mailed at other classes through the USPS (e.g. First-Class mail): average, 57; actual, 61. Free or nominal rate distribution outside the mail (carriers or other means): average, 156; actual, 150. Total free or nominal rate distribution: average, 213; actual, 211. Office use, leftover, unaccounted, spoiled after printing: average, 408; actual, 433. Total: average, 6,074, actual, 5,910. Percentage paid: average, 96.24; actual, 96.15.

If you have been searching for an easy way to authority control your library's current cataloging, try LTI's Authority Express service.

With Authority Express, a library uses the Internet to transmit a file of newly cataloged bibliographic records to LTI (via ftp). LTI immediately processes the records through its state-of-the-art authority control system. Then, at the library's convenience, it logs into LTI's ftp server to retrieve fully authorized catalog records, along with linked LC name and subject authority records.

Authority Express
- Keeps authority control current at an affordable price
- Integrates easily into existing workflows
- Lowers cost by reducing staff time spent on catalog maintenance
- Provides 1 hour turn around for up to 1,000 catalog records

“Authority Control for the 21st Century”

Library Technologies, Inc.
2300 Computer Avenue, Suite D-19  Willow Grove, PA 19090
(215) 830-9320  Fax: (215) 830-9422
(800) 795-9504  email: LTI@LibraryTech.Com

Visit our website at:
www.LibraryTech.Com