

EDITORIAL BOARD

Editor and Chair
Peggy Johnson

Members

- Steven B. Carrico, *University of Florida*
 Elise Calvi, *Indiana Historical Society*
 Allyson Carlyle, *University of Washington*
 Leslie Czechowski, *University of Pittsburgh Health Sciences Library*
 Lewis Brian Day, *Harvard University*
 October Ivins, *Ivins eContent Solutions*
 Edgar Jones, *National University*
 Steven A. Knowlton, *University of Memphis*
 Birdie MacLennan, *University of Vermont*
 Rebecca L. Mugridge, *Pennsylvania State University*
 Randy Roeder, *University of Iowa*
 Carlen Ruschoff, *University of Maryland*
 Sarah Simpson, *Tulsa City-County Library System*
 Anne M. Sleeman, *Community College of Baltimore County Catonsville*
 Lori Terrill, *University of Wyoming*
 Elaine L. Westbrook, *University of Nebraska-Lincoln*
 Lynn N. Wiley, *University of Illinois at Urbana-Champaign*

Ex-Officio Members

- Charles Wilt, Executive Director, ALCTS
 Alice Platt, *Southern New Hampshire University*, Editor, ALCTS Newsletter Online
 Norm Medeiros, *Haverford College* Book Review Editor, LRTS

Editorial

Peggy Johnson



I am writing this editorial in late December in very unusual weather for Minnesota. We have no snow on the ground and yesterday the temperature was the same in Minneapolis as it was in Atlanta. Many things are veering from the expected. Does that mean we should not plan for the future because the future is so unpredictable? I still have my winter emergency supplies loaded in the car even though this winter has been mild and is likely to continue milder than last year. I consider this a form of contingency planning. Contingency planning is directed toward preparing for one possible and usually undesirable future. Libraries prepare disaster contingency plans. Such a plan begins by identifying the possible disaster, such as a flood, and consequences for facilities, services, and collections, and develops appropriate steps to respond to those circumstances.

What other types of planning do libraries use? Master planning is top-down planning that begins in an administrative office. Unit plans are prepared consistent with the master plan. This approach, although it means that initiative at the unit level is constrained and the master plan may not account for the reality of day-to-day work, is fairly simple because not much is open to debate. Strategic planning has an external focus and requires continually monitoring changes in the environment to devise an appropriate response. It usually begins with a vision of the organization's future that serves as a guide to crafting goals, objectives, and strategies that form the plan. Strategic planning is broadly participative and remains an open-ended, continuous process revised at regular intervals. Scenario planning considers alternative futures and formulates strategies in each of those futures. A common example has been planning for various budget allocations—what would the library do with a 3 percent increase, a flat budget, a 3 percent decrease, a 5 percent decrease? Opportunistic planning, sometimes called entrepreneurial planning, is a laissez-faire approach and implies acting immediately when and while an opportunity arises. This approach can make librarians frantic because they do not know what is coming next and when or how they will have to do something completely different.

Drucker wrote that formal planning is improving the “futuraity” of decisions.¹ Some of the approaches to planning describes above do so, others do not. I have spent much of my professional career involved in planning—for a new automated system, a system upgrade, a revised cataloging code, a retrospective conversion project, collection moves, new facilities, a withdrawal project, and more. I have found that the more detailed the plans, the better the projects go. If, for example, a collection must be moved by a certain date, one develops a project calendar working backward from that date. Gantt charts are a useful tool for mapping the sequence of steps and the dependencies, clarifying tasks and responsibilities, and tracking the completion of each assignment. I love a plan because it can look so orderly. The caveat is that projects never go as planned. Eisenhower is credited with saying “In preparing for battle, I have always found that plans are useless but planning is indispensable.”² Most projects are not as orderly as one would like, but starting with a plan and recognizing that it will

have to be revised as the project advances will improve the futurity of any initiative.

One of my favorite gifts this past Christmas was a Lego kit to build the Frank Lloyd Wright–designed house, Fallingwater. It contained 801 pieces and took me nearly eight hours to complete. I loved it! I have been pondering why the project was so much fun. I think one reason I enjoyed it is because Lego kits are so orderly—the instructions are detailed and clear, one completes each step (which may have several steps within it) before moving on to the next, and, when one does the steps in order, the result is an impressive

project one can admire. I wish all library projects came with printed, spiral-bound instructions, and were as straightforward and as satisfying as a Lego kit.

References

1. Peter Drucker, *Management: Tasks, Responsibilities, Practices* (New York: Harper & Row, 1974): 125.
2. Notable Quotes, Dwight D. Eisenhower Quotes, www.notable-quotes.com/e/eisenhower_dwight_d.html (accessed Dec. 27, 2011).



ARCHIVAL.COM
INNOVATIVE SOLUTIONS FOR PRESERVATION

Call for a complete catalog

<i>Pamphlet Binders</i>	<i>Polypropylene Sheet & Photo Protectors</i>
<i>Music Binders</i>	<i>Archival Boards</i>
<i>Archival Folders</i>	<i>Adhesives</i>
<i>Manuscript Folders</i>	<i>Bookkeeper</i>
<i>Hinge Board Covers</i>	<i>Century Boxes</i>
<i>Academy Folders</i>	<i>Conservation Cloths</i>
<i>Newspaper/Map Folders</i>	<i>Non-Glare Polypropylene</i>
<i>Bound Four Flap Enclosures</i>	<i>Book Covers</i>
<i>Archival Binders</i>	<i>CoLibri Book Cover System</i>

ARCHIVAL PRODUCTS
P.O. Box 1413
Des Moines, Iowa 50306-1413
Phone: 800.526.5640
Fax: 888.220.2397
E-mail: custserv@archival.com
Web: archival.com

Cresting toward the Sea Change

Literature Review of Cataloging and Classification 2009–10

Sue Ann Gardner

Sue Ann Gardner (sgardner2@unl.edu) is Associate Professor/Scholarly Communications, Technical Services, University of Nebraska-Lincoln, Lincoln, Nebraska.

Submitted June 30, 2011; tentatively accepted August 19, 2011, pending revision; revision submitted October 10, 2011 and accepted for publication October 26, 2011.

This review was funded in part through a Carnegie-Whitney grant awarded by *Library Resources and Technical Services*.

The author wishes to thank the proposal review committee, especially the chair, Peggy Johnson, who generously granted the funds to help defray the cost of supplies, books, and wages for an assistant. Thanks, also, to Charles Wilt, Executive Director of the Association for Library Collections and Technical Services (ALCTS), for facilitating the grant.

The author appreciates Anna Sophia Cotton's expertise and diligence in turning up hundreds of potential citations.

Thanks to Sherab Chen and Susan Massey, the chair and vice chair of the ALCTS Cataloging and Classification Section, Cataloging and Classification Research Interest Group, who invited the author to speak about this project at the 2011 American Library Association Mid-winter Meeting.

Thanks are due to Mary Bolin, Chair of Technical Services at the University of Nebraska-Lincoln, who gave the author sufficient time to work on this large project.

This review covers cataloging and classification literature published in 2009 and 2010, including relevant explorations of knowledge organization systems and theory. Only English-language literature is reviewed, though not all of the literature covered is U.S.-based. Overarching themes presented in the literature include the merging of library metadata into the Web environment, the continuation of cooperative cataloging in libraries, the role of both controlled and uncontrolled headings in catalog records, and reconsiderations of workflow in light of impending changes to cataloging rules. Notably, several relevant foundational documents were either completed or revised during the review period.

The literature covering cataloging and classification in 2009 and 2010 reflects the dramatic changes taking place in cataloging and shows that cataloging as a means to resource discovery is evolving. With nascent efforts to integrate bibliographic data into the Web environment, in the form of linked data and the Semantic Web, nothing less than a sea change is emerging. The literature reflects the many creative approaches being taken to adapt to this potential reality, such as experimentation with *FRBR*-ized catalogs, based on the *Functional Requirements for Bibliographic Records (FRBR)*.¹ The limitations inherent in the more than forty year old Machine Readable Cataloging (MARC) format have been identified and shared, paving the way for acceptance of a more interoperable format. Several foundational documents, such as the *Statement of International Cataloging Principles* and others, were either revised or completed during the review period, eliciting analysis and commentary.² The use of controlled (i.e., subject, name, and series) headings and uncontrolled headings (i.e., tags) in records has been examined at length. Traditional concerns regarding cooperative cataloging and workflow also are well-represented in the literature, but often within the context of changes in the culture at large. Additionally, some excellent forays into the history of cataloging and classification were published.

Method

The author and an assistant identified some 450 possible publications to review that appeared in 2009 and 2010. Under the direction of the author, Anna Sophia

Cotton, a recent graduate of the University of Missouri-Columbia School of Information Science and Learning, compiled citations from August through October 2010 using *Library, Information Science, and Technology Abstracts*, and Google. Terms searched included: cataloging, bibliographical control, information organization, AACR2, RDA, MARC, authority control, classification, DDC, subject heading, LCSH, FRBR, metadata, OCLC, WorldCat, and Semantic Web. The author scanned the tables of contents for 2009-2010 issues of highly relevant journals to ensure that especially pertinent sources were revealed. The author subsequently performed circle searches by scanning the bibliographical references found in sources. The author also scanned tables of contents for appropriate publications listed in the Directory of Open Access Journals (DOAJ) under Social Sciences—Library and Information Science, such as *Ariadne* and *Code4Lib Journal*.

The focus of the selected sources is on the practice and theory of bibliographical cataloging and classification, including knowledge organization systems and theory. The author limited the universe considered to English language sources published as journal articles, conference papers, monographs, reports, and policy documents, electing those of greatest significant or representative of a topic. The volume of publications precluded being exhaustive. The literature review is organized into the following categories:

- Bibliographic Standards, Principles, Formats, and Codes
- Library Data in the Web World
- Workflow
- Cooperative Cataloging
- Personnel and Education
- History of Cataloging and Classification
- Classification
- Controlled and Uncontrolled Headings
- Cataloging Special Formats

Bibliographic Standards, Principles, Formats, and Codes

An internationally-constructed foundational document, built “on the great cataloging traditions of the world,” the *Statement of International Cataloging Principles (ICP)*, was released by the International Federation of Library Associations and Institutions (IFLA).³ This concise “statement replaces and broadens the scope of the Paris Principles from just textual works to all types of materials and from just the choice and form of entry to all aspects of bibliographic and authority data used in library catalogues.”⁴ The *ICP* incorporates the entire framework of *FRBR*, includes discussion of the entities from each *FRBR* Group, and reiterates the functions of the catalog as outlined in *FRBR*, namely, *find*,

identify, *select*, and *obtain*. Pointedly, the International Standard Bibliographic Description is referred to in the *ICP* as “the internationally agreed standard” for catalog record construction and display.⁵ Throughout the *ICP*, the convenience of the user is invoked.

Also fully incorporating the concepts put forth in *FRBR* was the draft consolidated edition of the *International Standard Bibliographic Description (ISBD)*.⁶ This draft states that “the main goal of the *ISBD* is, and has been since the beginning, to provide consistency when sharing bibliographic information.”⁷ A significant addition to the 2010 consolidated edition draft is *Area 0: Content Form and Media Type Area*, which takes the place of the general material designation (GMD) formerly included in *ISBD* Area 1. The recently-created MARC21 fields 336 (content type), 337 (media type), and 338 (carrier type) will contain the information that resides in the new *ISBD* Area 0.

Several papers analyzed these two publications. Guerrini wrote two. One described in detail the new *ICP*.⁸ The other, co-authored with Bianchini, considers the *ICP* in relation to *FRBR*, the *ISBD*, and Resource Description and Access (RDA).⁹ Bianchini and Guerrini make the point that the *ICP*, the *ISBD*, *FRBR*, and RDA have been formulated by disparate bodies, and consultation among them would have produced a more unified suite of documents. They see the *ICP*, the *ISBD*, and *FRBR* as coextensive and resting on long-held theoretical bases for the organization of information. While they declare, “we should support the RDA effort,” they have many critiques of RDA, both in its construction and in its fundamental underpinning.¹⁰ In their most forceful plea, they believe that the *ISBD* should be acknowledged directly and prominently in RDA as the internationally-agreed upon standard, not merely for purposes of consistent display, but also for the “instructions for data analysis, that is, it gives stipulations to the cataloger to search for and recognize data, to define the functions of each data element within the specific context and to ascertain the proper position for recording the data element within the areas of the description.”¹¹ They see the *ISBD* as more than a prescription for punctuation and placement of elements in a catalog record—it is the grammar of cataloging, giving catalog data meaning and, therefore, is essential to bring coherence to any cataloging standard.

Creider compares the Paris Principles with the 2009 *ICP*.¹² He chronicles some of the political machinations that led to the *ICP* and extols the international nature of its development. He notes, “in sum, eighty-one countries participated in the various IME ICC [IFLA Meeting of Experts on an International Cataloguing Code] sessions,” and affirms, “this is great progress.”¹³

Functional Requirements

Each of the three functional requirements documents was

issued (at least in draft) by 2010.¹⁴ After eighteen years of development, a study group under the auspices of IFLA produced an amended and corrected version of *FRBR* in February 2009.¹⁵ This clearly-written report includes the background of the initiative to produce the document as well as a thorough description of all of the elements that comprise *FRBR*, namely, *entities*, *attributes*, *relationships*, and an explication of *user tasks* (which mirror the functions of the catalog, as outlined above, namely, to *find*, *identify*, *select*, and *obtain*), and *basic requirements for national bibliographic records*.

Functional Requirements for Authority Data (FRAD) also was presented in a final form in 2009.¹⁶ *FRAD* is intended to be an adjunct to *FRBR*, further exploring and extending the Group 2 entities of *person* and *corporate body*. One main contribution of *FRAD* to the *FRBR* model is the addition of the element *family* to the Group 2 entities.

Functional Requirements of Subject Authority Data (FRSAD), released in June 2010, is highly conceptual and ultimately does not add much to the possible application of the general *FRBR* model.¹⁷ It suggests that the original Group 3 entities defined in *FRBR*, namely, *concept*, *object*, *event*, and *place* should remain, with no others added to the third group, though the authors suggest that Group 1 and Group 2 entities also be considered as potential subject entities. A key concept posited in *FRSAD* regards the definition of the phenomenon of *thema*, which are subject concepts, and associated *nomens*, which are signifiers by which the *themas* are known.

As an entity-relationship model, *FRBR* lends itself to Semantic Web applications. As Coyle says in *RDA Vocabularies for a Twenty-First Century Data Environment*,

The great value of using entities and relationships is that they allow the creation of a network of connections that goes beyond the description of a single item, more accurately reflecting the rich interaction between the intellectual creations that are being cataloged . . . it is these relationships that could transform library data into a true information network rather than a mere list of individual items.¹⁸

Coyle acknowledges that the application of *FRBR* is in what she terms beta mode, but she is confident that it holds a key to the Semantic Web.

Explanation and analyses of the *FRBR* family of concepts abound and some testing has been done. Copeland gives a clear summary of *FRBR* and RDA and Gemberling examines the concept of *thema* and *FRBR* Group 3.¹⁹ Gemberling talks about *FRSAD* in a concise and straightforward manner, and addresses the oft-vexing problem of whether names for buildings should be established as corporate entities or as thematic subject entities. He shows how decisions

regarding something such as this are arrived at arbitrarily and then the consequence of that arbitrary judgment gets carried forward with zeal, undeserving of the original intent behind the decision. His is a cautionary tale that should inform the eventual general application of the concepts outlined in *FRBR*, *FRAD*, and *FRSAD*.

Machine-Readable Cataloging (MARC)

MARC, as a legacy dataset, may never be fully extended into the wider Web, though some enterprising coders do their best to manipulate “a data format built to contain catalog records,” as opposed to bibliographic data in and of itself, as Thomale points out.²⁰ Thomale, from a coder-turned-cataloger’s perspective, sees that MARC catalog records are structured data rather than data records. That is, MARC records replicate information on a catalog card, in that format, and do not neatly encode for discrete bibliographic elements, which would make them eminently more extensible in the Web environment.

The RLG Partnership MARC Tag Usage Working Group considered some aspects of the implications of MARC tag usage on library metadata practices.²¹ They discovered that of nearly 200 available MARC tags, 102 are used with significant frequency and eighty-six are little used or not used at all. Through their analysis and from user studies, they have determined that the following tags are the most meaningful to users or are very highly used (or both): 245 (title and statement of responsibility), 260 (imprint), 300 (extent), 1XX/7XX (main/added entry), 65X (subject), 505 (contents note), 520 (summary note), 856 (online location and access), and 020 (ISBN). One conclusion they draw is that, “With more text indexed by search engines, focus should be on the authorized names, classifications, and controlled vocabularies that key word searching of full-text will not provide.”²²

Eklund and colleagues also conducted an empirical study of MARC content designation use in WorldCat records.²³ Their goal was to determine if MARC tag usage mirrored required elements in national-, core-, and minimal-level records, and found that they do not. The authors recommend that empirical studies, such as theirs, be consulted when standards are formulated, to ensure that cataloging practices emphasize the useful aspects of a catalog.

Mayernik enacted a power law analysis of the distributions of MARC fields in the Library of Congress (LC) online catalog.²⁴ With respect to bibliographic records, a power law analysis will show that a few fields will be used most frequently in records and the majority of fields will be used less frequently. Mayernik’s data show that the 1XX/7XX (main/added entry), 245 (title and statement of responsibility), 260 (imprint), 300 (extent), and 6XX (subject) fields were used most frequently, and the 5XX (notes) and 4XX (series) fields

were used next most frequently in his 1,500-record sample. He found that 76X-78X, 130, 240, and 250 fields were used infrequently.

Godby explains the specifics of crosswalking MARC with ONIX, EDItEUR's Online INformation eXchange communication standard.²⁵ She notes that "The two standards are structurally and semantically different because they support different needs and communities of practice."²⁶ Although because "the ONIX standard was proposed some thirty years after MARC was first adopted by libraries, it was informed by lessons learned from users of MARC."²⁷ As outlined in the Godby report, the set of semantic correspondences devised by OCLC staffers Renee Register and Bob Pearson is affected in part by the library community's adherence to *ISBD* presentation. Register and Pearson discovered that MARC-to-ONIX and ONIX-to-MARC converse mappings can suffer due to ONIX's presentation-agnostic aspect and MARC's facilitation of the *ISBD* display format. Godby describes in detail how these issues often have been overcome, thanks to clever mapping by developers who are both well-versed in library metadata and understand the perspective and intentions of the book-selling community.

Resource Description and Access (RDA)

Mayernik includes an important section in his paper on how RDA's "organizational scheme draws on the conceptual models found in the *FRBR* and *Functional Requirements for Authority Data (FRAD)* reports. . . . The material-specific rules—such as those for music, recorded sound, video, etc.—are mixed into each chapter of *RDA* rather than the *AACR2* practice of giving them their own chapters."²⁸ To catalog an item of printed music, he notes, a cataloger would have to consult pages 25, 56, and 194 of *RDA* chapter 2, and pages 30 and 134 of chapter 3, whereas in the *Anglo-American Cataloguing Rules*, 2nd ed., (*AACR2*), all of these rules are found within the twenty pages of chapter 5.

Knowlton examines how legacy formats such as reproductions, facsimiles, and microforms will be cataloged using *RDA*.²⁹ He has an excellent section on cataloging reproductions using earlier codes and outlines the multiple versions problem succinctly. He details the technical aspects of microforms that can be included in an *RDA* record, which are more extensive than those prescribed by *AACR2* in many cases. Knowlton, like Mayernik, decries the "dispersal of applicable rules throughout the code. . . . [thus] a cataloger needs to consult up to six separate chapters and appendices to be sure of applying the correct rules to a reproduction in hand. There is no rubric, such as that found in each chapter of *AACR2*, to guide catalogers through general rules as they apply to particular formats."³⁰

Hillman and colleagues wrote a paper containing details of the development and registration (in the National Science

Digital Library Registry) of the *RDA* vocabularies that will likely be used in Semantic Web applications.³¹ They explain that "a key aspect of a registry is that it can provide a unique identifier (URI) for each data element and for each member of a vocabulary as well as one for the vocabulary or element set as a whole. With registered elements and vocabularies, labels can be identified for different languages or different communities, though the identifier can remain the same."³²

Library Data in the Web World

Library metadata are currently highly textual in nature, as they are in the *MARC* format. *MARC* fields are not constructed like other machine-manipulable data, in strings of discrete data elements, arranged employing a simple syntax. Each *MARC* field employees varying syntax and cannot be machine-parsed consistently using algorithms. Library metadata will need to be more truly machine readable to be more readily incorporated into the Web.

If a reader is new to the concept of non-*MARC* metadata and the Semantic Web, a good place to start is Dunsire's "The Semantic Web and Expert Metadata: Pull Apart Then Bring Together."³³ He includes a historical explanation of bibliographic metadata representation, from catalog card through to a Semantic Web schematic. He notes that "information professionals trying to create structured, accurate and comprehensive metadata cannot keep up using 'traditional' methods. Instead . . . we need to get our machines to process metadata as effectively as they process data."³⁴ Dunsire explains how triplets in the Resource Description Framework (RDF) are constructed, these triplets serving as the basic units of the Semantic Web, and he elucidates the components and mechanisms that are involved in Semantic Web development. He uses dozens of terms and acronyms in his paper and succinctly describes each one.

A good place to progress from Dunsire's Semantic Web paper is to Coyle's series of papers that were published in *Library Technology Reports*.³⁵ Five of the six chapters in the January-March 2010 *Reports* provide a detailed overview of Semantic Web concepts, starting with the question "how can the library catalog move from being 'on the Web' to 'of the Web'?"³⁶ Coyle answers this question, in part, in "Changing the Nature of Library Data," by noting that catalog records need to be less textual and constructed of pure strings of data so that computers can recognize the data elements in them so those elements can be enfolded readily into the Web environment. Coyle observes that current catalogs present static data about resources, and these systems do not take advantage of Web technologies that afford dynamism, extensibility, and currency to metadata surrogates. Bradley also considers extending library metadata into the Web environment, specifically addressing the concept of linked data.³⁷

Yee offers a thoughtful exploration of whether bibliographic data can be put directly on the Web.³⁸ She wonders about issues such as authoritative provenance. For example, are the data we will be using going to be vetted some way, and assuredly correct? She brings up many other concerns. Will data points be brought together to correctly describe a bibliographic entity? Will the human-machine partnership be adequately considered when constructing Semantic Web-related protocols? How granular will the display of data be to the user? Yee points out that the concepts of entity/class and attribute/property are difficult to parse ideologically and will certainly, therefore, be difficult to implement in practice. She also points out that “bibliographic data [are] rife with hierarchy” and that she does not see how this issue is accounted for in Semantic Web development.³⁹

Tolkoff also lists some potential pitfalls with regard to interoperability including problems with transliteration and translation, as well as issues related to use of diverse subject heading systems internationally.⁴⁰ She notes that projects such as the Virtual International Authority File (VIAF) and Multilingual Access to Subjects (MACS), if successful, could ameliorate some of these problems. Tolkoff believes that linking of these files could offer an elegant solution, though the currently-available interface is inadequate and will need improvement to be most effective.

After reading numerous papers on non-MARC metadata, the Semantic Web, and interoperability, a reader may be ready to take in the graphic depiction of the metadata universe, *Seeing Standards: A Visualization of the Metadata Universe*, that Riley and Becker constructed.⁴¹ Accompanied by a glossary of terms, the graphic groups 105 currently-used, cultural heritage-related metadata phenomena based on *domain*, *function*, *purpose*, and *community*. Riley and Becker’s *Seeing Standards domains* include: cultural objects, datasets, geospatial data, moving images, musical materials, scholarly texts, and visual resources; *functions* include: conceptual model, content standard, controlled vocabulary, framework/technology, markup language, record format, and structure standard; *purposes* shown are: data, descriptive metadata, metadata wrappers, preservation metadata, rights metadata, structural metadata, and technical metadata; *communities* listed are: archives, information industry, libraries, and museums. The graphic is valuable for its breadth and for its attempt at categorization.

To what extent has the cataloging community embraced non-MARC metadata creation and interoperability? As the Yang, Lee, and Xu article reveals, not much.⁴² Similarly, Park and Tosaka conducted a survey that shows that only 9.9 percent of the 263 total respondents used more than three schemata during the metadata-creation process.⁴³ Ultimately, their results show that MARC continues to be the metadata schema of choice for digital collections.

Workflow

Papers by Ma, Toy-Smith, and Veve and Feltner-Reichert covered incorporating non-MARC metadata creation into catalogers’ workflow.⁴⁴ “What We’ve Learned from the RLG Partners Metadata Creation Workflows Survey” is an assessment compiled from a 2008 survey of 121 Research Libraries Group (RLG)-affiliated librarians showing that “the tools being used are very localized, and no one tool kit is being used,” and “institutional routines are not yet standard enough for inter-institutional collaboration.”⁴⁵ Ayers and her colleagues challenge the assumption that users will search using library sites instead of Web search engines, though the team learned that library data are available via multiple pathways, such as through various Web crawlers and the Open Access Initiative-Protocol for Metadata Harvesting (OAI-PMH) harvesters. Respondents reported that they are seeking to streamline workflows, especially for application of descriptive elements and for creation of procedures. The team concludes that this is still the initiation period for metadata creation in libraries and poses questions for further inquiry.

Valentino describes a project to integrate digital library metadata creation into the cataloging department at the University of Oklahoma Libraries.⁴⁶ She details the process, including the training period and the testing period, which may serve as a template for others wishing to do the same. She concludes that cataloging staff are logical partners with digital library staff and can free the digital projects librarians to work more on digitizing and less on time-consuming metadata creation.

In “Mountains to Molehills: The Past, Present, and Future of Cataloging Backlogs,” Howarth, Moore, and Sze conducted a review of the literature on cataloging backlogs.⁴⁷ One phenomenon they mention is something that Sarah Thomas called “bibliographic chicken,” in which catalogers leave an item in the backlog while they wait for someone else to create an original record for it and contribute the record to a shared database.⁴⁸ The authors point out that not all library managers wish to eliminate backlogs entirely because they offer a reserve of materials to work on when acquisitions have slowed.

Co-commissioned by the National Information Standards Organization (NISO) and OCLC Online Computer Library Center, Luther prepared a well-written and informative white paper, *Streamlining Book Metadata Workflow*.⁴⁹ This report encapsulates the methods used by publishers, metadata vendors, book sellers and wholesalers, national and other libraries, and Google, to exchange book-related data, primarily in ONIX and MARC. Luther includes statistics about the size of the market and information about how the stakeholders in the sector work together, though their metadata operations have not been fully standardized yet. For example, Nielsen Book, a commercial metadata vendor in the United Kingdom

processed 43.2 million records in 2008, the vast majority of which involved updates on price and status (the status might generate a “Hurry, there are only 3 left!”-type of message that a potential purchaser would see on a product page online)—critical elements for commercial entities, that must be absolutely current at all times. Luther reports that “OCLC hosted a Symposium for Publishers and Librarians to explore metadata needs and practices,” including the issue of various identifiers (such as the ISNI (International Standard Name Identifier), ISBN (International Standard Book Number), DOI (Digital Object Identifier)), series, related works, and subject schemes.⁵⁰ Many creative ideas for best practices were generated during the symposium.

WorldCat Local is the focus of a paper by Zhu of Washington State University Libraries, where they have integrated *WorldCat WSU* with their Innovative Interfaces integrated library system.⁵¹ Zhu lists the technical issues associated with configuring WorldCat Local and outlines the specific points at which technical services personnel were involved with the implementation of the product. Some local cataloging policies and procedures were reconsidered in light of the switch and the Griffin catalog had to be prepared in a variety of ways. Zhu feels that migrating to WorldCat Local will allow deferred projects to move forward, such as cataloging of hidden collections, retrospective conversion, digitization, and populating the institutional repository with content.

Kim describes an outsourcing effort at the Hanyang University Library in Seoul.⁵² The impetus for this project was to eliminate employees who conducted the “nonproductive and insignificant repetitive work” of cataloging and “to alleviate the bottleneck in employee promotion.”⁵³ All library-employed catalogers were fired and replaced with contractually-hired employees. Library staff supervised the new cataloger, and the replacement catalogers occupied the same space in the library as the former employees. Kim says, “since the contract workers are not library employees, they are not highly interested in the affairs of the library,” and “since [they] work according to the conditions specified in the contract agreement, it would be difficult to expect sincere participation from them.”⁵⁴ Kim says that “rather than forcing the library’s abstract vision onto the employees, the management should communicate with them to share its specific goals and vision,” then declares in turn that “every member of the department must share Hanyang University Library’s vision of becoming ‘the foundation for providing the latest knowledge and information for the nurturing of global leaders.’”⁵⁵

Cooperative Cataloging

In separate papers, Schuitema and Sellberg examine the likely role of cooperative cataloging in a post-online public access catalog (OPAC) world.⁵⁶ Both take a historical view

and Schuitema examines selected landmark developments. She wonders if librarians have reached an impasse in cooperative cataloging. Sellberg envisions librarians working cooperatively not to maintain and catalog redundant collections in each separate library but, instead, she says, “if one thinks about pooling the expertise of metadata experts and working together to facilitate use of the world’s information resources through well-designed and well-managed systems of access, then we have barely begun to realize the possibilities.”⁵⁷

Proving that cooperative cataloging activities are active in the United States today, articles by Banush, Charbonneau, and El-Sherbini cover the Program for Cooperative Cataloging (PCC) and its associated initiatives, BIBCO (Monographic Bibliographic Record Program), CONSER (Cooperative Online Serials Cataloging Program), NACO (Name Authority Cooperative Program), and SACO (Subject Authority Cooperative Program), and other cooperative cataloging ventures.⁵⁸ Citing funding challenges in light of continual changes to information technology and the library community’s inability to adapt quickly enough, Banush cautions that PCC is in a vulnerable position. He believes that NACO and SACO, with their purpose devoted to authority creation, may hold promise as one viable future for the PCC, and stresses that “the Program will have to change . . . if it chooses to be an important influence on the future of bibliographic control.”⁵⁹

Spiteri used the heuristics of *communication*, *identity*, and *perception* in examining and evaluating the social features and completeness of the catalog records of sixteen social cataloging websites.⁶⁰ She found that “although the bibliographic content of most of the catalog records examined was poor when assessed by professional cataloging practice, their social features can help make the library catalog a lively community of interest where people can share their reading interests with one another.”⁶¹

The economic aspects of cataloging, as they relate to cooperative endeavors and other considerations, have come under scrutiny. The LC commissioning of R2 Consulting to conduct a study of the North American MARC records marketplace is evidence of this.⁶² Nine hundred seventy-two libraries and seventy vendors contributed to the R2 study data set. Fischer and Lugg reveal that LC records are underpriced, the market does not pay well for original cataloging, excessive editing of records for use in cataloging silos occurs, and librarians are unaware of restrictions on MARC record use or redistribution. The authors point out that “the prevalence of open databases is a key factor in the economic confusion that plagues the MARC Record Market.”⁶³

Fischer and Lugg distinguish between the community value system that drives library operations and the commercial values of vendors, and report a disconnect between the desired activities of the cataloging community and the true costs associated with those activities. They note that the LC is

expected to perform a role as an über-provider of cataloging output that it cannot continue with its projected budget. The report also emphasizes that the LC's cataloging contributions are critically important to school and public libraries, which often have limited or no in-house cataloging expertise.

Regarding cooperative cataloging in the United States, a mere ten libraries contributed two-thirds of the BIBCO records in 2008. Fischer and Lugg point out that "a cooperative system only works well if everyone participates."⁶⁴ Additionally, according to their calculations of numbers of catalogers in North America, "we should not have backlogs."⁶⁵

Wolven challenges every aspect of cataloging mores, among them the fiction that librarians catalog items based on what users want.⁶⁶ He points out that librarians actually catalog what he terms units of commerce, such as books and serials, whereas patrons are really interested in specific poems in the books, certain articles in the serials, a particular song on a recording, and so on. Wolven suggests that "the scope of the library catalog that eventually emerged wasn't shaped by user needs . . . but by technological and economic limitations," and continues, "the incentive for libraries to trade descriptions of widely distributed published books drove standardization in ways that didn't apply to competing publishers of indexes and bibliographies, or to manuscripts and other unique materials."⁶⁷ He frames his comments in light of the changes on the horizon for cataloging and believes that "gradually, we will probably reach a new consensus on best practices, less grounded in 20th-century publishing patterns," and pronounces that "we must recognize that the questions are changing."⁶⁸

Personnel and Education

The evolution affecting bibliographic cataloging and classification extends to personnel. Those formerly called catalogers are increasingly referred to as metadata librarians and similar labels. This leads Hruska to consider, "Where Are We with the Staffing Transition from Cataloging to Metadata Management?"⁶⁹ She answers her own question in part when she says, "I believe that those operations that have traditionally been considered cataloging are well along in a transition that parallels the transformation of the content of library collections."⁷⁰ She continues, "there is a marked shift in research libraries to focus more staff effort on including products of local digitization in an integrated discovery experience, moving beyond the silo of a library catalog. It is this work that is actualizing metadata construction as an essential part of library operations, describing the growing collections of digital objects."⁷¹

Mitchell, Thompson, and Wu describe their experiences with transitioning technical services staff to better reflect emerging models of discovery.⁷² They say they want a more

fluid dynamic among the personnel in technical services so that staff are able to take on new roles readily, with autonomy and with the proper training behind them. At the University of Northern Colorado (UNC), Leffler and Newberg report that several possible organizational structures were considered after a key staff member in technical services resigned unexpectedly.⁷³ With broad staff input, UNC selected a final organizational model based on research conducted to determine future directions in technical services there.

Three studies examined the changing role of the metadata librarian. Each employed the technique of quantifying elements in job descriptions to assess trends. In their paper, Han and Hswe explain that they wanted to examine the different responsibilities and competencies that metadata librarians have compared with those of cataloging librarians.⁷⁴ They looked at job descriptions that were in advertisements for open positions between 2000 and 2008 and discovered that the skill set required of metadata librarians has evolved over that time, indicating that more than a name change is occurring. The other two studies were conducted by Park and her colleagues.⁷⁵ In both, they looked at advertisements posted in the mid-2000s, used co-term and co-citation analyses to construct co-occurrence matrices, and visualized these findings through multidimensional scaling and cluster analyses. In addition to noting that managerial-level positions are in high demand, they discovered that "the advancement of technology has affected every aspect of the cataloging profession: job titles, competencies/skills, and responsibilities."⁷⁶

Even for those whose titles remain simply cataloging librarian and variations thereof, changes are afoot in their daily work. Hitchens and Symons give a well-thought-out, detailed template for RDA training for catalogers at all levels.⁷⁷ They consider many pertinent issues and make suggestions such as mapping commonly used rules from *AACR2* to analogous ones in RDA. They point out that experienced catalogers will need to get into the habit of looking up rules in RDA for things with which they were very familiar with when using *AACR2*.

Cox and Myers conducted a 237-response survey of staff at Association of Research Libraries member libraries showed that one central perceived difference between paraprofessional and professional catalogers was expectation for involvement in professional development activities.⁷⁸ Professional catalogers frequently report that they are expected to produce articles and other research output and to serve on committees, while almost no paraprofessionals report having such expectations placed on them.

Two articles reference paraprofessionals' focus on training issues. Sapon-White gives a well-considered plan for training paraprofessionals to perform subject assignment of electronic theses and dissertations.⁷⁹ Shrinking budgets adversely affect numbers of professional catalogers

in academic libraries, so paraprofessional staff are left to perform higher level work. He details the pitfalls that can arise and gives measured suggestions for addressing the various issues. Using the skills she acquired as a German-language teacher, Valente describes an approach to training paraprofessionals that focuses on the beginning days of employment.⁸⁰ She includes long lists of catalogers' tasks and required abilities and includes some examples from work sheets that can be used in training.

Hudon considers the teaching of classification from 1990-2010 and is heartened to see that classification and knowledge organization are still taught in library schools, citing others who also feel it should remain in the curriculum.⁸¹ Of the many issues that have influenced the teaching of classification, one that stands out particularly is the change brought about by the use of online tools, like Classification Web and WebDewey. Hudon believes a focus on the theory of classification is necessary in library schools so a student can make informed decisions when he or she is applying classification schemes.

Some articles concern pre-professional cataloging experiences. In a paper with a comprehensive literature review on the topic, a team from Kent State University (KSU) Libraries reports that the Libraries have a formal program for students of the KSU School of Library and Information Science.⁸² The program is beneficial to both the Libraries and the students because, although the program has associated costs, the practicum students are not paid so the library sees a net cost savings. The students get formal training and hands-on experience cataloging before they complete their degrees. An apprentice cataloger gives his perspective at the end of the paper.

History of Cataloging and Classification

Some authors chronicled cataloging and classification history. Beall and Mitchell detail "the history of the representation of the Dewey Decimal Classification (DDC) in the . . . MARC formats, with special emphasis on the development of the MARC classification format."⁸³ As of 2009, the DDC is finally represented fully in the MARC21 format. They explain that concomitant incorporation of DDC fields into the authority format was rejected, however, as the authority format is not extensible enough. With recent changes to MARC, DDC facets can now be shown. Work is underway to expose DDC in the Semantic Web, with uniform resource indicators (URI) scheme development and other explorations.

McIlwaine "traces the history of the Universal Decimal Classification (UDC), its relationship with other schemes, and opportunities for further collaboration."⁸⁴ In summarizing the quest for a "universally acceptable system for retrieval of subjects," she avers that "the problem is to create

a structure that is both universally acceptable and sufficiently detailed to be useful without being overly complicated."⁸⁵ She explains that even the DDC, which is used worldwide, is not the ultimate system, and owes its success to external factors as much as to its robust construction. She discusses the early days of the UDC, whose development was interrupted by the First World War, and had a proviso placed on it that it would not—until much later—be published in English. She concurrently addresses the history of the DDC in her paper and describes how the two systems were developed, often-times at odds with one another.

Slavic contributed a chapter in *Library and Information Science in the Digital Age: Essays in Honor of M.P. Satija* in which she details the history of shelving and call number evolution for bibliographic collections in Europe.⁸⁶ She explains that even something as mundane as labeling of books took much wrangling to devise. She notes that it was not until late in the nineteenth century that call numbers were based on systematic book arrangement determined by a classification scheme.

In "Wholly Visionary," Yee examines the history of the American Library Association (ALA) and the catalog card distribution program at the LC.⁸⁷ She remarks, "the ALA is now dominated by library administrators with shrinking budgets who know very little about the complexities of bibliographic control (other than its expense) and who wonder if the fact that undergraduates are in love with Google might not provide an excuse for libraries to dispense with the information-organization part of their budget entirely."⁸⁸ Yee draws on the context of past challenges and opportunities to show that the decisions made today about how much access to provide via cataloging practices will have a real, on-going affect on people. She encourages librarians to be visionary and bold in their mission to serve the public as a whole and not kowtow to commercial forces, nor bow to uninformed leadership.

Knowlton gives an overview of a debate that took place in the pages of *Library Resources and Technical Services* in the 1950s and 1960s, just before the *Anglo-American Cataloguing Rules (AACR)* were enacted in 1967.⁸⁹ Seymour Lubetzky and others exchanged their thoughts on the best course of action over several years. Codes such as *AACR* do not arise out of a vacuum and with consensus. Knowlton notes that disagreement was resolved "in Paris, [where] 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."⁹⁰

Elkington details some of the ground-breaking research conducted by OCLC, much of which is ongoing and affects many day-to-day cataloging activities.⁹¹ OCLC helped to spearhead the creation of Dublin Core (DC), when they convened a workshop in 1995 with the aim of envisioning a "simple, modular, extensible metadata scheme for

Web-based resources.⁹² With various partners, OCLC also was a key player in the formation of PREMIS, the preservation metadata framework, VIAF, the Virtual International Authority File, and WorldCat Identities, a project that creates an automatically-generated summary page for every name in WorldCat, nearly 25 million names.

Classification

Classification in practice was the focus of a few articles. Mitchell and Vizin-Goetz chronicle the acquisition of the DDC by OCLC in 1988 and what OCLC has done to maintain it, including facilitating translations of it, and mappings of it with subject headings systems.⁹³ Since the acquisition, OCLC researchers have conducted research on DDC-related topics, including attaching URIs to elements in the DDC, and modeling it in Simple Knowledge Organization System (SKOS).

Bibliographic classification on the Semantic Web was considered at the Dublin Core Conference in 2009, when Panzer and Zeng revealed some problems with modeling classification schemes in SKOS, particularly with so-called centered entries and number spans in the DDC hierarchy.⁹⁴ They present the issues in detail in their paper and suggest possible work-arounds.

The DDC was the subject of a highly technical and comprehensive study undertaken by Wang, who explored the limitations and possibilities involved in the successful application of text categorization with the aim of automated assignment of DDC classes for bibliographic items.⁹⁵ After much experimentation, Wang sees the DDC, above all, as a human construct that is based on convention rather than a machine-friendly hierarchy. Wang sees the need to rework the scheme to satisfy the needs of the computer in order to afford the DDC more interoperability.

Classification theory was considered by some authors during the review period. Gnoli parallels, but more fluidly fleshes out, the ideas presented in the *FRSAD* document in his commentary on the distinctions between phenomena and their attendant signifiers.⁹⁶ Gnoli suggests that discrete phenomena encompassing a theme, for example, animals, ought to be taken to be the “basic units of classification,” because, “while disciplines are a traditional way of organizing knowledge, in many cases they also act as a superstructure adding unnecessary perspective to the content itself.”⁹⁷ In the manner that classification schemes are ordered currently, he explains, animals, for example, are considered relative to society’s relationship with them at any point, such as whether they are pets, livestock, wildlife, fossils, or laboratory subjects. Gnoli suggests that animals be presented in a classification scheme simply using their chosen signifier (such as the English word animals and equivalent words

in other languages) and that relationships are separately explicated in the classification scheme. The signifiers for the *phenomena* and, separately, the *relationships* could then be combined in endless ways, thus freeing the classification scheme from imposed perspectives.

Gnoli’s thesis ties in with what Lee reports in “Divination and the State: Classifying Technical Texts in Han China.”⁹⁸ Lee quotes Tsuen-Hsui Tsien to point out that “ancient writings in China were used for communication not only among human beings, but also between human beings and spirits.”⁹⁹ This act was not akin to current-day spiritual explorations, but more like contemporary scientific investigations. Lee also explains that the issuing body in ancient China was often the most salient aspect of writings. Similar to the U.S. Superintendent of Documents (SuDocs) arrangement, whatever written products a certain governmental body issued would be grouped together by virtue of their authorship. Understanding the context in which ancient texts were constructed is paramount in classifying them meaningfully. Without a modern-day equivalent to the activities from earlier times, current classification schemes may offer no proper place for those ancient texts. Gnoli’s proposed construct is one possible solution to this problem.

Not only ancient texts but also those created more recently may not find a place in current classification schemes. The ever-increasing corpus of Islamic texts is one example. Idrees and Mahmood conducted a small but well-thought out survey of ten Islamic and ten information science scholars asking how they believe an adequate Islamic classification might be devised because none exists.¹⁰⁰ The DDC number for Islam, for example, is simply 297. The majority of the scholars in the study favor the idea of creation of an extensive and current classification scheme for Islamic literature.

Some public librarians in the United States are reconsidering application of classification schemes in their libraries’ book arrangements. As Fister reports, some have taken to modeling their book layout according to a version of the Book Industry Standards and Communications (BISAC) system.¹⁰¹ Fister quotes a blogging mom who declares “the books, everywhere, but especially in the children’s room, have been shelved, labeled, and organized in a way that makes me feel less like a moron and more empowered to find what I’m looking for on my own.”¹⁰² According to the survey conducted by Fister, many public librarians share this sentiment, at least to some extent, with more than 85 percent acknowledging that public librarians would better serve patrons if a not-strictly-Dewey system were implemented in their shelving schemes.

Controlled and Uncontrolled Headings

Some authors examined controlled headings in general. Hearn looked at library catalogs at universities that are part

of the Committee on Institutional Cooperation (CIC) and the LC catalog to determine how much quality assurance of controlled headings is taking place in each.¹⁰³ He found a wide variation in the pace of upkeep of changed headings between the catalogs, concluding that a study of this kind can help institutions in a consortium or other selected group of libraries to measure catalog quality based on comparison with that of peers.

In anticipation of adoption of RDA, Burke and Shorten chronicle practices before the cataloging standard changes and compare how authority work is being done in U.S. libraries today.¹⁰⁴ They find that authority work varies more based on a library's size than on the type of library. Larger libraries outsource their authority work more often than is done at smaller libraries and larger libraries are more often NACO contributors. They also found that personal name headings were the most frequently controlled and that uniform titles were least likely to be constructed or edited.

Names

"The DeathFlip Project: Automating Death Date Revisions to Name Headings in Bibliographic Records" is an undertaking of the librarians at Kent State University in their response to the LC's decision to add death dates to name headings as they arise.¹⁰⁵ In early 2006, the LC began to add death dates to name authority records, changing about 500 records per month the first two years after the rule revision. Kent State staff used the OCLC RSS (Real Simple Syndication) feed of the headings changes to initiate a protocol in their catalog to automatically flip the records so the proper, newer headings are incorporated into the catalog.

Dragon and her colleagues have made a concerted effort to control the name headings associated with the Eastern North Carolina Postcard Collection at East Carolina University.¹⁰⁶ Under the premise that "minimization of the necessity of relying on chance for information discovery is the mark of quality metadata," their thoughtful method includes many important points, so that any other group wishing to control name headings for a small collection of images would do well to reference this paper.¹⁰⁷

Subjects

Chung, Miksa, and Hastings explored the use of text categorization (TC) to assign subject terms, using algorithms that are modified with each successive use.¹⁰⁸ They contend that humans do not use the full text when assigning subject headings, so computer algorithms perhaps should not, also. They argue that conceptual frameworks based on human-informed processes can best direct TC by computers. They discovered that keyword was shown to be a more effective source than full text when mining the article for subject

terms, and cited that works, article title and journal title were as effective as mining the full text when assigning headings based on performing TC on a document.

Yi and Chan explored whether LC Subject Headings (LCSH) may be employed as an effective subject access tool in a networked environment.¹⁰⁹ They examined the syntax and structure of the headings and determined that there is both a local relational structure as well as a global hierarchical structure. They conclude that the LCSH retains too many features of natural language to be used successfully as an interoperable controlled vocabulary and the LCSH needs to be more rigorously hierarchical. Ultimately, the syntactic structures are too diverse to parse algorithmically, making the LCSH of limited use in the networked environment.

Series

Sapon-White recounts how the cataloging staff at Oregon State University responded when the LC changed its policy to no longer trace series.¹¹⁰ They undertook the task of counting bibliographic records added over the year-and-a-half study period and noted the source of the records. The set of records with untraced series statements gathered from the LC during the study period was negligible (sixty out of 53,911). Only approximately 900 more than that came through with untraced series in total, thereby indicating that this change was not a burden on the staff and did not warrant a significant change in workflow.

Uncontrolled Headings

Uncontrolled headings are keywords attached to documents, websites, bibliographic records, or other content, that are not from a thesaurus or other list of controlled headings. The term "tag" is often used to denote keywords input by users of content, though expert metadata creators also can assign tags to a document. The aggregate of tags associated with a defined grouping of content is termed a folksonomy. Much research has been done to examine the use of uncontrolled headings in catalog records.

Bianco conducted a survey through which she found that social tagging has not been adopted widely by medical librarians.¹¹¹ In another study reported by Maggio and colleagues, instructors asked medical students at Boston University's Alumni Medical Library to assign tags to several digital objects.¹¹² They asked the students in an Introduction to Biomedical Literature class to answer the question, "What would you call it?"¹¹³ The exercise elicited synonymous terms, spelling mistakes and variations, and variations in specificity, thereby highlighting the importance of using controlled terms when searching for medical literature.

Griffis and Ford demonstrated that subject liaisons can act as partners with catalogers to provide helpful keywords

and descriptions of resources of electronic databases and media items.¹¹⁴ They observe that subject experts can offer high-quality user-generated uncontrolled terms in records that will enhance discovery. The article offers useful suggestions about embarking on an enterprise such as theirs.

Similarly, Strader examines “the overlap between author-assigned keywords and cataloger-assigned LCSH for a set of electronic theses and dissertations in Ohio State University’s online catalog.”¹¹⁵ She reports that the combination of controlled and uncontrolled terms in records improves retrieval, and notes that previous studies have consistently shown the same phenomenon.

Adler conducted a comparative study of controlled terms in LCSH and user-generated tags in LibraryThing for transgender books.¹¹⁶ She declares that “perhaps the greatest power of folksonomies, especially when set against controlled vocabularies like LCSH, lies in their capacity to empower user communities to name their own resources in their own terms.”¹¹⁷ Her research, comparing tags in LibraryThing for transgender-related books with LCSH assigned to the same books, revealed “a disconnect between the language used by people who own these books and the terms authorized by the LC and assigned by catalogers.”¹¹⁸ She concludes that the two sets of vocabulary each have their strengths and their limitations and thus complement each other.

Thomas, Caudle, and Schmitz also analyzed the tags assigned by users in LibraryThing.¹¹⁹ They found that “tag variations [are] the most prominent hindrance to search and retrieval,” variations referring specifically to, “tags which were the same except for tense, symbols, spelling, and capitalization . . . also included noun-adjective combinations and word combinations meaning the same thing or concept,” and acronyms and initialisms.¹²⁰ Ten books were chosen for the study to which 7,653 tags had been assigned and 59 percent of those tags were either variations or contained non-alphabetic characters. The authors determine that folksonomies can augment controlled headings, but cannot replace them.

Sharif refers to folksonomies as Web 2.0 technology and ontologies, or lists of controlled terms, as Web 3.0 technology and what will drive the success of the Semantic Web.¹²¹ Sharif declares that, to be effective, ontologies must be maintained rigorously to accurately reflect the most current approach to a subject. She believes that collaborations with users may be the most effective way to maintain an accurate and current ontology. She developed a model consisting of an ontology of a folksonomy in which relationships between tags, the objects they represent, and the tag assigners, are explicated. These relationships form patterns that then can be exploited and integrated into searching systems to better improve precision and recall.

Likewise, Kakali and Papatheodorou observe that

“the current state of the art on the semantic correlation between folksonomies and knowledge organization systems (KOS) centers upon the analysis and integration of user, creator, expert, and machine generated vocabularies.”¹²² The authors directed a tagging effort by a group of experts and had a group of catalogers inspect the tag choices. The authors also gathered tags from LibraryThing to supplement the tag set. The librarians approved of the majority of the tags input by expert users and a smaller quantity of the LibraryThing tags and found that tags supplement the subject heading assignment and improve searching.

Lawson also concludes that “while social tagging does consist of a great deal of subjective tagging, there is enough objective tagging available on bibliographic-related websites such as Amazon and LibraryThing that librarians can use to provide enriched bibliographic records,” and recognizes that user-generated tags can enhance subject cataloging.¹²³ Rolla concurs as he remarks in his paper on the topic, “user tags can enhance subject access to library materials, but they cannot entirely replace controlled vocabularies such as the LC subject headings.”¹²⁴

Cataloging Special Formats

Several helpful publications address the cataloging of specific material formats. Representative articles are reviewed.

Boock and Kunda compare past processes and workflows for print theses and dissertations with the current workflow for those that are electronic.¹²⁵ Part of the workflow involves student authors themselves entering basic metadata when they deposit their dissertations. Some of the metadata are generated automatically via DSpace, and the library staff add subject headings and verify the student-submitted metadata. Helpfully, the authors include some cost-saving estimates that their workflow proffers.

Wu and Mitchell describe how difficult using vendor-supplied metadata for hundreds of thousands of e-books from several different vendors is.¹²⁶ They use MarcEdit and the SerialsSolution MARC service for e-books to accomplish some cataloging tasks, some of which are batched, and they discuss the benefits of the provider-neutral record approach, namely, ultimately fewer and more-consistent records. They conclude by saying that e-book cataloging on the scale they describe is new for them, but that, with efficient solutions and an awareness of the resource supply chain, they have been able to absorb the increased throughput of materials. That said, they point to issues that still need to be resolved.

Recognizing the ongoing challenge to provide access to ever more digital resources Reese wrote “Automated Metadata Harvesting: Low-Barrier MARC Record Generation from OAI-PMH Repository Stores Using

MarcEdit.”¹²⁷ Reese points out that metadata arise from many sources, not just OCLC as was more common in the past. In describing two potential use cases, Reese shows how MarcEdit “offers default conversion support from OAI-MPH metadata to a number of different metadata formats.”¹²⁸

Beamer thoughtfully examines problem areas in map cataloging and, despite the inherent difficulty in cataloging maps, implores would-be map catalogers to “secure appropriate systems for retrieval and include geographical location information, specifically numerical co-ordinates.”¹²⁹ She looks not only at MARC21 and AACR2 as cataloging tools, but also Encoded Archival Description (EAD) and the Dublin Core in the Resource Description Framework format. She reports about the benefits and detriments of the proprietary Royal Scottish Geographical Society’s Images for All project’s Oracle-based system. Though imperfect, the United Kingdom Ministry of Defence Parsons Classification is used to denote locations in the system, which is useful because it includes non-extant geographic place names. Beamer also discusses the European Library’s open source DIGMAP retrieval system.

Kowal and Martyn consider cataloging a different set of maps, namely those found in books.¹³⁰ This kind of granularity in cataloging becomes feasible when digitizing collections, where image-based information can be easily uncovered and highlighted. After the theft of pages of maps from books, the Vulnerable Collections Item Project was initiated at the British Library and more than 3,000 maps from the fifteenth through the seventeenth century were selected. Maps were cataloged using MARC21 and AACR2 and the analytic records were linked to the parent record in each case. Using the ADAM module to link administrative and rights metadata to the catalog record, digital images of the maps also were linked to the records.

A group of Dutch and French librarians present “an experiment on enhancing the semantic interoperability of two digital iconographic collections: Mandragore, the iconographic database of the Manuscript Department of the French National Library (BnF), and the Medieval Illuminated manuscripts collection of the National Library of the Netherlands (KB).”¹³¹ They believe that the cultural importance of these collections and their similarity justify that “these collections need to be interconnected and made interoperable, in a way ‘smart’ enough to allow users to seamlessly interact with the resulting aggregates.”¹³² They report about an experiment they conducted that employed Simple Knowledge Organization System (SKOS) and an ontology matching technique that automatically identified semantic correspondences based on lexical alignment. They then built an interface to test the model and found that the results indicate that they could achieve their goal of interoperable searching using this method.

Conclusion

As 2010 was declared to be the Year of Cataloging Research by Carlyle and others, this was an especially auspicious time to review the literature.¹³³ Despite the enormous volume of literature published during the review period, some themes merit highlighting. A common research topic was the encroachment on traditional cataloging by non-MARC metadata and the potential for bibliographic metadata’s interoperability with the wider Web, the end-product of which is often referred to as the Semantic Web. Another popular topic was the use of uncontrolled tags in catalogs, often user-generated, invariably showing that tags supplement, but do not supplant, the use of controlled headings in records. The literature reflects a continued emphasis on cooperative cataloging, what it has meant in the past, and what it will look like in the future. Considerations such as these regarding the future of cataloging weigh heavily over the enterprise of bibliographic cataloging and classification.

Academic libraries are represented overwhelmingly in the literature from 2009-2010, though many important and useful papers were published about school libraries. Very few articles addressed the unique concerns of cataloging and classification in public and special libraries.

A couple of surprises arose as the author surveyed the literature. RDA was written about in the context of other topics, but few standalone papers analyzing RDA were published. With such a significant change to the fundamental tool used in cataloging looming on the horizon, the author had expected to locate numerous articles addressing RDA. Also, as never before, the business of cataloging was investigated in some detail during the past two years. Most notably, a large study of the MARC records marketplace study was conducted by R2 Consulting for the LC.

The large number of papers on personnel and workflow issues shows that catalogers are still considering workaday activities, but almost all were posited within the context of impending change. The collective output of 2009-2010 ultimately shows that cataloging and classification continue to be essential activities in libraries, even as catalogers partner more and more with those in increasingly-disparate disciplines.

References and Notes

1. IFLA Study Group on the Functional Requirements for Bibliographic Records, *Functional Requirements for Bibliographic Records: Final Report*, approved by the Standing Committee of the IFLA Section on Cataloging Sep. 1997, as amended and corrected through Feb. 2009 (The Hague: IFLA, Feb. 2009), www.ifla.org/files/cataloguing/frbr/frbr_2008.pdf (accessed Aug. 11, 2011).
2. International Federation of Library Associations and Institutions, *Statement of International Cataloging Principles* (The

- Hague: IFLA, 2009).
3. Ibid., 1.
 4. Ibid.
 5. Ibid., 4.
 6. International Federation of Library Associations and Institutions, *International Standard Bibliographic Description (ISBD)*, Consolidated ed. [draft], IFLA Series on Bibliographic Control v. 31 (The Hague: IFLA, 2010).
 7. Ibid., v.
 8. Mauro Guerrini, "In Praise of the Un-Finished: the IFLA Statement of International Cataloguing Principles (2009)," *Cataloging & Classification Quarterly* 47, no. 8 (2009): 722–40.
 9. Carlo Bianchini and Mauro Guerrini, "From Bibliographic Models to Cataloging Rules: Remarks on FRBR, ICP, ISBD, and RDA and the Relationships between Them," *Cataloging & Classification Quarterly* 47, no. 2 (2009): 105–24.
 10. Ibid., 114.
 11. Ibid., 113.
 12. Laurence S. Creider, "A Comparison of the Paris Principles and the International Cataloguing Principles," *Cataloging & Classification Quarterly* 47, no. 6 (2009): 583–99.
 13. Ibid., 590.
 14. Glenn E. Patton, ed., *Functional Requirements for Authority Data: A Conceptual Model*, IFLA Series on Bibliographic Control v. 34 (Munich: K.G. Saur, 2009); IFLA Study Group on the Functional Requirements for Bibliographic Records, *Functional Requirements for Bibliographic Records: Final Report*; Marcia Lei Zeng, Maja Žumer and Athena Salaba, eds., *Functional Requirements for Subject Authority Data: A Conceptual Model* (The Hague: IFLA, June 2010), www.ifla.org/node/1297 (accessed Oct. 19, 2011).
 15. IFLA Study Group on the Functional Requirements for Bibliographic Records, *Functional Requirements for Bibliographic Records: Final Report*.
 16. Patton, ed., *Functional Requirements for Authority Data: A Conceptual Model*.
 17. Zeng, Žumer, and Salaba, eds., *Functional Requirements for Subject Authority Data: A Conceptual Model*.
 18. Karen Coyle, "FRBR, the Domain Model," in *RDA Vocabularies for a Twenty-First Century Data Environment*, 20–25, *Library Technology Reports* 46, no. 2 (Feb.-Mar. 2010) 46(2): 21, 23.
 19. Jud H. Copeland, "RDA and FRBR: A Brave New World in Cataloging: A Guide to Understanding the Fundamental Changes in Cataloging," *Arkansas Libraries* 67, no. 2 (Summer 2010): 14–19; Ted Gemberling, "Thema and FRBR's Third Group," *Cataloging & Classification Quarterly* 48, no. 5 (2010): 445–49.
 20. Jason Thomale, "Interpreting MARC: Where's the Bibliographic Data?" *Code4Lib Journal* 11 (2010), http://journal.code4lib.org/issues/issue11 (accessed Oct. 19, 2011).
 21. Karen Smith-Yoshimura et al., *Implications of MARC Tag Usage on Library Metadata Practices* (Dublin, Ohio: OCLC, 2010), www.oclc.org/research/publications/library/2010/2010-06.pdf (accessed Mar. 14, 2011).
 22. Ibid., 27.
 23. Amy P. Eklund et al., "Comparison of MARC Content Designation Utilization in OCLC WorldCat Records with National, Core, and Minimal Level Record Standards," *Journal of Library Metadata* 9, no. 1/2 (2009): 36–64.
 24. Matthew Mayernick, "The Distributions of MARC Fields in Bibliographic Records: A Power Law Analysis," *Library Resources & Technical Services* 54, no. 1 (2010): 40–54.
 25. Carol Jean Godby, *Mapping ONIX to MARC* (Dublin, Ohio: OCLC Research, Apr. 2010), www.oclc.org/research/publications/library/2010/2010-14.pdf; ONIX-MARC Mapping (Crosswalk), a spreadsheet workbook containing thirteen tables related to *Mapping ONIX to MARC* is available at www.oclc.org/research/publications/library/2010/2010-14a.xls (accessed Mar. 14, 2011).
 26. Ibid., 10.
 27. Ibid., 16.
 28. Mayernick, "The Distributions of MARC Fields in Bibliographic Records: A Power Law Analysis," 49; *Anglo-American Cataloguing Rules*, 2nd ed., 2002 rev., 2005 update (Chicago: American Library Association; Ottawa: Canadian Library Association; London: Chartered Institute of Library and Information Professionals, 2005).
 29. Steven A. Knowlton, "How the Current Draft of RDA Addresses the Cataloging of Reproductions, Facsimiles, and Microforms," *Library Resources & Technical Services* 53, no. 3 (2009): 159–65.
 30. Mayernick, "The Distributions of MARC Fields in Bibliographic Records"; Knowlton, "How the Current Draft of RDA Addresses the Cataloging of Reproductions, Facsimiles, and Microforms," 163–64.
 31. Diane Hillmann et al., "RDA Vocabularies: Process, Outcome, Use," *D-Lib Magazine* 16, no. 1/2 (Jan./Feb. 2010) www.dlib.org/dlib/january10/hillmann/01hillmann.html (accessed Aug. 15, 2011).
 32. Ibid.
 33. Gordon Dunsire, "The Semantic Web and Expert Metadata: Pull Apart Then Bring Together," (presentation, Archives, Libraries, Museums 12 (AKM12), Poreč, Croatia, Nov. 26–28, 2008), eprints.rclis.org/bitstream/10760/7443/1/akm2008semanticweb.pdf (accessed Mar. 18, 2011).
 34. Ibid., 1.
 35. Karen Coyle, "Library Data in a Modern Context," in *Understanding the Semantic Web: Bibliographic Data and Metadata*, *Library Technology Reports* 46, no. 1 (Jan. 2010): 5–13; Karen Coyle, "Changing the Nature of Library Data," in *Understanding the Semantic Web: Bibliographic Data and Metadata*, *Library Technology Reports* 46, no. 1 (Jan. 2010): 14–29; Karen Coyle, "Library Data in the Web world," in *RDA Vocabularies for a Twenty-First Century Data Environment*, *Library Technology Reports* 46, no. 2 (Feb./Mar. 2010): 5–11; Karen Coyle, "Metadata Models of the World Wide Web," in *RDA Vocabularies for a Twenty-First Century Data Environment*, *Library Technology Reports* 46, no. 2 (Feb./Mar. 2010) 46(2): 12–19, 36; Karen Coyle, "RDA in RDF," in *RDA Vocabularies for a Twenty-First Century Data Environment*, *Library Technology Reports* 46, no. 2 (Feb./Mar. 2010): 26–36.
 36. Coyle, "Changing the Nature of Library Data," 11.
 37. Fiona Bradley, "Discovering Linked Data," *Library Journal*

- 134, no. 7 (Apr. 15, 2009): 48–50.
38. Martha M. Yee, "Can Bibliographic Data Be Put Directly onto the Semantic Web?" *Information Technology & Libraries* 28, no. 2 (June 2009): 55–80.
 39. *Ibid.*, 66.
 40. Ilana Tolkoff, "The Path toward Global Interoperability in Cataloguing," *Information Technology & Libraries* 29, no. 1 (Mar. 2010): 30–39.
 41. Jenn Riley (content author) and Devin Becker (design), Seeing Standards: A Visualization of the Metadata Universe (2009–2010), www.lib.unc.edu/users/jlriley/metadata-map/seeingstandards.pdf (accessed Oct. 19, 2011). This publication is accompanied by *Glossary of Metadata Standards*, www.lib.unc.edu/users/jlriley/metadata-map/seeingstandards_glossary_pamphlet.pdf (accessed Mar. 14, 2011).
 42. Sharon Yang, Yanyi Lee, and Amanda Xu, "The Semantic Web and Libraries in the United States: Experimentation and Achievements," in *Emerging Trends in Technology: Libraries between Web 2.0, Semantic Web and Search Technology: Proceedings. IFLA 2009 Milan-Italy; Satellite meetings in Florence* (CD-ROM ed.) (Florence, Italy: IFLA, 2009).
 43. Jung-ran Park and Yuji Tosaka, "Metadata Creation Practices in Digital Repositories and Collections: Schemata, Selection Criteria, and Interoperability," *Information Technology & Libraries* 29, no. 3 (Sept. 2010): 104–16.
 44. Jin Ma, "Metadata in ARL libraries: A Survey of Metadata Practices," *Journal of Library Metadata* 9, no. 1/2 (Jan. 2009): 1–14; Vicki Toy-Smith, "UALC Best Practices Metadata Guidelines: A Consortial Approach," *Journal of Library Metadata* 10, no. 1 (2010): 1–12; Marielle Vevé and Melanie Feltner-Reichert, "Integrating Non-MARC Metadata Duties into the Workflow of Traditional Catalogers: A Survey of Trends and Perceptions among Catalogers in Four Discussion Lists," *Technical Services Quarterly* 27, no. 2 (2010): 194–213.
 45. Leighann Ayers et al., *What We've Learned from the RLC Partners Metadata Creation Workflows Survey* (Dublin, Ohio: OCLC, 2009), www.oclc.org/programs/publications/reports/2009-04.pdf (accessed Mar. 30, 2011), 6, 7.
 46. Maura L. Valentino, "Integrating Creation into Catalog Workflow," *Cataloging & Classification Quarterly* 48, no. 6/7 (2010): 541–50.
 47. Lynne C. Howarth, Leslie Moore, and Eliza Sze, "Mountains to Molehills: The Past, Present and Future of Cataloging Backlogs," *Cataloging & Classification Quarterly* 48, no. 5 (2010): 423–44.
 48. As cited in Howarth, Moore and Sze: Sarah E. Thomas, "The Core Bibliographic Record and the Program for Cooperative Cataloging," *Cataloging & Classification Quarterly* (1996) 21(3/4): 91–108.
 49. Judy Luther, *Streamlining Book Metadata Workflow* (Ardmore, Pa.: Informed Strategies, June 30, 2009), www.niso.org/publications/white_papers/StreamlineBookMetadataWorkflowWhitePaper.pdf (accessed Mar. 21, 2011).
 50. Luther, *Streamlining Book Metadata Workflow*, 13.
 51. Lihong Zhu, "The Role of the Cataloging Department in the Implementation of OCLC WorldCat Local," *Library Collections & Technical Services* 34, no. 4 (2010): 123–29.
 52. Dong Suk Kim, "Using the Balanced Scorecard for Strategic Operation of the Cataloging Department," *Cataloging & Classification Quarterly* 48, no. 6/7 (2010): 572–84.
 53. *Ibid.*, 575, 576.
 54. *Ibid.*
 55. *Ibid.*, 579–580, 581.
 56. Joan E. Schuitema, "The Future of Cooperative Cataloging: Curve, Fork, or Impasse?" *Cataloging & Classification Quarterly* 48, no. 2/3 (2010): 258–70; Roxanne Sellberg, "Cooperative Cataloging in a Post-OPAC World," *Cataloging & Classification Quarterly* 48, no. 2/3 (2010): 237–46.
 57. Sellberg, "Cooperative Cataloging in a Post-OPAC World," 244.
 58. David Banush, "Cooperative Cataloging at the Intersection of Tradition and Transformation: Possible Futures for the Program for Cooperative Cataloging," *Cataloging & Classification Quarterly* 48, no. 2/3 (2010): 247–57; Mechaël D. Charbonneau, "Program for Cooperative Cataloging: The Indiana Experience," *Cataloging & Classification Quarterly* 48, no. 2/3 (2010): 113–25; Magda El-Sherbini, "Program for Cooperative Cataloging: Analysis of Quality," *Cataloging & Classification Quarterly* 48, no. 2/3 (2010): 221–36.
 59. Banush, "Cooperative Cataloging at the Intersection of Tradition and Transformation," 256.
 60. Louise F. Spiteri, "The Impact of Social Cataloging Sites on the Construction of Bibliographic Records in the Public Library Catalog," *Cataloging & Classification Quarterly* 47, no. 1 (2009): 52–73.
 61. *Ibid.*, 52.
 62. Ruth Fischer and Rick Lugg, *Library of Congress Study of the North American MARC Records Marketplace* (Contoocook, N.H.: R2 Consulting, Oct. 2009), www.loc.gov/bibliographic-future/news/MARC_Record_Marketplace_2009-10.pdf (accessed Mar. 16, 2011).
 63. *Ibid.*, 23.
 64. *Ibid.*, 29.
 65. *Ibid.*, 36.
 66. Robert Wolven, "Cataloging without Silos, or, Where Do We Go from Here?" *Technicalities* 29, no. 5 (Sept./Oct. 2009): 1, 6–8.
 67. *Ibid.*, 7.
 68. *Ibid.*, 8.
 69. Martha Hruska, "Where Are We with the Staffing Transition from Cataloging to Metadata Management?" *Technicalities* 29, no. 2 (Mar./Apr. 2009): 1, 14–16.
 70. *Ibid.*, 14.
 71. *Ibid.*, 15.
 72. Anne M. Mitchell, Michael Thompson, and Annie Wu, "Agile Cataloging: Staffing and Skills for a Bibliographic Future," *Cataloging & Classification Quarterly* 48, no. 6/7 (2010): 506–24.
 73. Jennifer J. Leffler and Pamela Newberg, "Re-Visioning Technical Services: A Unique Opportunity to Examine the Past, Access the Present, and Create a Better Future," *Cataloging & Classification Quarterly* 48, no. 6/7 (2010): 561–71.
 74. Myung-Ja Han and Patricia Hswe, "The Evolving Role of the Metadata Librarian: Competencies Found in Job Descriptions," *Library Resources & Technical Services* 54, no. 3 (2010): 129–41.

75. Jung-ran Park and Caimei Lu, "Metadata Professionals: Roles and Competencies as Reflected in Job Announcements, 2003-2006," *Cataloging & Classification Quarterly* 47, no. 2 (2009): 145-60; Jung-ran Park, Caimei Lu, and Linda Marion, "Cataloging Professionals in the Digital Environment: A Content Analysis of Job Descriptions," *Journal of the American Society for Information Science and Technology* 60, no. 4 (Apr. 2009): 844-57.
76. Park, Lu and Marion, "Cataloging Professionals in the Digital Environment," 848, 854.
77. Alison Hitchens and Ellen Symons, "Preparing Catalogers for RDA Training," *Cataloging & Classification Quarterly* 47, no. 8 (2009): 691-707.
78. Elizabeth J. Cox and Ann K. D. Myers, "What Is a Professional Cataloger? Perception Differences between Professionals and Paraprofessionals," *Library Resources & Technical Services* 54, no. 4 (2010): 212-26.
79. Richard Sapon-White, "Subject Analysis Training for Cataloging Paraprofessionals," *Technical Services Quarterly* 26, no. 3 (2009): 183-93.
80. Colleen Valente, "Training Successful Paraprofessional Copy Catalogers," *Library Resources & Technical Services* 53, no. 4 (2009): 219-30.
81. Michèle Hudon, "Teaching Classification, 1990-2010," *Cataloging & Classification Quarterly* 48, no. 1 (2010): 64-82.
82. Lisius et al., "Integrating Enhance and NACO Work into Pre-Professional Experiences," *Cataloging & Classification Quarterly* 48, no. 6/7 (2010) 610-33.
83. Julianne Beall and Joan S. Mitchell, "History of the Representation of the DDC in the MARC Classification Format," *Cataloging & Classification Quarterly* 48, no. 1 (2010): 48.
84. Ia C. McIlwaine, "Universal Bibliographic Control and the Quest for a Universally Acceptable Subject Arrangement," *Cataloging & Classification Quarterly* 48, no. 1 (2010): 36.
85. *Ibid.*, 36, 37.
86. Aida Slavic, "Call Numbers, Book Numbers & Collection Arrangements in European Library Traditions," *Library and Information Science in the Digital Age: Essays in Honour of M.P. Satija*, ed. Jagtar Singh, Indervir Malhan, and Trishanjit Kaur, 257-85 (New Delhi: Ess Ess, 2009).
87. Martha M. Yee, "'Wholly Visionary': The American Library Association, the Library of Congress, and the Card Distribution Program," *Library Resources & Technical Services* 53, no. 2 (2009): 68-78.
88. *Ibid.*, 75.
89. Steven A. Knowlton, "Criticism of Cataloging Code Reform, As Seen in the Pages of *Library Resources and Technical Services* (1957-66)," *Library Resources & Technical Services* 53, no. 1 (2009): 15-24.
90. *Ibid.*, 21.
91. Nancy E. Elkington, "OCLC Research: Past, Present, and Future," *Journal of Library Administration* 49, no. 7 (2009): 707-18.
92. *Ibid.*, 709.
93. Joan S. Mitchell and Diane Vizine-Goetz, "The DDC and OCLC," *Journal of Library Administration* 49, no. 6 (2009): 657-67.
94. Michael Panzer and Marcia Lei Zeng, "Modeling Classification Systems in SKOS: Some Challenges and Best-Practice Recommendations," *International Conference on Dublin Core and Metadata Applications CD-2009—Seoul Proceedings*, Seoul, Korea, Oct. 12-16, 2009, <http://dcpapers.dublincore.org/ojs/pubs/article/view/974> (accessed June 14, 2011).
95. Jun Wang, "An Extensive Study on Automated Dewey Decimal Classification," *Journal of the American Society of Information Science and Technology* 60, no. 11 (Nov. 2009): 2, 269-86.
96. Claudio Gnoli, "Classification Transcends Library Business," *Knowledge Organization* 37, no. 3 (2010): 223-29.
97. *Ibid.*, 223.
98. Hur-Li Lee, "Divination and the State: Classifying Technical Texts in Han China," *Library Resources & Technical Services* 54, no. 4 (2010): 200-211.
99. *Ibid.*, 200.
100. Haroon Idrees and Khalid Mahmood, "Devising a Classification Scheme for Islam: Opinions of LIS and Islamic Studies Scholars," *Library Philosophy & Practice* (Oct. 2009), <http://unllib.unl.edu/LPP/idrees-mahmood.htm> (accessed Oct. 25, 2011).
101. Barbara Fister, "The Dewey Dilemma: In the Search for Better Possibility, Librarians Are Putting Dewey in a Different Class," 134, no. 16 *Library Journal* (Oct. 1, 2009): 22-25.
102. Nicole Lyons, "The New Darien Library: It's for ME!" *All about Darien: The Insider Guide* [blog], Jan. 7, 2009, <http://all-aboutdarien.com/2009/01/the-new-darien-library-its-for-me/> (accessed June 8, 2011).
103. Stephen Hearn, "Comparing Catalogs: Currency and Consistency of Controlled Headings," *Library Resources & Technical Services* 53, no. 1 (2009): 25-40.
104. Susan K. Burke and Jay Shorten, "Name Authority Work Today: A Comparison of Types of Academic Libraries," *Library Resources & Technical Services* 54, no. 1 (2010): 4-20.
105. Michael Kreyche, Peter H. Lisius, and Amey Park, "The DeathFlip Project: Automating Death Date Revisions to Name Headings in Bibliographic Records," *Cataloging & Classification Quarterly* 48, no. 8 (2010): 684-695.
106. Patricia M. Dragon, "Name Authority Control in Local Digitization Projects and the Eastern North Carolina Postcard Collection," *Library Resources & Technical Services* 53, no. 3 (2009): 185-96.
107. *Ibid.*, 189.
108. EunKyung Chung, Shawne Miksa and Samantha K. Hastings, "A Framework of Automatic Subject Term Assignment for Text Categorization: An Indexing Conception-Based Approach," *Journal of the American Society of Information Science and Technology* 61, no. 4 (2010): 688-99.
109. Kwan Yi and Lois Mai Chan, "Revisiting the Syntactical and Structural Analysis of Library of Congress Subject Headings for the Digital Environment," *Journal of the American Society for Information Science and Technology* 61, no. 4 (Apr. 2010): 677-87.
110. Richard E. Sapon-White, "Series Authority Control at Oregon State University after the Library of Congress's Series Policy Change," *Library Resources & Technical Services* 53, no. 2 (2009): 79-85.
111. Cecile E. Bianco, "Medical Librarians' Uses and Perceptions

- of Social Tagging,” *Journal of the Medical Library Association* 97, no. 2 (Apr. 2009): 136–39.
112. Lauren A. Maggio et al., “A Case Study: Using Social tagging to Engage Students in Learning Medical Subject Headings,” *Journal of the Medical Library Association* 97, no. 2 (Apr. 2009): 77–83.
 113. *Ibid.*, 78.
 114. Patrick Griffis and Cyrus Ford, “Enhancing OPAC Records for Discovery,” *Information Technology & Libraries* 28, no. 4 (Dec. 2009): 191–93.
 115. C. Rockelle Strader, “Author-Assigned Keywords versus Library of Congress Subject Headings: Implications for the Cataloging of Electronic Theses and Dissertations,” *Library Resources & Technical Services* 53, no. 4 (2009): 243.
 116. Melissa Adler, “Transcending Library Catalogs: A Comparative Study of Controlled Terms in Library of Congress Subject Headings and User-Generated Tags in LibraryThing for Transgender Books,” *Journal of Web Librarianship* 3, no. 4 (2009): 309–31.
 117. *Ibid.*, 309.
 118. *Ibid.*
 119. Marliese Thomas, Dana M. Caudle, and Cecilia Schmitz, “Trashy Tags: Problematic Tags in LibraryThing,” *New Library World* 111, no. 5/6 (2010): 223–35.
 120. *Ibid.*, 223, 226.
 121. Atefeh Sharif, “Combining Ontology and Folksonomy: An Integrated Approach to Knowledge Representation,” Presented at IFLA 2009 Satellite Meeting in Florence, Italy, Aug. 19–20, 2009, www.eprints.rclis.org/bitstream/1070/15058/2/AtefehSharif.pdf (accessed Feb. 15, 2012).
 122. Constantia Kakali and Christos Papatheodorou, “Could Social Tags Enrich the Library Subject Index?” *Proceedings of the International Conference Libraries in the Digital Age (LIDA)*, held in Zadar, Croatia, May 24–28, 2010, eprints.rclis.org/bitstream/10760/14775/1/could_tags_Lida2010v2.3.pdf (accessed Mar. 21, 2011), 2.
 123. Karen G. Lawson, “Mining Social Tagging Data for Enhanced Subject Access for Readers and Researchers,” *Journal of Academic Librarianship* 35, no. 6 (2009): 580.
 124. Peter J. Rolla, “User Tags versus Subject Headings: Can User-Supplied Data Improve Subject Access to Library Collections?” *Library Resources & Technical Services* 53, no. 3 (2009): 174.
 125. Michael Boock and Sue Kunda, “Electronic Thesis and Dissertation Metadata Workflow at Oregon State University Libraries,” *Cataloging & Classification Quarterly* 47, no. 3/4 (2009): 297–308.
 126. Annie Wu and Anne M. Mitchell, “Mass Management of E-Book Catalog Records: Approaches, Challenges, and Solutions,” *Library Resources & Technical Services* 54, no. 3 (2010): 164–74.
 127. Terry Reese, “Automated Metadata Harvesting: Low-Barrier MARC Record Generation from OAI-PMH Repository Stores Using MarcEdit,” *Library Resources & Technical Services* 53, no. 2 (2009): 121–34.
 128. *Ibid.*, 122.
 129. Ashley Beamer, “Map Metadata: Essential Elements,” *Program: Electronic Library and Information Systems* 43, no. 1 (2009): 18.
 130. Kimberly C. Kowal and Christophe Martyn, “Descriptive Metadata for Digitization of Maps in Books: A British Library Project,” *Library Resources & Technical Services* 53, no. 2 (2009): 108–20.
 131. Anila Angjeli et al., “Semantic Web and Vocabulary Interoperability: An Experiment with Illumination Collections,” *International Cataloguing and Bibliographic Control* 38, no. 2 (Apr./June 2009): 25.
 132. *Ibid.*, 25.
 133. Allyson Carlyle, “Invited Editorial: Announcing 2010, Year of Cataloging Research,” *Cataloging & Classification Quarterly* 47, no. 8 (2009): 687–90.

How Much Do Core Journals Change over a Decade?

The Case of Communication Disorders

Steve Black

This study replicates the method of identifying core journals in the field of communication disorders published in the January 2001 issue of Library Resources and Technical Services for the purpose of determining the degree to which the ranked list changed after ten years. Two measures are used to assess the reliability of rankings over time: Spearman's rho rank correlations among the citing journals and coefficients of variation among cited journals. Rank correlations of groups of journals can mask important changes in rank for individual titles, so characteristics of the journals with the greatest movements in rank over a decade are explored. Major findings are that the discipline's literature grew substantially over the decade, and the core journals remained stable over ten years ($r_s = 0.73$). However, despite stability of core journals over time, some titles changed dramatically in rank. Coefficients of reliability calculated for this group of communication disorders journals suggests that approximately one-third of observed change in ranks is because of random variability in works cited.

The January 2001 issue of *Library Resources and Technical Services (LRTS)* published an article by this author on core journals in the field of communication disorders.¹ The purpose of the present study is to determine the degree to which the ranked list of frequently cited journals in communication disorders changed after ten years. Because serials collection development relies on choosing the most relevant journals among many, keeping track of changes in core journals over time is vital. An important related question is whether changes in rank observed over time were greater than variation in ranks between titles covering the same years. This is important because reliability (consistency of measurements) indicates whether changes in rank are indicative of journals' true rise or fall relative to others or the result of random movement caused by measuring different samples. Two measures are used to assess the reliability of rankings over time: rank correlations among the citing journals and coefficients of variation among cited journals. Correlations of journal rankings provide a useful measure of the reliability of core lists over time, but looking at journals as a group can mask important changes in the ranks of individual titles. Therefore this study includes a look at the characteristics of journals that experienced the greatest movement up or down the ranked list of most frequently cited journals in the field of communication disorders.

Steve Black (blacks@strose.edu) is Serials and Reference Librarian, Neil Hellman Library, The College of Saint Rose, Albany, New York.

Submitted December 6, 2011; tentatively accepted, pending modest revision December 15, 2011; revisions submitted December 22, 2011, and accepted for publication.

The author thanks Stan Lyle at the Rod Library, University of Northern Iowa, for sending the article count for the 1997 volume of *Journal of Medical Speech-Language Pathology*.

Literature Review

In her conclusion to a 2003 overview of journals in communication disorders, Shpilko recommended that future researchers use citation analysis to follow trends in the literature because core journals are likely to change over time.² Various criteria (including lists compiled by subject experts, extent of coverage by indexes, library holdings, surveys of faculty, and citation data) for selecting the best journals to support research within a discipline have been used by collection development librarians.³ Citation analysis is the study of relationships between authors, articles, journals, concepts, institutions, nations, and other elements by compiling and counting references in documents.⁴ Broadly stated, citation analysis is the study of relationships among published works based on the principle that a reference in a published paper implies a relationship between the cited and the citing documents.⁵ One type of citation analysis is to count citations to identify leading journals in a discipline. The present study employed the method of counting citations to identify the most frequently cited journals in a discipline, in this case communication disorders.

Using Published Impact Factors to Rank Journals

Garfield's pioneering work, comparing citation counts to objectively evaluate the relative quality of journals on the basis of how frequently they are referenced, led to the creation of the Institute for Scientific Information's (ISI) citation indexes.⁶ The citation indexes are now incorporated within Thomson Reuters' *Web of Knowledge*, which also includes *Journal Citation Reports*, a ranking of journals by topic area based on impact factors. A journal's impact factor is calculated "by dividing the number of current year citations to the source items published in that journal during the previous two years."⁷ The journal impact factors published by Thomson Reuters have become standard, albeit controversial, metrics for the measurement of importance or quality of journals, authors, and institutions. A voluminous literature has critiqued the methods of formulating impact factors and the (mis)applications of them in various contexts. A small but representative sample of these critiques are reviewed here to highlight primary concerns surrounding the validity and reliability of impact factors as calculated in the *Web of Knowledge*.

One line of criticism points out logical and practical weaknesses in the very concept of using impact factors to judge journals, scholars, or institutions. Martin emphasized the need for a conceptual distinction between the quality, importance, and impact of publications and cautioned that citation counts are only useful as a measure of impact.⁸ He wrote in the context of assessing the output of research

centers, but his logic applies equally well to journals. In his view, *quality* is how well the research is conducted and written, *importance* is the potential influence of the work, and *impact* is the actual influence. Importance and impact are closely related, but it takes time for importance to be reflected in citations, and for reasons such as language, prominence of the journal in which it was published, or narrowness of subspecialty, an important paper may receive few citations. Simply put, a high-quality paper may have the misfortune of never receiving much attention.

Rosenstreich and Wooliscroft wrote a particularly good summary of critiques of citation-based journal rankings.⁹ The critiques include arguments that frequency of citation is not equal to quality because a "halo effect" raises the profile of a journal containing a highly cited paper, and a "Matthew Effect" causes highly cited papers to continue to be more visible and thus more highly cited. The halo effect occurs when a highly cited article draws attention to other articles in the same issue of a journal, a phenomenon that is becoming less prevalent as more articles are viewed online in isolation from others published within an issue.¹⁰ The Matthew Effect was coined by Merton after the passage in Matthew (13:12), "for unto every one that hath shall be given, and he shall have abundance: but from him that hath not shall be taken away even that which he hath."¹¹ The more attention something receives, the more attention it will continue to receive. The presumption is that this Matthew Effect skews citation rates because the most highly cited papers or journals end up receiving more citations than they deserve based solely on their quality and importance. An issue related to such skewed citation is Seglen's point that little correlation exists between the quality of a paper and the quality of the journal that published it. He argues that because excellent papers appear in lesser-known journals, those judging a researcher's work should not consider the impact factors of the journals in which they have published.¹²

Much of the literature critiquing journal rankings specifically addressed perceived shortcomings in the data published by the Institute for Scientific Information (ISI). MacRoberts and MacRoberts described several problems with ISI's method of ranking journals in the *Journal Citation Reports*: influences not being cited, biased citing, inappropriate self-citation, no distinctions between negative and positive references, variation in citation patterns among disciplines, miscounts due to misspellings, multiple authors with the same name, and incomplete coverage of the literature.¹³ One should note that Thomson Reuters is aware of such critiques and works to minimize those errors they are able to control.¹⁴ A criticism of the Thomson Reuters citation data unrelated to the aforementioned issues is the high cost. Purdue and Piotrowski note that many researchers have budgets that are insufficient to gain access to the data.¹⁵

Ranked Lists Created Independently from Published Impact Factors

Since counting citations is an objective way to measure the impact of journals, but citation data published by Thomson Reuters have drawbacks, researchers have created rankings of journals using other sources of citation data. Studies based on data from sources other than the *Web of Knowledge* are particularly useful for measuring the impact of journals in multidisciplinary fields. This is because the field may not be a category in Thomson Reuters' *Journal Citation Reports (JCR)* and data on journals covered by *JCR* reflect citations from all disciplines. For example, the published impact factor for *Library Resources and Technical Services* makes no distinction among citations from papers in library science, management, or any other discipline. That may be of little concern in well defined and established disciplines, but measuring only citations from within the discipline is essential for determining the most frequently cited journals in new, multidisciplinary fields. Without a focused analysis, one cannot know the degree to which the published impact factor correlates with the real impact of a journal within a specific field such as communication disorders.

One approach for an independent study is to use citation index data but not rely on the *JCR*. In 1991 when the citation indexes were available as dial-up databases, McCain employed the Cited Work field in the online version of *Science Citation Index (SCISEARCH)* to identify core journals in genetics.¹⁶ Other researchers have chosen to count citations independently rather than rely on citation indexes. Independent studies may be global (measure citations published anywhere) or local (measure citations of publications by researchers from one institution, state, or nation). Because the present paper is based on citations published anywhere, only global citation analysis studies will be noted here. Kushkowsky, Gerhard, and Dobson advocated using a Simple Index Method to aid collection development librarians in their decision making.¹⁷ The Simple Index Method generates ranked lists of core journals in a topic area by counting citations to journals in indexes such as *Econlit*, *ABI/Inform*, or *PsycINFO*. Mack created a list of core women's studies journals by counting citations in *Signs*, a journal known to be prominent in that discipline, and concluded that citation analysis was superior to surveying faculty for identifying core journals.¹⁸ Sittig culled citation data from the *Medline* bibliographic database to identify top journals in the interdisciplinary field of medical informatics.¹⁹ Goss counted citations from textbooks and journals to rank top journals in optometry and ophthalmology.²⁰ Wray began with a set of five source journals to identify core journals in special education.²¹ Slater created a ranked list of journals in speech-language pathology by counting citations in the 1991–93 issues of *American Journal of Speech-Language*

*Pathology, Journal of Speech and Hearing Research, European Journal of Disorders of Communication, and Journal of Communication Disorders.*²² Slater's study was similar enough to the present work to draw comparisons and calculate correlations, so it will be referred to again below.

Despite these examples of independent work, the preponderance of journal rankings has been based on data published by Thomson Reuters or Elsevier. Thomson Reuters' *Web of Knowledge* is the best-known source, but impact factors also are available in Elsevier's *Scopus*, and citation analysis of ever-increasing sophistication can be accomplished with Google Scholar. In fact, Harzing and van der Wal argue that Google Scholar is a superior data source because it has the most comprehensive coverage of journals.²³

Reliability of Journal Rankings

Perhaps because of the widespread use of the *Web of Knowledge*, surprisingly little has been written about the reliability of journal rankings that are based on citation counts. Variation exists over time and because of sample selection, but very few published studies have addressed the degree of variation. Variability in what is cited has been recognized for a long time. Bradford's seminal 1934 article on a discipline's concentration of citations in a core of journals includes a discussion of the marked variability of journals cited from journals outside the core.²⁴ He noted that the number of journals cited at least once grows linearly with the period of years being surveyed. The fact that more journals are cited as sample size increases is evidence for a fundamental variability in what is being cited. Nieuwenhuisen and Rousseau modeled which elements of fluctuation in citations were due to random variation rather than a genuine rise or decline in citations received by a journal.²⁵ Their model was based on the coefficient of variation of total citations received and the journals' impact factors. They found that the rank of a journal high on a list can be expected to vary little over time, but that the rank of a journal lower on a list will fluctuate more in rank. This is because at lower citation rates, a relatively small change in times cited causes a relatively large change in rank. Line noted that, over time, the reliability of a journal's rank grows weaker as one moves down a list, stating "it is a matter of chance whether a little-used or little-cited title receives, in any one year, zero, one, two, three, four, or five uses or citations, although the rank order may be dramatically affected."²⁶ Therefore he advocated that studies be conducted to measure the reliability of citation counts at the fringes of use. Line's point of view was based on his research that found limited commonality in journal rankings from the *JCR* with lending patterns at the British Library Lending Division in 1975, 1980, and 1983. Overlap in titles within the *JCR* lists over time was in the range of 80 percent to 95 percent, but overlap within

serials requested by the British Library over time was in the range of 54 percent to 62 percent. In another article, Line speculated that the dramatic difference stemmed from less variation between serials used by researchers than variability between serials used by the borrowing public at large.²⁷ Broadus acknowledged the difficulty in making judgments for low-use titles but maintained that, despite issues of reliability, *JCR* rankings were a cost-effective starting point for deciding which journal subscriptions to cancel.²⁸

Global versus Local Use

A fundamental issue raised by Broadus' argument for using the *JCR* for collection development decision was the correlation of global use (citations in the published works measured by the *JCR*) versus local use by library patrons.²⁹ Simply put, can *JCR* data reliably predict what a library's patrons will need? For an overview of studies that looked at the correlation of global with local use, see Kelland and Young.³⁰ They note that citation is a useful evaluative tool even though faculty refer to journals beyond their library's collection and some uses of journals in the collection may go uncounted. A few studies of global versus local use have used rank correlation as the method for measuring the relationship. White and White correlated psychology journal rankings in the *Social Science Citation Index* with rankings by subjective evaluation by psychologists, finding Spearman rank correlations of $r_s = 0.39$ and $r_s = 0.56$.³¹ Ralston, Gall, and Brahmī correlated ranked lists of psychiatry journals to determine how reliably *JCR* data matched works cited in publications by faculty of the Indiana University School of Medicine and found a strong positive rank correlation of $r_s = 0.79$, concluding that the global *JCR* data effectively captured local psychiatry research emphases.³² Kreider compared works cited by faculty publications at the University of British Columbia with the *JCR* and found strong positive correlations for most disciplines.³³

Statistical Tests for Ranked Lists

Kreider performed logarithmic transformations on the data and then correlated results using Pearson's product-moment coefficients based on the advice of Bensman that such an approach was a superior way to analyze citation data.³⁴ He argued correctly that Pearson's product-moment coefficients are based on parameters that include a normal distribution. But citation data are virtually always highly skewed and thus better modeled by a negative binomial distribution, e.g., Bradford's Law of Scattering.³⁵ Other solutions besides logarithmic transformations address this problem. Lack of fit to a standard distribution is the reason nonparametric tests were developed that are not based on assumptions about the population being measured. Such nonparametric statistics

include the well-known and widely used Spearman rank correlation test of the degree of association between two sets of data listed in rank orders.³⁶ The nonparametric Spearman's rho (denoted r_s) is an appropriate measure for correlating ranked lists because it is not based on assumptions (parameters) regarding normal distribution or standard variation of numerical data.³⁷ The Merriam-Webster dictionary defines reliability as "the extent to which an experiment, test, or measuring procedure yields the same results on repeated trials."³⁸ Spearman's rho is the appropriate tool for determining the reliability of ranked lists because it measures the degree of association between two sets of data, whether the sets are from different samples taken within the same time period or from comparable samples taken at different times.

Method

Half of the data for this study come from the author's 2001 *LRTS* article on a core collection in communication disorders. To determine the degree to which the core list changed over a decade, works cited were compiled for the appropriate volumes of the two journals used in the initial study: *Journal of Speech, Language, and Hearing Research (JSLHR)* and *Journal of Communication Disorders (JCD)*. The broader sample of four journals used by Slater may better represent the literature, but the sample corresponding to this author's first study was used to minimize extraneous variables.³⁹ The volumes, number of articles published, number of works cited, and average number of works cited per article are shown in table 1.

For the 2001 study the author entered the data by hand from printed issues. Thankfully, new tools are now available to speed the process of compiling works cited. The journals used for this study are covered by the *PsycINFO* bibliographic database. *PsycINFO* has the very useful feature of including the works cited of every article whether the full text is available or not. The author exported the works cited from every issue of the sampled journals from *PsycINFO* (via EBSCOhost) into the Refworks bibliographic management program. A drawback of downloading works cited instead of entering them by hand is that *PsycINFO* tags works from any type of series as "journal article." The author was thus unable to re-create the proportion of citations from journals, conferences, tests, books, and others that appeared in figure 1 of the 2001 study.⁴⁰ Databases hosted by EBSCO can export directly into several bibliographic management programs, any of which could be used for compiling citation data. RefWorks was chosen as the most cost-effective choice. Exported citations were organized into one folder for each volume of each journal, and each folder was downloaded into a Microsoft Excel worksheet. Keeping data from each volume discrete allowed the analysis that appears in table 2.

Table 1. Journals in Communication Disorders and Growth in Scholarly Literature

Journals Used for This Study	Articles Published	Works Cited	Average Works Cited per Article
<i>Journal of Communication Disorders</i> 1997–99	76	2,756	36
<i>Journal of Speech, Language, and Hearing Research</i> 1997	103	4,336	42
<i>Journal of Speech, Language, and Hearing Research</i> 1999	103	4,714	46
<i>Journal of Communication Disorders</i> 2007–9	102	4,785	47
<i>Journal of Speech, Language, and Hearing Research</i> 2007	109	5,407	50
<i>Journal of Speech, Language, and Hearing Research</i> 2009	117	5,439	47

Table 2. Spearman's rho (r_s) Rank Correlation Coefficients among the Six Volumes

	<i>Journal of Communication Disorders</i> 1997–99	<i>Journal of Speech, Language, and Hearing Research</i> 1997	<i>Journal of Speech, Language, and Hearing Research</i> 1999	<i>Journal of Communication Disorders</i> 2007–9	<i>Journal of Speech, Language, and Hearing Research</i> 2007	<i>Journal of Speech, Language, and Hearing Research</i> 2009
<i>Journal of Communication Disorders</i> 1997–99	1					
<i>Journal of Speech, Language, and Hearing Research</i> 1997	0.52	1				
<i>Journal of Speech, Language, and Hearing Research</i> 1999	0.61	0.71	1			
<i>Journal of Communication Disorders</i> 2007–9	0.67	0.65	0.69	1		
<i>Journal of Speech, Language, and Hearing Research</i> 2007	0.56	0.60	0.71	0.66	1	
<i>Journal of Speech, Language, and Hearing Research</i> 2009	0.61	0.64	0.68	0.67	0.70	1

The author used Excel to compile and analyze the works cited data from the six journals. A time-consuming but important task was fixing discrepancies in titles. For example, title variants *Audiol Neurootol*, *Audiol Neurotol*, *Audiology & Neuro-Otology*, *Audiology & Neurotology*, and *Audiology and Neuro-Otology* were edited to match the title as it appears in OCLC WorldCat, *Audiology & Neuro-otology*. Even with careful editing, the author deemed alphabetizing lists and counting citations to be better rather than relying on Excel's COUNTIF function, because any overlooked discrepancy in punctuation or spelling could cause a citation to go uncounted. Citations to books, standards, tests, chapters in series, and anything other than citations to journals were ignored. Title changes

posed something of a dilemma. Does the new title count as a new work to be treated separately or as a continuation of the same work? The author chose to aggregate counts for journals that changed title, but the author treated splits and mergers as separate works.

Counts for journals cited at least 5 times were copied to a separate worksheet for each citing journal. Worksheets were created to aggregate results by decade, one for 1997–99 and one for 2007–9. All lists were sorted by citation count to create ranked lists of cited journals. The core list published in 2001 included 104 journals that represented 80 percent of citations to journals. This base was used to establish 104 as the maximum number of journals in the ranked lists used here.

Table 3. Variability in Times Cited

Journals cited in <i>Journal of Speech, Language, and Hearing Research</i>	Times Cited 1997	Times Cited 1999	Times Cited 2007	Times Cited 2009	Coefficient of Variation (%)
<i>Journal of Speech, Language, and Hearing Research</i> (including preceding titles)	757	821	820	762	4
<i>Journal of the Acoustical Society of America</i>	303	275	476	315	27
<i>Journal of Communication Disorders</i>	41	27	43	60	32
<i>American Journal of Speech-Language Pathology</i>	22	38	34	56	38
<i>Journal of Child Language</i>	36	46	94	85	44
<i>Quarterly Journal of Experimental Psychology</i>	14	6	5	14	51
<i>Language</i>	7	21	19	6	59
<i>Dysphagia</i>	7	6	23	36	79
Average of 57 titles that received ≥ 5 citations in each of the 4 years					38

The Spearman's rho rank coefficient statistical test was used to measure whether rankings vary more over time than they do between the titles. The data for each pair of journals shown in table 2 were assigned ranks. For example, articles in the *JCD* 1997–99 had 89 journals cited at least 5 times. Many of these received an equal number of citations, but to calculate r_s , one must have 89 ranks for 89 journals. Ties were summed and averaged, e.g., 2 titles receiving 19 citations each were both ranked 16.5, and 6 titles receiving 11 citations were each ranked 34.5. Each pair of ranked lists included titles with no match. The nonmatched titles were given a rank one below the bottom. For instance, when correlating with the ranked list of 89 titles from *JCD* 1997–99, journals from the other lists that did not match were ranked 90. The correlations of ranked lists are shown in table 2.

Calculating rank correlations is a valid way to measure reliability; analyzing data from more than one perspective can be helpful. Another way to examine whether rankings vary more over time than they do between the titles is to calculate coefficients of variation. Coefficient of variation is the standard deviation divided by the mean. Standard deviation is only meaningful in the context of sample size as a measure of the dispersal of scores. Dividing standard deviation by the mean produces a relative dispersal that allows comparison of the spread of scores among different sample sizes.⁴¹ Coefficients of variation were calculated for the cited journals appearing in all four ranked lists from the volumes of *JSLHR*. The average variation and examples representing the range of variation are shown in table 3.

The formula for Spearman's rho takes into account journals that change dramatically in rank. But because it treats subjects as a group, the statistical measure can obscure significant changes at the level of individual journals. To account for this, journals that moved most in rank between 1997–99 and 2007–9 were identified and scrutinized. Myriad characteristics might affect the impact of a

journal. The author chose to investigate dates of publication, publisher, number of articles published in 1997, number of articles published in 2009, price, and availability online. The author retrieved dates of publication from WorldCat and determined publishers from the journals' websites. Article counts came from databases including *Medline*, *PsycINFO*, and *Communication and Mass Media Complete*, with the exception of *Journal of Medical Speech-Language Pathology*, which was determined by examining the journal. The author used EBSCOnet to look up prices. For purposes of comparison the author used prices for print subscriptions when available. For titles with tiered or otherwise differentiated pricing, subscription prices applicable to a masters-level college were recorded. Online availability was found in EBSCOnet and confirmed from Serials Solutions' data, searching in the identified databases, and checking publishers' websites.

Results

The comparison of works cited between 1997–99 and 2007–9 shows growth in both the number of articles published per volume and the number of works cited per article. The most obvious case of growth in the literature is *JCD*'s substantial increase in both the number of articles published and the average number of works cited per article. The *JSLHR* also experienced growth in articles and works cited. The sample as a group experienced a 32 percent increase in works cited over the decade.

The list of most frequently cited journals for 2007–9 from the *JSLHR* and *JCD* is shown the appendix. The sample contained 15,631 works cited, a total that includes citations to books, conferences, tests, and other resources. Ranks are based on the percentage of citations received, and the rightmost column indicates the change in rank relative to

the 2001 list of 104 journals representing 80 percent of citations to journals.⁴² The method of downloading works cited from *PsycINFO* obscured the type of resource being cited, so the percentages of works cited shown in the appendix total less than 80 percent. Titles that were not on the list of 104 titles published in 2001 were assigned a rank of 104, thus counting them as ties for last place. The Spearman rank correlation coefficient of the 1997–99 ranks against the 2007–09 ranks is $r_s = 0.73$. In essence, the Spearman correlation formula squares the differences in rankings and divides that by how many items are being ranked. By design, dramatic changes in rank by many items in the measured group will cause r_s to approach zero, i.e., no relationship between the two ranked lists. No changes in rank between lists would result in $r_s = 1$, a perfect correlation and total reliability between the two measurements. Any coefficient above 0.70 is generally considered a very strong positive correlation that indicates a high degree of reliability. However, the $r_s = 0.73$ must be considered within the context of two key factors: random variability in the journals authors cite and possible variation due to the composition of the sample.

Some of the change in which journals are cited is due to random variability. This makes determining the degree to which changes in rank are due to real changes in the literature or simply due to natural variability in which journals authors choose to cite difficult. Table 2 shows reliability of ranked lists among the volumes comprising the sample. As one might intuitively expect, the strongest correlations are between volumes of the same journal. Reliability as measured by Spearman's rho rank correlations for volumes of *JSLHR* range from $r_s = 0.60$ for the 1997 and 2007 volumes to $r_s = 0.71$ for both 1997–99 and 1999–2007. The weakest correlations are for the 1997–99 volumes of the *JCD*. This may be due to differences in topic coverage, but is likely because *JCD* 1997–99 contained the fewest number of works cited. The smaller sample size may have caused the lower reliability. Striking is that the strong positive $r_s = 0.73$ between the entire 1997–99 and 2007–9 groups indicates greater reliability than exists from any of the volumes shown in table 2. In that context, the simplest answer to the question “how much does the core change over a decade?” is “not much.” In fact, the between-group correlation of $r_s = 0.73$ compared to the within-group correlations shown in table 2 allows one to reasonably infer that most of the measured change in core journals over one decade was due to random variability.

The effect of composition of the sample can be inferred from the correlations shown in table 2. Each volume represents its own sample, so one can see that the reliability of ranked lists from these six samples ranges from $r_s = 0.52$ to $r_s = 0.71$, all indicators of strong, but not perfect, correlations. Another way to analyze the effect of sample selection is to compare these results to Slater's list of frequently cited

journals in communication disorders.⁴³ Slater used a method very similar to this authors' except that her sample was based on four journals. In addition to *JSLHR* and *JCD*, she also counted works cited in the *American Journal of Speech-Language Pathology* and the *British Journal of Disorders of Communication/European Journal of Disorders of Communication*. The period for Slater's study was 1991 through 1993. The correlation coefficients for Slater's results versus the present results are $r_s = 0.54$ for the 1997–99 list and $r_s = 0.38$ for the 2007–9 list. The difference in samples makes it difficult to determine how much of the drop in reliability is because of true changes in the literature. However, the fact that reliability of ranked lists drops from $r_s = 0.54$ after six years to $r_s = 0.38$ after sixteen years suggests a continual, albeit rather gradual, change in core journals over time.

Spearman's rho measures reliability on the basis of correlations of ranked lists. Another way to look at the degree of variability is to calculate coefficients of variation using the method explained above. Because the scope of *JCD* is somewhat different from that of *JSLHR*, only ranked lists from *JSLHR* were used for this analysis. Fifty-seven cited journals appeared on all four rankings from the volumes of *JSLHR*. As shown in table 3, the average coefficient of variation was 38 percent. One can infer from this that roughly one-third of the variation in times cited is because of random variation. But this too varies, from the highly reliable 4 percent rate for the top journal to the very unpredictable 79 percent for *Dysphagia*. One sees a general trend for the most-cited journals to have the lowest coefficients of variation. That would come as no surprise to Line, who emphasized the difficulty of predicting use of journals on the fringe of a core.⁴⁴

A reliability coefficient of $r_s = 0.73$ indicates that the core list of communication disorders journals changed relatively little over a decade, but this does not mean ranks of individual titles all stayed about the same. Table 4 shows the 10 journals that moved up the most in rank and selected characteristics that might suggest reasons for the changes. *Trends in Cognitive Sciences* and *Developmental Science* changed for the obvious reason that they were new launches. Length of time in publication also could be a major reason for increased citations to *Journal of Medical Speech-Language Pathology*, *NeuroImage*, and *Psychological Science*.

The case of the *Journal of Medical Speech-Language Pathology* is very interesting. It defies conventional wisdom to become so well established in the discipline despite being published only in print by a smaller publisher. Moreover, it is only indexed in *Cumulative Index to Nursing and Allied Health Literature* and *Linguistics Language and Behavior Abstracts*, not in the commonly used databases for communication disorders such as *Medline*, or *PsycINFO*, or *Communication and Mass Media Complete*. This shows that researchers manage to find their way to highly relevant

Table 4. Journals with Greatest Increases in Rank

Title	Rank Change	Pub. Dates	Publisher	Articles 1997	Articles 2009	Price	Online
<i>Trends in Cognitive Sciences</i>	+76	1997–	Elsevier	109	80	\$2,043	ScienceDirect
<i>Journal of Medical Speech-Language Pathology</i>	+63	1993–	Delmar	25	19	\$177	no (published in print only)
<i>Language and Cognitive Processes</i>	+60	1985–	Taylor & Francis	34	69	\$2,013	with 12-month embargo in Communication & Mass Media Complete
<i>Journal of Neuroscience</i>	+57	1981–	Society for Neuroscience	897	1595	\$4,240	OA after 6-month embargo
<i>Proceedings of the National Academy of Sciences of the United States of America</i>	+52	1915–	National Academy of Sciences	2681	4026	\$2,150	OA after 6-month embargo
<i>Journal of Applied Physiology</i>	+50	1985–	American Physiological Society	554	621	\$1,425	OA after 1-year embargo, plus “Author Choice” OA option
<i>NeuroImage</i>	+47	1992–	Elsevier	61	745	\$2,483	ScienceDirect
<i>Cleft Palate Journal + Cleft Palate-Craniofacial Journal</i>	+44	1964–90 +1991–	American Cleft Palate-Craniofacial Association	84	100	\$347	Publisher’s website by subscription, 1964–89 free from U Pitt Digital Library
<i>Psychological Science</i>	+38	1990–	Sage	85	236	\$5,538	Publisher’s website by subscription
<i>Developmental Science</i>	+35	1998–	Wiley-Blackwell	n/a	117	\$1,203 (online)	Academic Search Premier after 1-year embargo

literature. That also could be an explanation for the success of *Cleft Palate-Craniofacial Journal*. A possible factor in the success of those two journals is their relatively low subscription prices.

The remaining titles on the list of journals with greatest increases in rank are *Language and Cognitive Processes*, *Journal of Neuroscience*, *PNAS*, and *Journal of Applied Physiology*. All of these are either Open Access (OA) or in full-text in *Communication and Mass Media Complete* after an embargo period. Knowing the exact effect of OA on use by researchers is impossible, but of interest is that OA is more prevalent among journals that moved up in table 4 than it is among journals that moved down in rank. However, the four journals all experienced substantial increases in the number of articles published per year, which may be a more significant factor than online availability.

Several titles decreased in rank; see table 5. The journal experiencing the greatest drop was for the blindingly obvious reason that it ceased publication. It merged into *ASHA Leader*, the newspaper of the American Speech-Language-Hearing Association. Three of the journals that dropped in rank published fewer articles in 2009 than in 1997: *Seminars in Hearing*, *Language*, and *Annals of Neurology*. Two others had no significant growth in articles published: *Otology and Neurotology* and *Topics in Language Disorders*. Decreases

in rank for *Archives of Neurology*, *Neurology*, and *Annals of Neurology* hint at less use of general neurology journals by communication disorders researchers, but natural variability in what is cited prevents drawing any conclusions about that.

Discussion

One may conclude on the basis of the method used here that the core journals in communication disorders have changed little over a decade. But even though the list as a whole was stable, some journals did change significantly in rank. This study has several implications for collection development. One is that a core of journals carefully selected to support researchers can be expected to need only minor adjustments over time. If the case of communication disorders can be generalized to other fields, librarians can rest assured that core journals change gradually. Normal ongoing collection development activities (regrettably including cancellation projects) should be sufficient to keep the core aligned with researchers’ needs.

A second implication is that collection development is somewhat easier for libraries serving a range of disciplines than it is for libraries serving a special population. A discipline like communication disorders draws from the literature

Table 5. Journals with Greatest Decreases in Rank

Title	Rank Change	Pub. Dates	Publisher	Articles 1997	Articles 2009	Price	Online
ASHA	-85	1959–99	Am. Speech & Hearing Assoc.	71	n/a	ceased	13 selected articles
<i>Seminars in Hearing</i>	-65	1983–	Thieme	36	27	\$486	publisher's website by subscription
<i>Electroencephalography and Clinical Neurophysiology + Clinical Neurophysiology</i>	-58	1949–98 +1999–	Elsevier	248	584	\$1397	ScienceDirect
<i>American Journal of Otology + Otology & Neurotology</i>	-49	1979–2000 +2001–	Lippincott Williams & Wilkins	213	219	\$719	publisher's website by subscription
<i>Archives of Neurology</i>	-49	1960–	Am. Medical Assoc.	194	282	\$630	publisher's website, select OA articles
<i>Journal of Pediatrics</i>	-41	1932–	Mosby/Elsevier	982	1541	\$767	ScienceDirect
<i>Neurology</i>	-41	1951–	Lippincott Williams & Wilkins	883	1025	\$1152	publisher's website by subscription
<i>Topics in Language Disorders</i>	-40	1980–	Lippincott Williams & Wilkins	29	31	\$346	publisher's website, select OA articles
<i>Language</i>	-36	1925–	Linguistic Soc. of Am.	270	98	\$172	Project MUSE
<i>Annals of Neurology</i>	-37	1977–	Wiley	263	256	\$852	publisher's website by subscription, author "OnlineOpen" OA option

of many fields. A research library will probably have many heavily cited journals not specific to the discipline, such as *Nature*, *Science*, *Cognition*, and *Brain*. Smaller, specialized libraries are less likely to have such a broad-based collection to build upon.

A third implication is that variability in what is cited as shown in tables 2 and 3 warns librarians and scholars to treat with skepticism all ranked lists of journals. A ranked list can be a good starting point for collection development, selecting a journal to submit an article to, or suggesting appropriate places to browse. The principle of using ranks as only one source of information among many important considerations has been established in court. Judge Leonard B. Sand heard the case of publisher Gordon and Breach against Henry Barschall, who had published a ranked lists of journals based on cost relative to impact.⁴⁵ Sand wrote that “[Gordon and Breach] have proved only the unremarkable proposition that a librarian would be ill-advised to rely on Barschall’s study to the exclusion of all other considerations in making purchasing decisions.”⁴⁶ One should use rankings with the understanding that they have limited reliability. A different selection of sampled journal titles or years of coverage will result in a different ranked list. Rankings drawn from small samples should be treated with increased

caution. Lower reliability due to a limited sample is an inherent problem with ranked lists in a subspecialty not categorized in *JCR* or *Scopus*.

Limitations of Study

A cynical interpretation of this study of core journals in communication disorders could be that the inherent unreliability of ranked lists renders it all a waste of effort. But ranked lists are still useful starting points for identifying core journals, and these data have allowed a quantified analysis of the reliability of rankings in one field. The degree to which the case of communication disorders can be generalized to other disciplines is unknown.

Although this sample size was a robust $n = 15,631$ drawn from two of the most cited journals in the field, a broader selection of source journals would probably yield a ranked list that more accurately reflects the literature in communication disorders. The author was unfortunately unaware of Slater’s work when he wrote the 2001 article published in *LRTS*.⁴⁷ Replicating that sample of four journals would have expanded representation of the literature and allowed direct comparisons across three time periods. The author advises anyone conducting similar research to develop the sampled

journals in an iterative process, starting with a journal known to be a leader in the field, then expanding the selection of source journals based on preliminary rankings. A sample of at least four source journals would be superior to using only one or two.

Further Research

Additional research could investigate whether the characteristics of the journals in communication disorders discovered here can be generalized to other fields. Of particular interest is whether other disciplines' core journals experience similar reliability over time. This could be relatively easily done for disciplines categorized in *JCR* by a researcher with access to historical and current editions of the *JCR*. A fertile area for further research is to develop methods for using Google Scholar for citation analysis.

An area deserving attention is how large a sample of source journals is sufficient to yield a ranked list of a desired degree of reliability. Does one volume of one journal consistently produce a ranked list with about $r_s = 0.65$ reliability? How small a sample is capable of identifying the top ten journals in a field with $r_s = 0.50$ reliability, or $r_s = 0.75$? What is the maximum attainable reliability with large samples—can any sample size attain $r_s = 0.90$? Answers to questions like these would help librarians determine the degree of confidence they should have that ranked lists accurately reflect true use of the literature.

Conclusions

How reliable is a ranked list over time? The primary purpose of this study was to determine how much a ranked list of core journals changed during a decade. This was accomplished by replicating a 2001 study and calculating the Spearman's rho rank correlation coefficient of the two ranked lists. A second research question was whether variability in ranked lists was greater over time than it was among citations in source journals from the same timeframe. To answer that question, correlations were calculated to compare rankings derived from citations in two leading journals. In the process of answering those questions, discoveries were made regarding growth in literature and journals that changed most in rank were identified. This study supports these conclusions for the case of core journals in communication disorders:

- Journal literature grew substantially during the decade both in articles published and works cited per article.
- As a group, core journals remained quite stable over time ($r_s = 0.73$).
- Despite high reliability of core journals over time, some titles changed dramatically in rank.

- Approximately one-third of change in ranked lists derived from works cited in single journal volumes is due to random variability.

The stability of a core list over time means that normal collection development activities should be sufficient to keep a collection aligned with a discipline's core of most frequently cited journals. While ranked lists are useful collection development tools, random variability in rankings cautions librarians to treat any ranked listing with some skepticism.

References and Note

1. Steve Black, "Using Citation Analysis to Pursue a Core Collection of Journals for Communication Disorders," *Library Resources & Technical Services* 45, no. 1 (2001): 3–9.
2. Inna Shpilko, "An Overview of the Journal Literature on Communication Disorders," *Collection Building* 22, no. 3 (2003): 102–19.
3. Thomas E. Nisonger, "Journals in the Core Collection: Definition, Identification, and Applications," *Serials Librarian* 51, no. 3/4 (2007): 51–73.
4. R. E. Rice et al., "Journal-to-Journal Citation Data: Issues of Validity and Reliability," *Scientometrics* 15, no. 3 (1989): 257–82.
5. Dean F. Sittig, "Identifying a Core Set of Medical Informatics Serials: An Analysis Using the Medline Database," *Bulletin of the Medical Library Association* 84, no. 2 (1996): 200–204.
6. Eugene Garfield, "Citation Analysis as a Tool in Journal Evaluation," *Science* 178 (1972): 471–79.
7. Thomson Reuters, "The Thomson Reuters Impact Factor," http://thomsonreuters.com/products_services/science/free/essays/impact_factor (accessed Dec. 14, 2011).
8. B. R. Martin, "The Use of Multiple Indicators in the Assessment of Basic Research," *Scientometrics* 36, no. 3 (1996): 343–62.
9. Daniela Rosenstreich and Ben Wooliscroft, "Measuring the Impact of Accounting Journals using Google Scholar and the G-Index," *British Accounting Review* 41 (2009): 227–39.
10. John Hudson, "Be Known by the Company You Keep: Citations—Quality or Chance?" *Scientometrics* 71, no.2 (2007): 231–38.
11. Robert K. Merton, *The Sociology of Science: Theoretical and Empirical Investigations* (Chicago: University of Chicago Press, 1973).
12. Per O. Seglen, "Why the Impact Factor of Journals Should Not be Used for Evaluating Research," *British Medical Journal* 314, no. 7079 (1997): 498–502.
13. Michael H. MacRoberts and Barbara R. MacRoberts, "Problems of Citation Analysis: A Critical Review," *Journal of the American Society for Information Science* 40, no. 5 (1989): 342–49.
14. Stephen C. Hubbard and Marie E. McVeigh, "Casting a Wide Net: The Journal Impact Factor Numerator," *Learned Publishing* 24, no. 2 (2011): 133–37.
15. Bob Perdue and Chris Piotrowski, "Citation Searching: A

- New Feature in PsycINFO," *Perceptual & Motor Skills* 96 (2003): 1237–38.
16. Katherine W. McCain, "Core Journal Networks and Cocitation Maps: New Bibliometric Tools for Serials Management," *Library Quarterly* 61, no. 3 (1991): 311–36.
 17. Jeffery D. Kushkowski, Kristin H. Gerhard, and Cynthia Dobson, "A Method for Building Core Journal Lists in Interdisciplinary Subject Areas," *Journal of Documentation* 54, no. 4 (1998): 477–88.
 18. Thura Mack, "A Model Methodology for Selecting Women's Studies Core Journals," *Library & Information Science Research* 13 (1991): 131–45.
 19. Dean F. Sittig, "Identifying a Core Set of Medical Informatics Serials: An Analysis Using the Medline Database," *Bulletin of the Medical Library Association* 84, no. 2 (1996): 200–204.
 20. David A. Goss, "Citation Patterns in the Optometric and Ophthalmologic Clinical Binocular Vision Literature," *Optometry & Vision Science* 83, no. 12 (2006): 895–902.
 21. Christina C. Wray, "The Journey Starts Here: Finding Special Education Research in Subscription Databases," *Reference Librarian* 52 (2011): 231–43.
 22. Linda G. Slater, "Mapping the Literature of Speech-Language Pathology," *Bulletin of the Medical Library Association* 85, no. 3 (1997): 297–302.
 23. Anne-Wil Harzing and Ron van der Wal, "A Google Scholar H-Index for Journals: A Better Metric to Measure Journal Impact in Economics and Business?" (paper submitted to the 2008 Academy of Management Annual Meeting August 8–13, 2008, Anaheim, California).
 24. S. C. Bradford, "Sources of Information on Specific Subjects," *Engineering* 137 (1934): 85–86.
 25. P. Nieuwenhuysen and R. Rousseau, "A Quick and Easy Method to Estimate the Random Effect on Citation Measures," *Scientometrics* 13, no. 1/2 (1988): 45–52.
 26. Maurice B. Line, "Use of Citation Data for Periodicals Control in Libraries: A Response to Broadus," *College & Research Libraries* 46, no. 1 (1985): 36.
 27. Maurice B. Line, "Changes in Rank Lists of Serials Over Time: Interlending versus Citation Data," *College & Research Libraries* 46, no. 1 (1985): 77–79.
 28. Robert N. Broadus, "A Proposed Method for Eliminating Titles from Periodical Subscription Lists," *College & Research Libraries* 46, no. 1 (1985): 30–35.
 29. Ibid.
 30. John Laurence Kelland and Arthur P. Young, "Citation as a Form of Library Use," *Collection Management* 19, no. 1/2 (1994): 81–100.
 31. Murray J. White and K. Geoffrey White, "Citation Analysis of Psychology Journals," *American Psychologist* 32, no. 5 (1977): 3015.
 32. Rick Ralston, Carole Gall, and Frances A. Brahma, "Do Local Citation Patterns Support Use of the Impact Factor for Collection Development?" *Journal of the Medical Library Association* 96, no. 4 (2008): 374–78.
 33. Janice Kreider, "The Correlation of Local Citation Data with Citation Data from Journal Citation Reports," *Library Resources & Technical Services* 43, no. 2 (1999): 67–77.
 34. Stephen J. Bensman, "Probability Distributions in Library and Information Science: A Historical and Practitioner Viewpoint," *Journal of the American Society for Information Science & Technology* 51, no. 9 (2000): 816–33.
 35. S. C. Bradford, "Sources of Information on Specific Subjects."
 36. Sidney Siegel, "Nonparametric Statistics," *American Statistician* 11, no. 3 (1957): 13–19.
 37. Maurice Kendall and Jean Gibbons, *Rank Correlation Methods*, 5th ed. (New York: Oxford Univ. Pr., 1990).
 38. Merriam-Webster, "Reliability," www.merriam-webster.com/dictionary/reliability (accessed Dec. 21, 2011).
 39. Linda G. Slater, "Mapping the Literature of Speech-Language Pathology."
 40. Black, "Using Citation Analysis to Pursue a Core Collection of Journals for Communication Disorders," 6.
 41. Murray R. Spiegel, *Schaum's Outline of Theory and Problems of Statistics* (New York: Schaum, 1961): 73.
 42. Black, "Using Citation Analysis."
 43. Slater, "Mapping the Literature of Speech-Language Pathology."
 44. Line, "Use of Citation Data."
 45. See Gordon and Breach v. American Institute of Physics and American Physical Society, <http://barschall.stanford.edu> (accessed Dec. 14, 2011) for a thorough and well organized treatment of the Gordon and Breach case against Barschall and the publishers of his studies.
 46. Gordon and Breach v. American Institute of Physics and American Physical Society, 973 F. Supp. 414 (U.S. District Court for Southern New York), Aug. 26, 1997, <http://barschall.stanford.edu/opinions/970826OPIN.html> (accessed Dec. 14, 2011).
 47. Slater, "Mapping the Literature of Speech-Language Pathology."

**Appendix. Core Journals in Communication Disorders
Based on Times Cited 2007–9 (N = 15,631)**

Title	% of Cites	Δ Rank since 1999
<i>Journal of Speech, Language, & Hearing Research</i> + <i>Journal of Speech & Hearing Research</i> + <i>Journal of Speech & Hearing Disorders</i>	16.25	0
<i>Journal of the Acoustical Society of America</i>	6.74	0
<i>Journal of Communication Disorders</i>	1.52	+3
<i>Journal of Child Language</i>	1.49	+6
<i>Ear & Hearing</i>	1.44	+2
<i>Journal of Fluency Disorders</i>	1.38	-1
<i>Journal of Child Psychology & Psychiatry</i> + <i>Journal of Child Psychology & Psychiatry & Allied Disciplines</i>	1.24	+26
<i>Brain & Language</i>	1.20	-4
<i>British Journal of Disorders of Communication</i> + <i>European Disorders of Communication</i> + <i>International Journal of Language & Communication Disorders</i>	1.17	+9
<i>Journal of Voice</i>	1.17	+5
<i>Child Development</i>	1.07	0
<i>American Journal of Speech-Language Pathology</i>	1.06	+1
<i>Applied Psycholinguistics</i>	0.99	-1
<i>Clinical Linguistics & Phonetics</i>	0.92	+1
<i>Language, Speech, & Hearing Services in Schools</i>	0.92	-7
<i>Cognition</i>	0.86	0
<i>Developmental Psychology</i>	0.75	+10
<i>Journal of Autism & Developmental Disorders</i>	0.75	+28
<i>Journal of Phonetics</i>	0.74	-11
<i>Journal of Memory & Language</i>	0.69	+4
<i>Hearing Research</i>	0.63	+29
<i>Folia Phoniatica</i> + <i>Folia Phoniatica et Logopaedica</i>	0.61	0
<i>Annals of Otology, Rhinology, & Laryngology</i>	0.59	+11
<i>Journal of the American Academy of Audiology</i>	0.57	+11
<i>Science</i>	0.55	-8
<i>Journal of Applied Physiology</i>	0.47	+50
<i>Journal of Experimental Child Psychology</i>	0.46	+19
<i>Brain</i>	0.45	+30
<i>Trends in Cognitive Sciences</i>	0.45	+76
<i>Language & Speech</i>	0.44	-1
<i>Dysphagia</i>	0.44	+14
<i>Nature</i>	0.44	-3
<i>Psychological Review</i>	0.44	-12
<i>Laryngoscope</i>	0.41	-11
<i>Acta Otolaryngologica</i>	0.41	+14
<i>Journal of Neurophysiology</i>	0.41	+1
<i>Aphasiology</i>	0.40	-11
<i>Developmental Medicine & Child Neurology</i>	0.36	+5
<i>Perceptual & Motor Skills</i>	0.35	+9
<i>Neuropsychologia</i>	0.35	-8

**Appendix. Core Journals in Communication Disorders
Based on Times Cited 2007–9 (*N* = 15,631) (continued)**

Title	% of Cites	Δ Rank since 1999
<i>Journal of Medical Speech-Language Pathology</i>	0.34	+63
<i>Perception & Psychophysics</i>	0.33	-17
<i>Journal of Neuroscience</i>	0.32	+57
<i>Language & Cognitive Processes</i>	0.31	+60
<i>Memory & Cognition</i>	0.30	+30
<i>Psychological Bulletin</i>	0.30	+6
<i>Cleft Palate Journal + Cleft Palate-Craniofacial Journal</i>	0.29	+44
<i>Cognitive Psychology</i>	0.29	+11
<i>Journal of Experimental Psychology: Human Perception & Performance</i>	0.29	-8
<i>Pediatrics</i>	0.29	-21
<i>Augmentative & Alternative Communication</i>	0.28	+4
<i>Cortex</i>	0.28	+20
<i>Proceedings of the National Academy of Sciences of the United States of America</i>	0.28	+52
<i>Archives of Otolaryngology + Archives of Otolaryngology—Head & Neck Surgery</i>	0.27	-18
<i>Audiology</i>	0.27	-10
<i>Journal of Experimental Psychology: Learning, Memory, & Cognition</i>	0.27	+13
<i>Journal of Learning Disabilities</i>	0.27	-18
<i>NeuroImage</i>	0.27	+47
<i>British Journal of Audiology</i>	0.26	+12
<i>Journal of Verbal Learning & Verbal Behavior</i>	0.26	+3
<i>Topics in Language Disorders</i>	0.26	-40
<i>Phonetica</i>	0.25	-32
<i>Seminars in Speech & Language</i>	0.25	+11
<i>Monographs of the Society of Research in Child Development</i>	0.24	+21
<i>Speech Communication</i>	0.24	+21
<i>Experimental Brain Research</i>	0.24	-3
<i>Journal of Educational Psychology</i>	0.24	+10
<i>Psychological Science</i>	0.24	+38
<i>Developmental Science</i>	0.22	+35
<i>Quarterly Journal of Experimental Psychology</i>	0.21	-8
<i>Scandinavian Audiology</i>	0.21	-3
<i>Developmental Neuropsychology</i>	0.21	+30
<i>First Language</i>	0.21	-15
<i>International Journal of Audiology</i>	0.21	+30
<i>Journal of Cognitive Neuroscience</i>	0.21	+30
<i>Journal of the American Academy of Child & Adolescent Psychiatry</i>	0.21	+30
<i>Infant Behavior & Development</i>	0.20	-23
<i>Journal of Experimental Psychology: General</i>	0.20	+26
<i>Language</i>	0.20	-36
<i>Volta Review</i>	0.20	-11
<i>Audiology & Neuro-Otology</i>	0.19	+23

**Appendix. Core Journals in Communication Disorders
Based on Times Cited 2007-9 (N = 15,631) (continued)**

Title	% of Cites	Δ Rank since 1999
<i>Behavioral & Brain Sciences</i>	0.19	+23
<i>American Journal of Audiology</i>	0.18	-11
<i>American Journal of Human Genetics</i>	0.18	+18
<i>Brain Research</i>	0.18	+6
<i>Journal of Psycholinguistic Research</i>	0.18	-31
<i>Language Acquisition</i>	0.18	-18
<i>Neuroreport</i>	0.18	+18
<i>Reading & Writing: An Interdisciplinary Journal</i>	0.18	+18
<i>Archives of Physical Medicine & Rehabilitation</i>	0.17	+14
<i>Reading Research Quarterly</i>	0.17	+14
<i>American Journal of Mental Retardation</i>	0.16	-8
<i>Cognitive Development</i>	0.16	+11
<i>Journal of Neurology, Neurosurgery, & Psychiatry</i>	0.16	-1
<i>American Journal of Medical Genetics</i>	0.16	-4
<i>Electroencephalography & Clinical Neurophysiology + Clinical Neurophysiology</i>	0.16	-58
<i>American Journal of Otolaryngology + Otolaryngology & Neurotology</i>	0.15	-49
<i>American Journal of Psychiatry</i>	0.15	+4
<i>International Journal of Pediatric Otorhinolaryngology</i>	0.15	+4
<i>Journal of Neurolinguistics</i>	0.15	+4
<i>Lancet</i>	0.15	+4
<i>Lingua</i>	0.15	+4
<i>Neuroscience Letters</i>	0.15	+4
<i>Annals of Dyslexia</i>	0.14	-18

Notes on Operations

Need Exceeds Space

A Serials Withdrawal Project at the University of Rhode Island University Libraries

Brian T. Gallagher and Andrée J. Rathemacher

In 2009, the University of Rhode Island's main library, the Robert L. Carothers Library and Learning Commons in Kingston, initiated a pilot serials withdrawal project when the need for space for new services exceeded the space available in the library. A joint venture between Acquisitions and Access Services, the successful pilot led to a subsequent withdrawal project in summer 2010 to free additional space in the library. The print journals targeted for withdrawal were ones to which the library had online access through licensed journal archives. Considerations on what to withdraw, the process of identifying titles for withdrawal, and the logistics of managing the withdrawal of more than 35,000 volumes are described.

As library collections move online in the digital era, institutions are repurposing space formerly used to house collections for student use and enhanced services in support of teaching and learning.¹ As reported by Payne, “the ‘library as place’ movement has redefined what space within the library should be used for. As a result, libraries are coming to be seen primarily as centers for independent and collaborative study and learning rather than as housing for physical collections.”² Schonfeld and Housewright acknowledge that “the information/learning commons movement to create suitable learning spaces and bring new services into the library has been transforming the physical space of library after library.”³ As a result, “libraries turn to the deaccessioning of print as a key tactic for finding the needed space.”⁴

In response to plans to construct a learning commons in the University of Rhode Island's (URI) main library in Kingston and the consequent need to free space for new services, the library initiated a pilot serials-discard project during the summer of 2009. Other options for creating space to accommodate the learning commons, such as constructing an addition to the library or transferring collections to off-site storage, were beyond the library's financial resources. Furthermore, a tight schedule for construction required a timely response.

Successfully completed well before the construction deadline, the serials-discard pilot was continued in summer 2010 as a means to create additional strategic space in the library. The extension of the serials-discard pilot offered the opportunity to refine procedures and reflect on broader considerations surrounding the withdrawal of library collections in the digital era, such as the

Brian T. Gallagher (btg@uri.edu) is Head of Access Services and **Andrée J. Rathemacher** (andree@uri.edu) is Head of Acquisitions, University Library, University of Rhode Island, Kingston, Rhode Island.

Submitted July 19, 2011; returned with request for revision August 7, 2011; revised and resubmitted October 7, 2011; tentatively accepted Nov. 7, 2011 pending modest revision; final revision submitted November 15, 2011, and accepted for publication.

desirability of collaboration with faculty and other library stakeholders as well as the importance of coordination with other libraries to ensure that unique materials are not lost to the library community. The process followed can inform similar projects in other academic libraries.

Literature Review

As academic libraries face space constraints and budget pressures within the context of the mass migration of serials content to online format, many libraries have made the decision to withdraw print serials from their collections. O'Connor and Jilovsky point out that "the rate of discard of materials from academic libraries in recent years has become a fast growing statistic" driven primarily by lack of space and "new directions for library physical space."⁵ Corroborating this view is the *Ithaka S+R Library Survey 2010: Insights from U.S. Academic Library Directors*, which found that "most libraries have become comfortable with deaccessioning or moving offsite their print journal collections after they have reliable digital access to copies of these materials: 91% have already done so or are planning to do so in the future."⁶ Besides space needs, an important factor in decisions to withdraw print journals from library collections are the costs involved in the long-term storage of print format materials. Estimates of these costs vary widely on the basis of research methods, with annual storage costs in open library stacks calculated to range from \$4.26 to \$56.20 per volume per year.⁷ Montgomery and King estimate the cost-per-use for bound, print journals to be \$30, compared to \$2 for electronic journals.⁸

Despite this trend to remove legacy print journal collections from libraries, a coordinated strategy for the systemic withdrawal of print serials has largely been lacking. The *Ithaka S+R*

Library Survey 2010 found that only a minority of library directors confirmed having specific strategies and policies for deaccessioning print journals that they have access to electronically.⁹ As O'Connor and Jilovsky note, "the loss of so much material in a largely uncoordinated, unrecorded fashion is a matter of considerable long-term concern."¹⁰ Instead, libraries have tended to develop criteria for which serials to remove from their collections locally and on an ad-hoc basis.

Before widespread electronic access to journal content was a factor in withdrawal decisions, these decisions were based primarily on anticipated low levels of use of the removed materials and availability of the material at other nearby libraries. For example, in the late 1980s, librarians at Texas A&M University developed criteria for selecting infrequently used serial titles to move to a storage area within the library.¹¹ They targeted serial titles with large back runs, especially ceased or canceled titles in science and technology and titles with outdated subject matter, outdated directories and superseded annuals, and titles duplicated in other libraries on campus. Similarly, librarians at Purdue University Libraries conducted a pilot study in which journal titles were identified for deselection using criteria that included publication date, completeness of holdings, and, primarily, the availability of the journals at specified benchmark institutions.¹²

As the availability of journal content online has become increasingly acceptable as a substitute for print, criteria for withdrawing print journals with online equivalents have focused on the reliability of the online content. A number of studies have shown that the use of print journals decreases when online versions are made available.¹³ In addition, recent surveys have found that the majority of faculty members now prefer to access journals in online format.¹⁴ In her review of the literature, Sorensen notes that

"by 2004, there is evidence that libraries were actually withdrawing print backruns, but only if the online versions met certain standards."¹⁵ At the University of Saskatchewan, JSTOR-available print journals were flagged for disposal in 2006/2007 "because of the stable, comprehensive, and secure nature of this collection of scholarly electronic journals," while publisher-available print journals were placed into storage "given that these online journals were generally complete and stable."¹⁶

The University of Arizona's Science-Engineering Library likewise decided to remove print journals as online archival backfiles that met their "standards for perpetual access and commitment to digital preservation" were purchased.¹⁷ Backfiles are "evaluated for completeness, quality, and publisher commitment to digital preservation" and "publishers must include a license that allows for perpetual access."¹⁸ In a similar vein, to solve acute space pressure, Imperial College in London devised a strategy to withdraw print holdings for which the library had "sustainable" electronic access. Sustainable meant the anticipated availability of free or affordable online access for as long as required through one of the following: perpetual access rights to the content via the web, the journal being permanently open access, or the journal being in a trusted service such as JSTOR, the Association for Computing Machinery (ACM) digital archive, or a JISC-funded archive.¹⁹

In response to this growing movement to reclaim library space by removing back runs of print journals from library collections, Ithaka released a report in 2009 titled *What to Withdraw? Print Collections Management in the Wake of Digitization*.²⁰ Schonfeld and Housewright, the report's authors, clearly define criteria for responsibly withdrawing print journals from library collections. Journals that are good candidates for withdrawal

have these characteristics: they were digitized with high standards of quality, scanning errors are being actively corrected, digital copies are reliably preserved, they are not image-intensive, and they are available through reliable license terms. The report concludes that JSTOR-digitized, text-only journals that are held in at least two print repositories meet these criteria and can be responsibly withdrawn from library collections. The conclusion that follows from Ithaka's report, as well as from the literature cited above, is that given reliable, affordable, and secure access to quality online journal content, libraries can comfortably make the decision to dispose of corresponding print volumes.

Background

The URI is a land-grant, sea-grant, and urban-grant institution founded in 1892 as the Rhode Island College of Agriculture and Mechanic Arts. It holds the Carnegie Classification of RU/H (research university with high research activity), enrolls more than 16,000 students, and employs close to 600 tenured or tenure-track faculty working in a broad range of disciplines. The University Libraries consist of the main library on the Kingston campus and two small branch libraries in Narragansett and Providence. The University Libraries are part of the HELIN Library Consortium, which is composed of eleven academic libraries, twelve hospital libraries, and one law library in Rhode Island and nearby Massachusetts. The university's main library in Kingston was built in 1964 and has had two major renovations since that time: one in 1976 and one in 1993. The building currently provides shelf space for 1.4 million volumes and seating capacity for 1,300 people. As is the case in many academic libraries, space is at a premium as collections expand and new services for library users are developed.

Growth in the URI library's collection of print periodicals has been slowing since the early 1980s as the result of ten major serials cancellation projects in which more than 3,000 subscriptions were cancelled as well as a steady shift from print to online-only current periodical subscriptions.²¹ Despite this slow rate of growth, the library's bound periodicals collection, classified and stored on the building's lower level, was overcrowded by the early 2000s. Efforts to shift material to create additional room were only minimally successful and little free shelf space remained. In many cases, shelvers had no choice but to shelve journals on top of other journals.

The opposite situation existed in the current periodicals area on the main floor. Since 2002, when the library's management team adopted a policy of converting print subscriptions to online-only format whenever practical, the number of current periodicals received in print format declined steadily to approximately 600 in 2011.²² In fiscal year 2010/11, library staff checked in 5,271 periodical issues, down from 17,227 in 2001/02, a decrease of 69 percent. Meanwhile, the number of online-only subscriptions continued to grow, reaching more than 57,000 titles in 2011, or about 90 percent of current titles. Including journals in JSTOR, online publisher backfiles, and full-text aggregator databases, the total number of paid online journals available to the URI in 2011 was 26,471.²³ As a result of this shift to online journals, the space designed in the early 1990s to house thousands of current periodicals was little-used, with the remaining print subscriptions scattered across mainly empty shelves.

2009 Pilot Project

An opportunity to address the overcrowding on the lower level and the excess space in the current periodicals

area came with the departure of Robert L. Carothers, the URI's tenth president, who retired in 2009 after eighteen years. As the result of a campaign to name the University Library in his honor, the URI's main library officially became the Robert L. Carothers Library and Learning Commons in May 2009. After renaming the library, university administrators began planning the construction of a learning commons in the area of the main level, where the current periodicals were housed. Construction would begin during the summer of 2010, before which the current periodicals would need to be relocated to the already overcrowded lower level with the bound periodicals.

The dean of libraries dismissed the idea of eliminating study space to accommodate the current periodicals, so the head of Acquisitions suggested that the library withdraw bound journals duplicated online. Eliminating digital duplication would open up physical space in the library without diminishing the total content available to library users, as would, for example, withdrawing low-use periodicals not available online. Furthermore, no reliable use data were available for the library's bound journals, because they do not circulate and reshelfing statistics by title are not maintained. Thus removing print volumes available online appeared to be the most rational and expedient strategy, especially given the time constraint.

After a decision was made about what to withdraw, the dean of libraries, head of Access Services, and head of Acquisitions sought to identify a suitable target area for weeding that would provide enough space for the relocated current periodicals plus additional space for study tables. They identified an appropriate space, very visible and therefore easy for library users to find, next to the main stairwell on the lower level. Because a deadline was involved and this was the first large-scale serials withdrawal project undertaken at

the University Libraries, the dean of libraries and heads of Access Services and Acquisitions decided to limit the project to the withdrawal of journals with Library of Congress call numbers beginning with Q (Science) and R (Medicine). Not only were the Qs and Rs physically close to the area that needed to be cleared to accommodate the current periodicals, but much of the library's online journal content was in the sciences, so many potential titles likely could be identified for discard. Though limited to two Library of Congress classifications, this withdrawal project could serve as a pilot for more extensive weeding in the future.

The heads of Access Services and Acquisitions chose the summer months of July and August to complete the first phase of the pilot: discarding the journals. The next phase, creating a new space for current periodicals through shifting volumes in the stacks, would take place during the fall semester. The shifting implications, according to the stacks supervisor, would be imperceptible and kept to a small area.

What to Withdraw

In deciding which titles to withdraw, the head of Acquisitions suggested specifically targeting volumes for which the library had online access through licensed journal archives or backfiles. The details of this purchasing model vary by publisher and vendor.

Targeting print volumes duplicated in online archives for withdrawal had both advantages and disadvantages. The most obvious advantage was that the material in the archives corresponded to the bound volumes that occupied the space that needed to be emptied. The online archives offer complete coverage of every issue of every title in the archive in the form of page images that can be viewed online, printed, and downloaded, thus offering a workable facsimile of the printed volumes. In addition, the library would be unlikely to cancel any

of these backfiles after having invested in the initial purchase price, especially because the annual access fees for all archives, totaling less than \$100,000, comprised a very small proportion of the library's \$3.2 million materials expenditures.

The disadvantages included the possibility of losing access to content for reasons other than nonpayment of annual access fees. A title can disappear from a publisher's online archive if the publisher sells the title to another publisher. In such cases, the library could lose access to the online backfiles of titles already withdrawn in print unless the publisher has endorsed the Transfer Code of Practice, developed by UKSG to address the challenges and implications of titles moving between publishers in the online environment.²⁴ The Transfer Code of Practice stipulates that "the Transferring Publisher must ensure continued access to its customers where it has granted perpetual access rights even if the Transferring Publisher will cease hosting the online version of the journal after the Effective Transfer Date. Either the Transferring or the Receiving Publisher, or both, could fulfill perpetual access obligations." Often ambiguous in license agreements for archives of a collection of journals is whether perpetual access applies to the collection in aggregate or to a specific, named set of titles.

The library also could lose access to backfile content if a publisher were to go out of business. To insure against such loss, many libraries use the LOCKSS (Lots of Copies Keep Stuff Safe) or Portico digital preservation services to guarantee their ongoing access to subscribed digital content. LOCKSS, based at Stanford University Libraries, is an "international community initiative that provides libraries with digital preservation tools and support so that they can easily and inexpensively collect and preserve their own copies of authorized e-content."²⁵

Libraries install LOCKSS software on a local server and program their "LOCKSS box" to harvest subscribed content from more than 500 participating publishers, including a number of smaller publishers not archived by Portico.²⁶

In contrast to LOCKSS' distributed model, Portico (a part of Ithaka), is a centralized repository of e-journals and other electronic content that was certified in 2009 by the Center for Research Libraries as a "trusted digital repository."²⁷ Portico maintains formal preservation agreements with 135 publishers and ingests their content into its archive.²⁸ In the case of a trigger event such as cessation of a publisher's operations, discontinuation of a title by a publisher, or back issues no longer offered by a publisher, participating libraries may access preserved content through Portico.²⁹

The URI Libraries do not participate in LOCKSS or Portico, as the support required for LOCKSS and the expense of Portico are not feasible given the library's current staffing and budget levels. Since the library's purchased archives are primarily from large and well-established publishers and vendors, the head of Acquisitions concluded that losing access to content as the result of publisher failure was a small risk. If a publisher were to go out of business, the library could join Portico at a later time to regain access to lost content. At the time of the withdrawal project, content from many, but not all, of the library's backfile collections was available through Portico, though holdings in Portico were often incomplete, with numerous missing volumes and issues.

Generating a List of Titles

Before any work could begin in the stacks, the head of Acquisitions needed to identify the specific titles that were candidates for withdrawal. The first step in doing so was to generate a list of all titles available to the

Table 1. Setup of Initial Spreadsheet for Serials Withdrawal Project

Column	Column Label	Data	Data Source
A	Call no.	QA	Millennium library system
B		11	Millennium library system
C		A	Millennium library system
D		.1	Millennium library system
E		J	Millennium library system
F		.68	Millennium library system
G	Title	<i>Journal for Research in Mathematics Education</i>	Serials Solutions
H	Archive	JSTOR A&S IV	Serials Solutions
I	Archive dates	1970 to 2003	Serials Solutions
J	URI discard(ed)	1977–2003	Millennium library system
K	Remaining on shelf	v. 35(2004)–v. 39(2008)	Physical inventory
L	.c record	c1021096	Millennium library system
M	# vols w/d	20	Physical count
N	Inches	28	Physical measurement

library through licensed online journal archives. The head of Acquisitions used the library's Serials Solutions Client Center to download the titles and years available in each of the library's online journal backfiles.

Using Serials Solutions instead of vendor websites to gather title lists offered one significant advantage in addition to convenience—Serials Solutions tracks title changes and provides a separate listing, with holdings, for each iteration of a journal, while vendors and publishers often collapse the holdings of previous titles under the most recent title. When the time came to match titles and holdings downloaded from Serials Solutions with library holdings using the library catalog, which follows the *Anglo-American Cataloguing Rules*, 2nd ed., standards on successive entry cataloging, having each version of a journal's title listed would save time and avoid frustration.³⁰

The head of Acquisitions merged the titles and holdings downloaded from Serials Solutions into a single spreadsheet that contained journal titles, names of online archives, and dates of coverage. She added blank columns for call number, years to

discard, volumes and years remaining on shelf, library system record number, number of volumes withdrawn, and number of inches withdrawn.

The next step was to determine which of the 3,169 titles on the spreadsheet were held by the URI in print format. To do this, the head of Acquisitions used the library's Innovative Interfaces Millennium integrated library system (ILS) to create a review file of bibliographic records in the target call number range on the library's lower level. The resulting list of bibliographic records was sorted by title.

A graduate student employee compared the spreadsheet with the list of bibliographic records, both of which were sorted by title, and identified matching titles. When she found a matching title, she recorded on the spreadsheet (table 1, row L) the record number in URI's Millennium library database where holdings information for each title was stored. This would allow the database record for each title to be retrieved quickly after the withdrawal project was complete to update holdings. She also entered the journal's call number into the spreadsheet, using a separate row for each segment of the call number

to facilitate sorting. The student then compared the journal holdings in the online archive with the physical holdings listed in the catalog, recording overlapped years on the spreadsheet as volumes to discard. Table 1 presents a sample record with complete data for one title recorded on the spreadsheet. The head of Acquisitions made sure that the graduate student understood that if the library held bound volumes that were published after the volumes available in the online archive, only the volumes included in the archive should be discarded.

After the student finished matching titles in the library catalog with titles available through the online archives, she deleted spreadsheet rows containing online archive titles not held by the library in print format. This resulted in a spreadsheet (see figure 1, which displays an excerpt) containing 314 titles with one or more volumes to be withdrawn, only 10 percent of the titles originally listed. The student sorted the spreadsheet by call number and created a duplicate copy in a large, bold font for printing that would allow adequate space for her to write down the number of volumes and inches discarded and

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	BACK FILES OWNED BY URI													
2	Last updated: 6/29/09													
3														
4	Call no						Title	Archive	Archive dates	URI discard(ed)	Remaining on shelf	c. record	# vols w/d	inches
5	Q	1	N	.2			Nature	Nature.com	1869-1996	1869-1996	v.385(1997)-v.443(2006)	c1050345	335	631
6	QA	1	A	.18			Acta mathematica	Springer OJA	12/28/1882 to 12/31/1996	1975-1996	v.176-179(1996/97)	c1023251	27	38
7	QA	1	A	.51			American journal of mathematics	JSTOR A&S I	1878 to 2003	1927-2001	0	c1025636	96	186
8	QA	1	A	.515			American mathematical monthly	JSTOR A&S I	1894 to 2005	1894-2005	v.113(2006)-v.115(2008)	c1026288	135	256
9	QA	1	A	.6			Annals of mathematics	JSTOR A&S I	1884 to 2003	1884-2003	v.159(2004)-v.162(2005)	c1028182	142	272
10	QA	1	A	.522			Transactions of the American Mathematical Society	JSTOR A&S I	1900 to 2003	1921-2003	v.356(2004)	c1128139	269	600
11	QA	1	C	.6			Rendiconti del Circolo Matematico di Palermo	Springer OJA	01/01/1952 to 12/31/1996	1952-1988	v.48(1924)-v.58(1934)	c1033864	26	46
12	QA	1	I	.92			Israel Journal of Mathematics	Springer OJA	03/26/1963 to 12/31/1996	1967-1996	v.104(1997)-v.156(2006)	c1043201	69	55
13	QA	1	J	.85			Journal d'Analyse Mathématique	Springer OJA	12/23/1951 to 12/31/1996	1954-1996	v.71-73(1997)	c1043225	32	55
14	QA	1	J	.9763			Journal of the American Mathematical Society	JSTOR A&S I	1988 to 2003	1998-2003	v.17(2004)	c1530938	7	15
15	QA	1	M	.426			Mathematical gazette	JSTOR A&S VII	1894 to 2003	2001-2003	0	c1694820	26	35
16	QA	1	M	.767			Mathematical Systems Theory	Springer OJA	03/01/1967 to 11/01/1996	1967-1996	0	c1011492	26	35
17	QA	1	M	.86			Mathematische Annalen	Springer OJA	03/21/1869 to 12/31/1996	1960-1995	0	c1047899	101	180
18	QA	1	M	.88			Mathematische Zeitschrift	Springer OJA	01/01/1918 to 12/31/1996	1928-1996	v.224-226(1997)	c1047929	76	95
19	QA	1	N	.35			Mathematics magazine	JSTOR A&S II	1947 to 2005	1963-2005	v.79(2006)-v.81(2008)	c1047851	42	51
20	QA	1	S	.73			Journal Of The Society For Industrial And Applied Mathematics	JSTOR A&S I	1953 to 1965	1953-1965	0	c1429048	10	20
21	QA	1	S	.73			SIAM Journal on Applied Mathematics	JSTOR A&S I	1966 to 2003	1966-2002	0	c1059294	66	143
22	QA	1	S	.2			SIAM review	JSTOR A&S I	1959 to 2003	1959-2002	0	c1058782	43	76
23	QA	1	Z	.37			Zeitschrift für Angewandte Mathematik und Physik (ZAMP)	Springer OJA	01/01/1950 to 12/31/1996	1962-1996	v.48(1997)-v.53(2002)	c1064393	36	84

Figure 1. Spreadsheet Extract Listing URI Backfiles to Be Withdrawn

which volumes, if any, would remain on the shelves. With the spreadsheet of titles to withdraw completed, the student proceeded through the stacks in call number order to mark volumes for discard. After trying various techniques, including round stickers and colored paper taped to the volumes' spines, she found that the most effective method was to mark the spine labels of the volumes with fluorescent, oil-based paint markers. The student noted the total count of volumes to be withdrawn for each title and recorded it on the spreadsheet. She measured the shelf length in inches of each run to be withdrawn to track progress toward the goal of clearing the approximately four ranges of shelf space required to accommodate the current periodicals. Volume numbers and years of any holdings to retain (or lack thereof) were recorded. The student noted any irregularities, such as missing volumes, problems, or questions about a title, and reported them to the head of Acquisitions for investigation. This entire process, from populating the spreadsheet to marking, counting, and measuring 10,805 volumes, took approximately six weeks.

Disposing of Withdrawn Volumes

Members of the HELIN Library Consortium have agreed that libraries will

offer withdrawn materials to other consortium libraries before discarding them. Though time was short, the head of Acquisitions e-mailed the spreadsheet with titles and volumes to be discarded to all academic and research libraries in Rhode Island so that they might fill gaps in their serials collections, with a one-week deadline to respond. In response to requests from 3 libraries, the graduate student coordinator shipped 236 volumes from 8 journals. Most of the volumes sent were from the journal *Nature*. The student marked these volumes with a "No Longer the Property of University of Rhode Island Library" stamp and sent them to the receiving libraries through the state-wide library delivery service.

To dispose of the withdrawn volumes, the university's recycling and solid waste coordinator made arrangements with a local recycling company to accept the bound volumes for recycling with the covers still on. This had been a concern because another library in the HELIN Consortium had been required by their recycler to remove front and back covers from the text block of every withdrawn volume, which would have been an unmanageable task given the volume of material to discard at the URI.

Although the process of transporting the volumes to be discarded from the lower-level stacks to a main

floor staging area commenced before the one-week deadline given to other Rhode Island libraries for indicating interest in the discarded volumes, the expectation was that any volumes requested could be easily located if they were stacked on the floor of the staging area in orderly piles. Given the large volume of material moved, this assumption proved to be unrealistic. Fortunately, the graduate student received only one request for two volumes after the physical moving process began. In total, the pilot project resulted in recycling 10,805 bound volumes weighing 8.65 tons and freeing 1,604 feet of shelf space.

Updating the Library Catalog and Shifting the Stacks

The final records work was updating the holdings statement in the library catalog for each discarded title. Some bound volumes remained in the library's collection for slightly more than half the 314 titles withdrawn. The graduate student was able to update holdings information for these 161 titles in her final week working on the project. All volumes for the other 153 titles had been discarded, and the holdings for these titles were deleted from the catalog and from OCLC by an Acquisitions staff member during the next month.

The second phase of the project—shifting volumes in the stacks to create space for the new current periodicals area—began immediately after the targeted volumes were removed from the shelves. The process of shifting was slow because the withdrawn volumes had been removed from 55 ranges and 5,998 shelves; each of the remaining volumes and others shelved near them had to be moved. As a result, the shift was not completed until the winter intersession of 2009–10, at which time two empty ranges, consisting of twenty-one double-sided shelving units, remained to house the current periodicals. An additional two empty ranges were removed to create a seating area near the current periodicals that accommodated three study tables. Most important, the twelve ranges on the main floor of the library that had formerly held the current periodicals were removed, allowing construction of the new learning commons to begin.

2010 Project

After gaining experience in 2009 with the pilot withdrawal project, the dean of libraries and heads of Access Services and Acquisitions were eager to extend the withdrawal process to the entire collection during the summer of 2010. The goal for this larger project was to relieve space pressure in the serial stacks on the lower level and potentially create storage space for special collections or other materials.

In 2010 the heads of Acquisitions and Access Services realized that the withdrawal process could be streamlined by gathering and disposing of materials incrementally instead of in one big push at the end of the summer. This would allow for the withdrawal of a far greater amount in the same period of time because disposal of some titles could begin while additional titles were being marked and measured. The criteria for withdrawal remained the same—volumes for which the library

had online access through licensed journal archives to a workable facsimile of the printed edition. In 2010, the head of Acquisitions compared the library's physical holdings against the same set of online journal archives that had been used in 2009 with the addition of some newly purchased backfiles. She decided that any art journals (Library of Congress Classification N) would be carefully evaluated before withdrawal, as the illustrations in the online versions might be inadequate for user needs. Art titles with significant image content would be retained in print, following the recommendations of Ithaka's *What to Withdraw?* report, which was released in September 2009 after the library's pilot project was completed.³¹ The previous year, the heads of Acquisitions and Access Services had not been attentive to this consideration and therefore had not applied the same scrutiny to image-intensive scientific and medical journals. This may have been a mistake; however, the library has not received complaints from users about lost access to print titles that were withdrawn. The quality of the scanned digital versions also was not taken into account either year in deciding what to withdraw, with one exception: the historical *Moody's Manuals*. The head of Acquisitions determined that these would be retained in print, despite the fact that the library had access to the complete online archive because some of the earlier digitized material was of poor quality and unreadable.

The graduate student returned at the end of June to manage the project; her experience in 2009 greatly increased her efficiency in 2010. She followed a similar process to the previous year's, with minor improvements. The most significant differences in 2010 were the incremental gathering and disposing of materials and improved staging for the recycler that required each volume to be handled fewer times.

By the end of August 2010, the

URI library withdrew an additional 24,130 print journal volumes from 808 titles throughout the entire collection, clearing 3,674 feet of shelf space in the serial stacks. As a result of subsequent shifting, which had consolidated the remaining volumes in the serial stacks, the dean of libraries and head of Access Services identified an additional five to six ranges of freed space for future use.

Discussion

The withdrawal project at the URI was consistent with the practice of many other college and university libraries facing space pressures. As recommended in the Ithaka report *What to Withdraw?*, the titles withdrawn are available through reliable license terms, offering ongoing access rights to all content discarded.³² As the report also recommended, the library was sensitive to the level of quality of the digitized materials, avoiding the withdrawal of image-intensive art titles and the *Moody's Manuals*, though a more extensive survey of the library's online backfiles could have been undertaken to look for other exceptions.

Because of the limited timeframe of the initial withdrawal pilot (summer 2009), library managers were unable to consider several issues as thoroughly as might otherwise have been desirable. The first was the input of faculty: the library did not consult with faculty about the withdrawal project or announce it before it took place. Besides expedience, one reason for the library's silence was an assumption based on experience that the majority of faculty prefer online access to the journal literature over print. For example, URI faculty regularly complain that the library does not have a journal when in fact the library has a print-only subscription.

Corroborating faculty preference for online journals on a national level, Ithaka's *Faculty Survey 2009*

concluded that “in the eyes of faculty, electronic versions of journals are now utterly mainstream. While print journals may continue to play a limited role for faculty with specific needs that are otherwise poorly met, digital versions are clearly the medium of choice for most faculty members, even among humanists.”³³ The Ithaka survey does acknowledge, however, that “faculty attitudes toward backfiles are somewhat more mixed.”³⁴ In 2009, only half of all faculty responded at least somewhat positively to the statement that they would be “happy” to see hard-copy collections of scholarly journals discarded and replaced entirely by electronic collections, yet fewer than 40 percent believed that it will always be crucial for their own library to maintain hard-copy collections of journals. Ithaka concluded that, “while faculty continue to value print preservation, they seem to be feeling less of a need to have immediate access to print journals locally.”³⁵ Ithaka’s findings are consistent with those of Newby, who determined that the withdrawal of print journals available through JSTOR from the University of Arizona Libraries did not in any way affect the teaching and research of most mathematics faculty and graduate students, and that the majority of survey respondents preferred electronic access to journals.³⁶ Similarly, Dubicki reported that Monmouth University Library’s success in expanding online access to journal titles after a mold problem had eliminated access to all print periodicals led to a decision to migrate the periodical collection to electronic format.³⁷ Indeed, as predicted, the URI Library did not receive any complaints after the withdrawal of print volumes.

In both the 2009 pilot and the 2010 withdrawal project, library managers did not consider alternatives to recycling the withdrawn volumes, such as storing them in an off-site repository, selling them through vendors such as B-Logistics or Periodicals Service Company, or offering them to other

libraries (beyond the HELIN Library Consortium) through back issues online discussion lists, like BACKSERV, BACKMED, or the Association for Library Collections and Technical Services Duplicates Exchange Union. The library lacks the resources to establish an off-site storage repository, and it does not have access to a shared storage facility like those run by the Research Collections and Preservation Consortium (ReCap) in New York/New Jersey, the Five Colleges in Massachusetts, or the Western Regional Storage Trust (WEST) in California, for example.

In addition, the library did not check OCLC or consider the holdings of local or regional libraries as a factor in the decision about what to discard. The library managers assumed that because all the materials to be discarded were available through JSTOR or backfiles provided by major publishers, no unique or rare material would be withdrawn and that the material discarded would be widely held in other library collections. However, as more libraries withdraw print volumes duplicated online, efforts to coordinate disposal and preserve a minimum number of print copies is becoming increasingly important. Schonfeld and Housewright, in *What to Withdraw?* explain that preservation in print of digitized material is important to fix scanning errors and improve scanning quality of digitized versions, serve as a backup in the case of digital preservation failure, ensure access in cases when the digital version is subject to restrictive licensing terms, and provide for unique scholarly needs requiring access to printed materials.³⁸ Yet without coordination between libraries, “there is a very real risk that so many copies may be discarded as to threaten the availability of certain materials in their original format.”³⁹ To this end, Ithaka has developed a Print Collections Decision-Support Tool to help librarians determine “which JSTOR-digitized journals meet the criteria

outlined in the *What to Withdraw?* report and therefore may be responsibly deaccessioned in print from any library.”⁴⁰ In the coming years, Ithaka plans to add additional, non-JSTOR titles to the tool. Similarly, Malpas at OCLC has noted that the absence of a shared infrastructure for disclosing print preservation commitments is an obstacle to libraries’ ability to cooperatively manage legacy print collections and withdraw redundant print serial holdings.⁴¹ To address this problem, OCLC is developing methods for libraries and repositories to record the condition and print archiving status of local journal holdings.⁴²

The URI library has no definite plans for additional serials withdrawal projects in the near future, though such efforts are inevitable in the coming years as more archival print content becomes available digitally and additional demands are made on the library’s space. The authors hope that by the time of URI’s next withdrawal project, the library community, along with organizations such as OCLC and Ithaka, will have further developed shared archival repositories of print journals as well as online tools for verifying the preservation status of specific titles. Checking volumes to be withdrawn against such databases will be worth the additional effort and permit contributing needed volumes to shared collections. In this way the URI library would be able to participate in the shared responsibility for preserving legacy materials and prevent the loss of unique items.

Conclusion

At the end of the two-year process, the library had withdrawn 35,729 volumes and cleared 5,277 feet of shelf space—almost one mile. The URI began with a pilot project in 2009, which focused on the need to create space for a learning commons. This project identified titles in the Q and R classification ranges

because these titles were close to the area in which the learning commons was planned and much of the library's online content was in the sciences. The library used its Serials Solutions Client Center to identify titles and years for which online backfiles were available. This list was compared to a file of bibliographic records extracted from its ILS. Ultimately, 10,805 volumes from 314 titles were withdrawn and either recycled or sent to a HELIN Library consortium member.

The success of the pilot project and the processes developed gave the URI the confidence to implement a larger project in 2010, which reviewed all serials across the collection for volumes that could be withdrawn using the same criteria—those volumes withdrawn had to be reliably available online. Lessons learned in the pilot prompted the library to be attentive to materials with significant image content, which were retained. In this larger project, an additional 24,130 print volumes from 808 titles were withdrawn. The URI heads of Acquisitions and Access Services carefully documented every step taken and developed specific strategies and policies for the withdrawal of print journal back volumes held by the library electronically. The procedures developed and the lessons learned will serve as a guide for staff at the URI and, the authors hope, for other libraries to follow when the question of space for services versus collections is raised. As collections continue their migration to digital format and academic libraries further evolve to meet user needs, the deaccessioning of print collections is likely to continue, resulting in more freed space that can be redesigned for other purposes.

References

1. Roger C. Schonfeld and Ross Housewright, *What to Withdraw? Print Collections Management in the Wake of Digitization* (New York: Ithaka S+R, 2009): 7, www.ithaka.org/ithaka-s-r/research/what-to-withdraw (accessed Feb. 12, 2011).
2. Lizanne Payne, *Library Storage Facilities and the Future of Print Collections in North America* (Dublin, Ohio: OCLC Online Computer Library Center, 2007): 16, www.oclc.org/programs/publications/reports/2007-01.pdf (accessed Sept. 7, 2011).
3. Schonfeld and Housewright, *What to Withdraw?* 7.
4. Ibid.
5. Steve O'Connor and Cathie Jilovsky, "Approaches to the Storage of Low Use and Last Copy Research Materials," *Library Collections, Acquisitions, & Technical Services* 32, no. 3/4 (2009): 121–222.
6. Matthew P. Long and Roger C. Schonfeld, *Ithaka S+R Library Survey 2010: Insights from U.S. Academic Library Directors* (2010): 6, www.ithaka.org/ithaka-s-r/research/ithaka-s-r-library-survey-2010 (accessed Feb. 10, 2011).
7. Paul N. Courant and Matthew "Buzzy" Nielson, "On the Cost of Keeping a Book," in *The Idea of Order: Transforming Research Collections for 21st Century Scholarship*, 81–105 (Washington, D.C.: Council on Library and Information Resources, 2010): 91, www.clir.org/pubs/reports/pub147/pub147.pdf (accessed Aug. 29, 2011); Michael D. Cooper, "The Costs of Providing Electronic Journal Access and Printed Copies of Journals to University Users," *Library Quarterly* 76, no. 3 (2006): 337.
8. Carol Hansen Montgomery and Donald W. King, "Comparing Library and User Related Costs of Print and Electronic Journal Collections," *D-Lib Magazine* 8, no. 10 (Oct. 2002), www.dlib.org/dlib/october02/montgomery/10montgomery.html (accessed Mar. 29, 2011).
9. Long and Schonfeld, *Ithaka S+R Library Survey 2010*.
10. O'Connor and Jilovsky, "Approaches to the Storage," 122.
11. Suzanne D. Gyeszly, Marifran Bustion, and Jane Treadwell, "Infrequently Used Serials: A Space Utilization Project," *Collection Management* 12, no. 1/2 (1990): 109–24.
12. Suzanne M. Ward and Mary C. Aagard, "The Dark Side of Collection Management: Deselecting Serials from a Research Library's Storage Facility Using WorldCat Collection Analysis," *Collection Management* 33, no. 4 (2008): 272–87.
13. Ziming Liu, "Print vs. Electronic Resources: A Study of User Perceptions, Preferences, and Use," *Information Processing & Management* 42, no. 2 (2006): 583–92; Steve Black, "Impact of Full Text on Print Journal Use at a Liberal Arts College," *Library Resources & Technical Services* 49, no. 1 (2005): 19–26; Brian E. C. Schottlaender et al., "Collection Management Strategies in a Digital Environment: A Project of the Collection Management Initiative of the University of California Libraries" (Jan. 2004), www.ucop.edu/cmi/final-report (accessed Aug. 31, 2011); K. T. L. Vaughan, "Changing Use Patterns of Print Journals in the Digital Age: Impacts of Electronic Equivalents on Print Chemistry Journal Use," *Journal of the American Society for Information Science & Technology* 54, no. 12 (2003): 1149–52; Pongracz Sennyey, Gillian D. Ellern, and Nancy Newsome, "Collection Development and a Long-Term Periodical Use Study: Methodology and Implications," *Serials Review* 28, no. 1 (2002): 38–44; Sandra L. De Groote and Josephine L. Dorsch, "Online Journals: Impact on Print Usage," *Journal of the Medical Library Association* 89, no. 4 (2001): 372–78; David H. Morse and William A. Clintworth, "Comparing Patterns of Print and Electronic Journal Use in an Academic Health Science Library," *Issues in Science & Technology Librarianship* 28 (2000), www.istl.org/00-fall/refereed.html (accessed Aug. 31, 2011).
14. Kitty McClanahan et al., "Embracing Change: Perceptions of E-Journals by Faculty Members," *Learned Publishing* 23, no. 3 (2010): 209–23; Roger C. Schonfeld and Ross Housewright, *Faculty Survey 2009: Key Strategic Insights for Libraries, Publishers, and Societies* (New York: Ithaka S+R, 2010), www.ithaka.org/ithaka-s-r/research/faculty-surveys-2000-2009/faculty-survey-2009 (accessed Aug. 29, 2011); Carol Tenopir et al.,

- "Electronic Journals and Changes in Scholarly Article Seeking and Reading Patterns," *Aslib Proceedings: New Information Perspectives* 61, no. 1 (2009): 5–32; Ian Rowlands, "Electronic Journals and User Behavior: A Review of Recent Research," *Library & Information Science Research* 29, no. 3 (2007): 369–96.
15. Charlene Sorensen, "The 5K Run Toolkit: A Quick, Painless, and Thoughtful Approach to Managing Print Journal Backruns," *Serials Review* 35, no. 4 (2009): 228.
 16. Marianne Stowell Bracke and Jim Martin, "Developing Criteria for the Withdrawal of Print Content Available Online," *Collection Building* 24, no. 2 (2005): 61
 17. Marianne Stowell Bracke and Jim Martin, "Developing Criteria for the Withdrawal of Print Content Available Online," *Collection Building* 24, no. 2 (2005): 61–64.
 18. *Ibid.*, 62.
 19. Ruth Cooper and David Norris, "To Bin or Not to Bin? Deselecting Print Back-Runs Available Electronically at Imperial College London Library," *Serials* 20, no. 3 (2007): 208–14.
 20. Schonfeld and Housewright, *What to Withdraw?*
 21. University Libraries, University of Rhode Island, Historical Serials Cancellations at the University Library, www.uri.edu/library/serials/serials_cuts/serialscuts.html (accessed Aug. 31, 2011).
 22. Andrée J. Rathemacher and Michael C. Vocino, "Perspectives on Using E-Journal Usage Statistics in a Serials Cancellation Project," in *Library Data: Empowering Practice and Persuasion*, ed. Darby Orcutt, 91–102 (Santa Barbara, Calif.: Libraries Unlimited, 2010): 92.
 23. Andrée J. Rathemacher, Technical Services Reports and Statistics, Acquisitions Unit Annual Report 2010–2011, digitalcommons.uri.edu/ts_rpts/112 (accessed Aug. 31, 2011).
 24. UKSG, Transfer Code of Practice: Version 2.0, 2008, www.uksg.org/sites/uksg.org/files/TRANSFER_Code_of_Practice_2_0.pdf (accessed Sept. 1, 2011).
 25. LOCKSS, What Is the LOCKSS Program? (2008), <http://lockss.stanford.edu/lockss/Home> (accessed Sept. 6, 2011).
 26. Michael Seadle, "Archiving in the Networked World: By the Numbers," *Library Hi Tech* 29, no. 1 (2011): 189–97.
 27. Center for Research Libraries, Portico Audit Report 2010, Executive Summary, 2010, www.crl.edu/archiving-preservation/digital-archives/certification-and-assessment-digital-repositories/portico (accessed Sept. 6, 2011).
 28. Portico, Evaluating Your Preservation Options, www.portico.org/digital-preservation/services/evaluating-your-preservation-options (accessed Sept. 6, 2011).
 29. Portico, Access to Archived Content, www.portico.org/digital-preservation/the-archive-content-access/access-to-archived-content (accessed Sept. 1, 2011).
 30. *Anglo-American Cataloguing Rules*, 2nd ed., 2002 rev., 2005 update (Chicago: American Library Association; Ottawa: Canadian Library Association; London: Chartered Institute of Library and Information Professionals, 2005).
 31. Schonfeld and Housewright, *What to Withdraw?*
 32. *Ibid.*
 33. Schonfeld and Housewright, *Faculty Survey 2009*, 11.
 34. *Ibid.*, 5.
 35. *Ibid.*, 7.
 36. Jill Newby, "An Emerging Picture of Mathematicians' Use of Electronic Resources: The Effect of Withdrawal of Older Print Volumes," *Science & Technology Libraries* 25, no. 4 (2005): 65–85.
 37. Eleonora Dubucki, "Surviving the Loss of Access to Print Periodicals," *Technical Services Quarterly* 23, no. 2 (2005): 1–17.
 38. Schonfeld and Housewright, *Faculty Survey 2009*.
 39. *Ibid.*, 8.
 40. Ithaka, Print Collections Decision-Support Tool, www.ithaka.org/ithaka-s-r/research/what-to-withdraw/print-collections-decision-support-tool (accessed Sept. 7, 2011).
 41. Constance Malpas, "Shared Print Update: Current OCLC Research in Cooperative Print Management," July 13, 2009, www.slideshare.net/RLGPrograms/shared-print-update-ala-2009-1744906 (accessed Aug. 20, 2011).
 42. OCLC, "OCLC Print Archives Pilot Project Update 2011 05-26," May 26, 2011, www.slideshare.net/oclc/print-archives-pilot-project-update-2011-0526 (accessed Sept. 7, 2011).

Notes on Operations

Integration of a Research Management System and an OAI-PMH Compatible ETDs Repository at the University of Novi Sad, Republic of Serbia

Lidija Ivanović, Dragan Ivanović, and Dušan Surla

This paper discusses the extension of the Current Research Information System (CRIS) at the University of Novi Sad, Republic of Serbia, to incorporate electronic theses and dissertations (ETDs). Data describing ETDs is entered using a web application that enables researchers to input their own data through a webpage without knowing the standards on which the system is based. The ETDs repository can exchange data with CRIS institutional repositories and Networked Digital Library of Theses and Dissertations members. In this way, the international visibility of theses and dissertations created at the University of Novi Sad is enhanced without duplicating data entry in various systems. This approach has been verified and tested on a dataset of theses and dissertations at the University of Novi Sad.

Public access to theses and dissertations via the Internet is important for the development of a knowledge-based society. A knowledge-based society relies on the knowledge of its citizens to drive entrepreneurship, innovation, and vitality of that society's economy. A knowledge-based society possesses a community of scholars, researchers, research networks, engineers, technicians, and businesses engaged in research and the production of high-technology goods and provision of services. It forms a national innovation and production system, which is integrated into international networks of knowledge production. Its communication and information technological tools make vast amounts of human knowledge easily accessible. This paper describes a test bed project at the University of Novi Sad (UNS), Republic of Serbia, which aims to improve international access to UNS research. The approach described here can inform projects at other institutions.

One approach to achieving a knowledge-based society can be through depositing electronic dissertations and theses (ETDs) in a freely accessible digital repository. Assigning appropriate metadata to ETDs can improve discoverability by increasing their visibility. Furthermore, visibility of ETDs can be increased by putting the digital object or its descriptive metadata (or both) into systems containing theses and dissertations, such as digital libraries, research management systems, institutional repositories (IRs), the Networked Digital Library of Thesis and Dissertations (NDLTD), DART-Europe E-thesis portal, Digital Repository Infrastructure for European Research (DRIVER), and others. These initiatives

Lidija Ivanović (lidija.ivanovic@pef.uns.ac.rs) is a teaching assistant, Faculty of Education, University of Novi Sad, Sombor, Republic of Serbia; **Dragan Ivanović** (chenejac@uns.ac.rs) is an assistant professor, Faculty of Technical Sciences, University of Novi Sad, Novi Sad, Republic of Serbia; and **Dušan Surla** (surla@uns.ac.rs) is a professor emeritus, Faculty of Sciences, University of Novi Sad, Novi Sad, Republic of Serbia.

Submitted October 14, 2011; tentatively accepted November 20, 2011, pending modest revision; revision submitted December 6, 2011, and accepted for publication.

This paper is part of the research project "Infrastructure for Technology Enhanced Learning in Serbia," supported by the Ministry of Education and Science of the Republic of Serbia (Project No. 47003).

and related terms are explored in detail later in this paper.

Current Research Information System (CRIS) at the University of Novi Sad (UNS), Republic of Serbia, is a Common European Research Information Format (CERIF)-compatible research management system that has been in development since 2008 at UNS.¹ CERIF is “a comprehensive metadata standard and data exchange model that can be used for a very broad range of purposes involving the management and exchange of research data” developed by the European Organization for International Research Information (www.eurocris.org).² This system has been extended at UNS with a module for storing ETDs. The primary motivation for this expansion of CRIS UNS has been to increase the international visibility of theses and dissertations by UNS scholars. Increasing the visibility of ETDs can be achieved in the following ways:

- Exchanging data between the CRIS UNS system and other research management systems according to the CERIF standard.
- Exchanging data between the CRIS UNS system and IRs in Dublin Core (DC) format according to the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH).
- Membership in the NDLTD network, i.e., exchanging data between CRIS UNS and other members of the NDLTD network in the Interoperability Metadata Standard for Electronic Theses and Dissertations (ETD-MS) format according to OAI-PMH protocol.

Another motivation for building this module was the creation of a unique system with all relevant data for scientific research activities. The system described here is a research management system integrated

with an IR. The system architecture enables easy integration with library information systems, which are based on MARC 21 format and also can hold metadata about ETDs.³

The goal of the integrated system, developed at UNS in accordance with CERIF, DC, ETD-MS, and OAI-PMH, is to avoid or reduce duplicated inputs on the two platforms and increase metadata quality, reliability, and reusability.

Literature Review

Scientific research is an important component of knowledge. Much contemporary scientific research along with its associated metadata is available in digital format via various means, such as digital libraries, research management systems, IRs, and publishers' platforms. Open access to scientific research enhances further development of science.⁴ Maximizing the visibility of scientific research is essential for scientific advancement. Visibility can be enhanced by putting research into IRs that are OAI-PMH interoperable.⁵ The OAI-PMH protocol was primarily developed as a low-barrier method for interoperability between metadata repositories and provides an interoperability framework based on metadata harvesting by defining two classes of participants: data providers that expose metadata and service providers that harvest metadata. The IR's metadata schema has a key role in increasing interoperability of the repository, i.e., maximizing visibility of theses and dissertations that are stored in the repository.⁶

A rich metadata schema enables establishing relations between various systems that contain scientific research. Collaboration between those systems has been discussed in recent years. According to Joint, sharing data between institution repositories and research management systems to avoid duplication of efforts is necessary.⁷ Joint

recommends a single point of entry to research articles regardless of whether it is through an institution repository or a research management system. He notes that three institutions (Glasgow University, Southampton University, and Kingston University) have already implemented this approach.

Krause suggests the creation of a virtual library that aims to enable users to gain integrated access to all relevant information in their special scientific field, irrespective of the location of metadata and digital form of documents.⁸ A virtual library includes a single point for creation of queries that are sent to all systems that are part of the virtual library and integrates results retrieved from the systems.

“NARCIS: The Gateway to Dutch Scientific Information,” by Dijk and colleagues, describes the National Academic Research and Collaborations Research System (NARCIS) portal (www.narcis.nl), which provides access to all scientific research information in the Netherlands.⁹ That system is an integration of the Netherlands research management system and the Digital Academic Repositories in the Netherlands (DARENET).¹⁰ Olivier describes collaboration between the research management system and the digital library at Pretoria University.¹¹

The general objective of the CRIS-IR group is “to work out an optimal solution for the interoperability of Research Management Systems on the one hand and Institutional Repositories on the other, on a European scale, taking into account all relevant aspects.”¹² The aim of the Current Research Information Systems and Open Access Repository (CRIS/OAR) interoperability project is to increase the interoperability between research management systems and open access repositories “by defining and proposing a metadata exchange format for publication information with an associate common vocabulary.”¹³

The aim of integrating systems that contain scientific research is to

maximize visibility of scientific research and avoid duplicating input of the same metadata in various systems. Many libraries worldwide store metadata in the MARC 21 format. Those libraries have electronic services that enable downloading metadata about bibliographic records related to scientific research, thus the interoperability of a repository of scientific research with those libraries is important for increasing visibility. Frequently, electronic services will enable metadata exchange in DC, but the fact that DC is not strict a standard can cause problem in metadata interoperability. However, the metadata defined by this format are a subset of metadata defined by MARC 21. When searching databases of scientific research, users can better express their information needs if the research is described in a richer set of metadata. A CERIF-compatible data model based on the MARC 21 format makes CRIS systems interoperable with library information systems.¹⁴ In this model some CERIF data are stored in the MARC 21 format data model. As noted earlier, CERIF defines a data model that enables interoperability between CRIS implementations; MARC 21 is a standard for storing data for library systems. That model includes all entities and attributes of the CERIF data model and preserves the existing references between the CERIF data model entities. Furthermore, that model enables input of multilingual data prescribed by the CERIF standard. The MARC 21 format is rich in metadata and enables more detailed description of entities in CRIS systems. A MARC 21 record can store all metadata prescribed by DC and ETD-MS format.¹⁵ An information system based on the CERIF-compatible data model can exchange data with other systems using XML documents (which have XML schemas prescribed by CERIF standard) and can exchange data with LIS based on the MARC 21 formats and with IRs based on DC or ETD-MS format.

Context: CRIS UNS

CRIS UNS is a CERIF-compatible research management system under development since 2008 at UNS. The first phase of CRIS UNS development was the implementation of a system for entering metadata about published scientific research including papers published in journals, papers from scientific conferences, monographs, and papers published in monographs.

CRIS UNS is built on the CERIF-compatible data model based on the MARC 21 format described in the previous section. CRIS UNS was implemented as a web application based on “best-of-breed” open-source components written in Java.¹⁶ The system has three-tier architecture. Three-tier architecture contains a client tier (the presentation logic, including simple control and user input validation), middle tier (the business processes logic and the data access), and data tier (the data server provides the business data).¹⁷ Any web browser supporting HTML 4 and JavaScript can be used for application access.

The server side of the system is executed within the Apache Tomcat (<http://tomcat.apache.org>) application server. Apache Tomcat is an open-source software implementation of the Java Servlet and JavaServer Pages (JSP) technologies. A servlet is “used to extend the capabilities of servers that host applications accessed via a request-response programming model. Although servlets can respond to any type of request, they are commonly used to extend the applications hosted by Web servers.”¹⁸ JSP technology “provides a simplified, fast way to create dynamic web content.”¹⁹

The presentation tier is developed using the JavaServer Faces (JSF) development environment (www.jcp.org/en/jsr/detail?id=252) and RichFaces (www.jboss.org/richfaces) library. JSF technology simplifies building user interfaces for JavaServer applications, and RichFaces is a library

of Ajax-enabled components for JSF. The Apache Lucene (<http://lucene.apache.org>) library is an open-source information retrieval library written in Java and is used for indexing and searching text contents. Text indexing and query processing include a Cyrillic to Latin transliteration algorithm. All index entries are stored as Latin text, thus enabling the use of both scripts in searching. On the other hand, database contents hold information as it was entered by the user, preserving the correct script. Cyrillic to Latin transliteration is unambiguous. This means that every character of Cyrillic has an appropriate character in the Latin (or Roman) alphabet—and that every word written using Cyrillic can be unambiguously translated to a word using Latin characters. The MySQL (www.mysql.com) database management system is used for data preservation.⁴ The system data model and architecture enable easy integration of the system with LIS and interoperability with other CERIF-compatible national CRIS systems.

Published results from the system are available to anonymous users via the Internet. Moreover, the system is in accordance with the CERIF standard and meets requirements prescribed by the Republic of Serbia Ministry of Science and Technological Development in the field of scientific research results evaluation. Therefore the system data model is extended with necessary entities.²⁰ The system is implemented as a web application that enables authors to input metadata about their own research without knowledge of the CERIF standard and the MARC 21 format.

Research Method

The first step in this project was analysis of various systems that contain metadata about theses and dissertations. The following are international initiatives:

- NDLTD (www.ndltd.org) is an international organization that aims to create a worldwide network of ETDs. Each digital repository that is a network member has to enable metadata exchange in the ETD-MS format (developed by DNLTD) in accordance with OAI-PMH.²¹
- DART-Europe E-Thesis Portal (www.dart-europe.eu/contributors/how.php) aims to collect details of the open access research theses stored in Europe's digital repositories (doctoral and master theses). It collects metadata in DC using OAI-PMH.
- DRIVER (www.driver-community.eu) is an international organization co-funded by the European Commission with the goal of creating a network of freely accessible digital repositories with content across all academic disciplines. Each digital repository that is a network member has to enable metadata exchange in DC in accordance with the OAI-PMH protocol.

In addition, many academic and research institutions and research communities may implement and manage the following approaches to collecting, preserving, accessing, and disseminating research:

- IRs are online systems that collect, preserve, and disseminate the intellectual output in digital form of an institution. IRs may use open-source software, such as DSpace (www.dspace.org) and Fedora (<http://fedora-commons.org>), or hosted, proprietary software, such as Digital Commons (<http://digitalcommons.bepress.com>) and SimpleDL (www.simpledl.com). Many IRs support the exchange of data in DC via OAI-PMH.
- A CRIS is a database of other

information system for storing data on current research (e.g., data about institutions, researchers, research projects, equipment, published results, etc.). The European Union encourages the development of national research management systems in accordance with the CERIF standard.²² CERIF-compatible research management systems are called CRIS. Due to specific local or national requirements, CRIS systems are built on different modifications (or extensions) of CERIF data model.²²

- A library information system (LIS) is a software system for acquiring, cataloging, and circulating library holdings. LIS are built on various bibliographic standards; most are based on MARC 21 formats.

Across these systems, different standards and protocols—CERIF, OAI-PMH, DC, ETD-MS, and MARC—enable interoperability.

After analysis was completed, a comprehensive metadata set was defined to develop a repository that is compatible with all previously mentioned systems. Then the authors extended the CRIS UNS data model to store all metadata about ETDs as well as the ETDs as digital objects. Finally, the authors expanded CRIS UNS with a module for storing ETDs along with associated metadata. An object-oriented method was used for the module modeling. Object-oriented modeling creates models using object-oriented diagrams (class diagram, sequence diagram, etc.), which is the starting point for implementing a system using object-oriented programming language. The modeling was carried out using the Sybase PowerDesigner tool that supports OMG's Unified Modeling Language (UML) 2.0 (www.omg.org/spec/UML/2.0). The module model can be obtained by contacting the authors. The implementation

was realized using “best-of-breed” open-source components written in Java. After the authors developed the module, it was verified and tested on EDTs by researchers in the Faculty of Sciences, UNS. After migration of the existing dataset containing ETDs along with associated metadata from the DIGLIB UNS system to the CRIS UNS system, UNS researchers verified the migrated data about their theses and dissertations and supplied additional data; the CRIS UNS metadata set is richer than the DIGLIB UNS metadata set. These steps are covered in detail in the following sections.

Data Model Definition

After analysis of various systems that contain metadata about theses and dissertations (NDLTD, DART-Europe E-thesis portal, DRIVER, IRs, CRISs, LIS, DIGLIB UNS), a comprehensive metadata set was defined to create a repository that is compatible with various ETDs systems. DIGLIB UNS is the IR at UNS and contains theses and dissertations from the university. This system allows input of metadata about theses and dissertations as required by the UNS rule book, which defines key words that all the university theses and dissertations must have. Table 1 presents the list of metadata elements selected for CRIS UNS and indicates their presence or absence in CERIF, DC, and ETD-MS. This metadata set unites metadata describing EDTs, drawing from all standards used in the DIGLIB UNS (diglib.uns.ac.rs).²³ The set of metadata about EDTs adopted for the CRIS-UNS system unites the metadata sets prescribed by CERIF, DC, and ETD-MS format, extended by metadata that are used in DIGLIB UNS to meet the needs of the UNS.

Data Model Extension

As already stated, the CRIS UNS data model holds data about scientific

research in MARC 21 format. MARC 21 records are stored using an attribute of the MARC 21 record entity that holds a string representing a MARC 21 record serialized according to the International Standards Organization (ISO) 2709 standard, which sets out the format for information exchange.²⁴ Upon serializing the MARC 21 record in an ISO 2709 string, the record is stored in the database and its contents are indexed using the Apache Lucene information retrieval library. MARC 21 records can be classified using the entity MARC 21 Record_Class: master thesis, PhD dissertation, and so on. Also, that entity can be used for the definition of the scientific field and scientific discipline of the research, such as mathematics, computer sciences, biology, information systems, and artificial intelligence. Using that entity, records can be divided in sets and the OAI-PMH “ListRecords” requirement, which mandates the ability to download only records that belong to a defined set, can be met.

The earlier CRIS UNS data model was extended by adding four attributes to the MARC 21 Record entity. These added attributes are *creator*, *dateOfCreation*, *modifier*, and *dateOfLastModification*. Date of creation and date of the last modification are necessary to meet all requirements prescribed by the OAI-PMH protocol; the OAI-PMH ListRecords request must be able to download only records that are processed in a certain period.

Furthermore, the previous CRIS UNS data model is extended by adding the *File_Storage* entity that is intended to hold data related to the digital form of theses or dissertations. Each instance of the *File_Storage* entity is connected to an instance of the *MARC 21 Record* entity that holds bibliographic metadata about the thesis or dissertations. The *uploader* attribute holds the e-mail address of the user who uploaded the digital content.

Table 1. Metadata about Theses and Dissertations Adopted for the CRIS-UNS System

CRIS-UNS	CERIF	Dublin Core	ETD-MS
author	+	+	+
advisor	-	-	+
chair	-	-	+
committee member	-	-	+
title	+	+	+
alternative title	-	-	+
subtitle	+	-	-
keywords	+	+	+
abstract	+	+	+
extended abstract	-	-	-
note	+	-	+
language	-	+	+
ISBN	+	-	-
physical description	+	-	-
UDC	-	-	-
publisher	+	+	+
publication date	+	+	+
record type	-	+	+
content format	-	+	+
URI	+	+	+
access rights	-	+	+
thesis type	+	-	-
name of author degree after defense	-	-	+
level of education	-	-	+
scientific field	-	-	+
scientific discipline	-	-	-
accepted by competent scientific institution on	-	-	-
institution	+	+	+
defended on	-	-	-
holding data	-	-	-

The attributes *fileName*, *mime*, and *length* store metadata describing the digital content that is stored in a folder of the file system of the CRIS UNS server. The folder is not directly accessible through the Internet, but digital contents can be downloaded using a Java servlet. In this way, access to digital content is controlled, i.e., the Java servlet controls who can download digital content.

Table 2 shows mappings of adopted metadata about theses and

dissertations shown in table 1 to the extended CRIS UNS data model. The first column holds names of metadata and the second column holds location in MARC 21 bibliographic record. The first three characters of a MARC 21 record present a field code; the next two characters present the first and the second indicator, respectively; and the last character presents a subfield code. The character “#” indicates that indicator is not defined. The last column shows some notes

Table 2. Mappings of Metadata to Data Model

Metadata	MARC 21	Note
author	1001# a	All data about <i>authors/advisors/chair/committee members</i> are stored in a MARC 21 authority record; relation of thesis or dissertation with the authority record is established using the subfield 0 of data field 100/700 of MARC 21 bibliographic record. The subfield e of data field 100/700 holds relationship type: <i>author, mentor, thesis/dissertation defend board chair, thesis/dissertation defend board member</i> .
advisor	7001# a	
chair	7001# a	
committee member	7001# a	
title	24500 a	Translations of those metadata are stored in the field 880 as described in "CERIF Compatible Data Model Based on MARC 21 Format." ^o
alternative title	2460# a	
subtitle	24500 b	
keywords	653 ## a	
abstract	5203# a	
extended abstract	520 ## a	
note	500 ## a	
language	008	Language is stored using three letters from 35th to 37th character positions of the control field 008. Character positions starts from 0.
ISBN	020 ## a	
physical description	300 ##	Physical description is stored using subfields of the data field 300.
UDC	080 ## a	
publisher	260 ## b	The metadata holds a value <i>author's reprint</i> or name of the appropriate institution.
publication date	260 ## c	Year of publication are additionally stored in character positions 7–10 of the control field 008.
record type	LDR	Record type is stored in 6th character position of the leader of MARC 21 record. Character positions starts from 0.
content format	856 ## q	The metadata holds one of the following values: pdf, doc, docx, odt.
URL	856 ## u	The subfield holds the URL of a thesis or dissertation in digital form.
access rights	540 ## a	
thesis type	655 #4 a	Also stored using the <i>MARC 21Record_Class</i> entity of the CRIS UNS data model.
name of author degree after defense	502 ## a	Name of degree is prescribed at the institution where author defends his or her thesis or dissertation. For example: master of electrical engineering, doctor of technical sciences, etc.
level of education	502 ## b	The element holds level of education: bachelor, master, doctoral, post-doctoral, etc.
scientific field	65024 a	Also stored using the <i>MARC 21Record_Class</i> entity of the CRIS UNS data model.
scientific discipline	65014 a	
accepted by competent scientific institution on	502 ## g	The metadata are stored in the subfield g in the following format: 502 ## \$gTheme of thesis or dissertation accepted on date.
institution	502 ## c	That subfield holds the name and address of the institution. All data about institutions are stored in a MARC 21 authority record, the relation of thesis or dissertation with the authority record is realized using entity <i>MARC 21Record_MARC 21Record</i> .
defended on	502 ## g	The metadata are stored in the subfield g in the following format: 502 ## \$gThesis or dissertation defended on date.
holding data	852 ## a	

^o Dragan Ivanović, Dušla Surla, and Zora Konjović, "CERIF Compatible Data Model Based on MARC 21 Format," *Electronic Library* 29, no. 1 (2011): 52–70.

about metadata and methods of their storing.

CRIS UNS Extension with ETDs

The next phase of the development of CRIS UNS was to extend it with a subsystem that enables uploading ETDs and inputting their metadata. The authors identified the basic

information requirements of this subsystem as the following:

- Uploading ETDs. The system supports pdf, doc, docx, and odt file formats. Furthermore, the system has to backup files and provides long-time preservation of those files.
- Migrating existing data from

DIGLIB UNS to the system.

- Entering all metadata about EDTs that that CERIF standard prescribes and all metadata that are necessary for exchange in accordance with the OAI-PMH protocol within ND LTD. User interface has to be as simple as possible so that it can be used by users without

Figure 1. Form for Input of Metadata

the knowledge of standards and protocols.

- Exchanging metadata about EDTs with other CRIS systems. In this way, researchers from European countries using national CRIS systems can find EDTs from the CRIS UNS system.
- Exchanging metadata about EDTs in accordance with the OAI-PMH protocol. In this way, theses and dissertations from CRIS UNS can be visible through a various IRs as well as through web applications for searching the NDLTD Union Catalogue: SCIRUS ETD Search (www.ndltd.org/service/providers/scirus-etd-search), VTLIS Visualizer (www.vtls.com/products/visualizer), etc.

The system architecture was extended with a file server component that manages storing and

downloading files from the server's file system. This component also is used to preserve digital contents of other scientific research, such as papers published in journals, monographs, and papers published in conference proceedings. This digital content is not freely accessible and access to those digital materials is controlled through the Java servlet. The file server component will be integrated with an open-source solution for long-term file preservation such as Lots of Copies Keep Stuff Safe (LOCKSS) (www.lockss.org). The file server component also extracts textual content from uploaded files using open-source Apache Tika library (<http://tika.apache.org>). After extraction, text goes through a Cyrillic to Latin transliteration algorithm and then is indexed using the Apache Lucene library. Query processing also includes a Cyrillic to Latin transliteration algorithm and

thus enables the use of both scripts (Cyrillic and Latin) in searching.

Furthermore, the system user interface is extended with user forms for uploading ETDs and entering metadata about ETDs. All textual user interface elements are stored in external files that facilitate the translation of the user interface to other languages. The first step is uploading the digital content, which uses a dialog that prompts the user to find the file to be added from his or her own computer. After uploading the digital content, the next step is input of the metadata listed in table 1. The form for input of metadata is shown in figure 1. Translations of multilingual metadata can be entered using this form and invoking (clicking on) the boxes to the right (e.g., Title translations, Subtitle translations, and so on).

All data about authors, advisors, chair, and committee members are stored in a MARC 21 authority record. The relation of a thesis or dissertation with the authority record is established using the subfield "0" of the MARC 21 record field 100/700. The subfield "0" contains the control number of the authority record that stores data about a researcher (thesis author, mentor, and so on). Subfield "e" of the field 100/700 holds the relationship type between a thesis and researcher (relation is established by subfield "0"), e.g., author, mentor, thesis or dissertation defense board chair, or thesis or dissertation defense board member. This approach to establishing relationships allows various reports to be generated, such as

- thesis and dissertations in which a researcher has been a mentor, thesis defense board chair, or thesis defense board member; and
- thesis and dissertations in which researchers from some departments have been a mentor, thesis defense board chair, or thesis defense board member.

Because some metadata are multi-lingual, information retrieval measures (precision, recall, and F-measure) are improved, i.e., visibility of ETDs are increased. Furthermore, visibility of ETDs is improved by using fuzzy search that is enabled through Apache Lucene library. Fuzzy search retrieves all theses and dissertations that meet a set of criteria that define similarity. For example, similarity criteria for two strings (string from a query and string from a thesis or dissertation title stored in the CRIS UNS database) are defined as follows:

- Each word in one string does not differ by more than two letters from a word in another string.
- If one string contains more than five words, the previous criterion is satisfied for at least 80 percent of the words.
- It is case insensitive and Cyrillic-Latin script insensitive (i.e., lower case and upper case are equal, as well as Cyrillic and Latin scripts).

Data Verification

This application was verified and tested on data about theses and dissertations of researchers employed at Faculty of Sciences, UNS. After migration of the existing dataset containing ETDs along with associated metadata from the DIGLIB UNS system to the CRIS UNS system, researchers from the University of Novi Sad verified and supplied migrated data about their theses and dissertations. The Faculty of Sciences employs more than 300 researchers and has written approximately 900 master theses and 500 PhD dissertations through 2011. The test set included metadata about all 1,400 theses and dissertations. Hard-copies of all 1,400 theses and dissertations can be found in the faculty library. In time of this writing, 200 of them also can be found in digital form. Transforming

the remaining 1,200 from hard-copy to ETDs by scanning is in progress. Researchers did not complain about the migrated data or the user interface. Adding theses and dissertations from the additional fourteen UNS faculties is also in progress. After this process is finished, an additional effort to consolidate data will be necessary; this will include such activities as removing duplicated items and consolidating scientific fields and disciplines.

Conclusion

This paper describes the implementation of a digital repository of EDTs within the CRIS UNS system. Metadata about theses and dissertations are stored in the MARC 21 bibliographic format. The implementation is based on open-source components. The system architecture allows an easy transition to other bibliographic standards and easy integration with LIS based on the adopted bibliographic standard.

The system can exchange ETDs metadata with other CRIS systems, IRs, the NDLTD network members, and LIS. Interoperability with previous stated systems maximizes visibility of ETDs from the repository without duplicate entry of ETDs metadata in various systems. Metadata are entered once, but metadata are stored in various systems across the Internet. High international visibility of theses and dissertations of researchers from University of Novi Sad enhances the further development of science and raises public awareness of UNS research.

The system for inputting of ETDs has been verified and tested on a dataset containing EDTs by researchers at Faculty of Sciences, UNS. The addition of ETDs from additional fourteen UNS faculties is in progress. After this process is finished, further effort to consolidate data will be necessary; this will include such activities as removing duplicated items and consolidating scientific fields and disciplines.

After this step, web services for data exchange will be made available for public access. Finally, an audit will be performed to assess whether the visibility of scientific research from UMS has increased after this repository implementation.

References and Notes

1. EuroCRIS, CERIF 2008—Final Release (1.2), www.eurocris.org/Index.php?page=CERIF2008&xt=1 (accessed Nov. 16, 2011). CERIF 1.3 Release was available for preview until early December 2011, www.eurocris.org/Index.php?page=CERIF-1.3&xt=1 (accessed Nov. 16, 2011).
2. CERIFY, What is the CERIFY Project? What is CERIF?, <http://cerify.ukoln.ac.uk/node/196> (accessed Nov. 14, 2011).
3. Library of Congress, MARC Standards, MARC 21 Formats, www.loc.gov/marc/marcdocz.html (accessed Nov. 16, 2011).
4. Steve Lawrence, "Free Online Availability Substantially Increases a Paper's Impact," *Nature* 411 (May 2001): 477, www.nature.com/nature/debates/e-access/Articles/lawrence (accessed Nov. 14, 2011); Stevan Harnad and Tim Brody, "Comparing the Impact of Open Access (OA) vs. Non-OA Articles in the Same Journals," *D-Lib Magazine* 10, no. 6 (2004), www.dlib.org/dlib/june04/harnad/06harnad.html (accessed Aug. 22, 2011); Kristin Antelman, "Do Open-Access Articles Have a Greater Research Impact?" *College & Research Libraries* 65, no. 5 (Sept. 2004): 372–82.; Kent Anderson et al., "Publishing Online-Only Peer-Reviewed Biomedical Literature: Three Years of Citation, Author Perception, and Usage Experience," *Journal of Electronic Publishing* 6, no. 3 (Mar. 2001), <http://quod.lib.umich.edu/cgi/t/text/text-idx?c=jep;view=text;rgn=main;idno=3336451.0006.303> (accessed Aug. 22, 2011); Gunther Eysenbach, "Citation Advantage of Open Access Articles," *PLoS Biology* 4, no. 5 (May 2006): 692–98, www.plosbiology.org/article/info:doi/10.1371/journal

- .pbio.0040157 (accessed Aug. 22, 2011).
5. Mohammad Hanief Bhat, "Interoperability of Open Access Repositories in Computer Science and IT—An Evaluation," *Library Hi Tech* 28, no. 1 (2010): 107–18.
 6. Eun G. Park and Marc Richard, "Metadata Assessment in E-Theses and Dissertations of Canadian Institutional Repositories," *The Electronic Library* 29, no. 3 (2011): 394–407; Sevim McCutcheon et al., "Morphing Metadata: Maximizing Access to Electronic Theses and Dissertations," *Library Hi Tech* 26, no. 1 (2008): 41–57.
 7. Nicholas Joint, "Current Research Information Systems, Open Access Repositories and Libraries: ANTAE-US," *Library Review* 57, no. 8 (2008): 570–75.
 8. Jürgen Krause, "Current Research Information As Part of Digital Libraries and the Heterogeneity Problem Integrated Searches in the Context of Databases with Different Content Analyses," in *Gaining Insight from Research Information: 6th International Conference on Current Research Information Systems*, ed. Wolfgang Adamczak and Annemarie Nase, 21–31 (Kassel, Germany: Kassel University Press, 2002), www.uni-kassel.de/hrz/db4/extern/dbupress/publik/abstract_en.php?978-3-933146-84-7 (accessed Nov. 14, 2011).
 9. Elly Dijk et al., "NARCIS: The Gateway to Dutch Scientific Information," in *Digital Spectrum: Integrating Technology and Culture: Proceedings of the 10th International Conference on Electronic Publishing held in Bansko, June 14–16, 2006*, ed. Bob Martens and Milena Dobрева, 49–57 (Sofia: FOI-COMMERCE, 2006), http://elpub.scix.net/data/works/att/233_elpub2006.content.pdf (accessed Nov. 15, 2011).
 10. Astrid van Wesenbeeck, "Digital Academic Repositories in the Netherlands: Built with the DARE Program (2003–2006)" (presentation, Valencia, Spain, June 20, 2006), <http://cde.uv.es/documents/2007-VANWESBEECK.pdf> (accessed Nov. 14, 2011).
 11. Elsabé Olivier, "Open Scholarship and Research Reporting in Tandem: Creating More Value" (presentation, The African Digital Scholarship & Curation Conference, May 12–14, 2009, Pretoria, South Africa), www.ais.up.ac.za/digi/docs/olivier_paper.pdf (accessed Aug. 22, 2011).
 12. EuroCRIS, Operation Work Plan for the CRIS-IR Task Group, www.eurocris.org/Index.php?page=CRIS-IR_workplan&t=1 (accessed Nov. 14, 2011).
 13. KE: Knowledge Exchange, CRIS/OAR Project, www.knowledge-exchange.info/Default.aspx?ID=340 (accessed Nov. 15, 2011).
 14. Dragan Ivanović, Dušla Surla, and Zora Konjović, "CERIF Compatible Data Model Based on MARC 21 Format," *Electronic Library* 29, no. 1 (2011): 52–70.
 15. Lidija Ivanović, Dragan Ivanović, and Dušlan Surla, "A Data Model of Theses and Dissertations Compatible with CERIF, Dublin Core, and ETD-MS," *Online Information Review* (forthcoming).
 16. Dragan Ivanović et al., "A CERIF-Compatible Research Management System Based on the MARC 21 Format," *Program: Electronic Library & Information Systems* 44, no. 3 (2010): 229–51; Gordana Milosavljević et al., "Automated Construction of the User Interface for a CERIF-Compliant Research Management System," *Electronic Library* 29, no. 5 (2011): 565–88; Aleksandar Kovačević et al., "Automatic Extraction of Metadata from Scientific Publications for CRIS Systems," *Program: Electronic Library & Information Systems* 45, no. 4 (2011): 376–96.
 17. Ariel Ortiz Ramires, "Three-Tier Architecture," *Linux Journal* 75 (July 1, 2000), www.linuxjournal.com/article/3508 (accessed Aug. 22, 2011).
 18. The H2EE Tutorial, What Is a Servlet? http://java.sun.com/j2ee/tutorial/1_3-fcs/doc/Servlets2.html#75087 (accessed Nov. 19, 2011).
 19. Oracle, JavaServer Pages Technology, www.oracle.com/technetwork/java/javasee/jsp/index.html (accessed Nov. 19, 2011).
 20. Dragan Ivanović, Dušlan Surla and Miloš Racković, "A CERIF Data Model Extension for Evaluation and Quantitative Expression of Scientific Research Results," *Scientometrics* 86, no. 1 (2011): 155–72.
 21. Networked Digital Library of Theses and Dissertations, ETD-MS: An Interoperability Metadata Standard for Electronic Theses and Dissertations, version 1.00, rev. 2, www.ndltd.org/standards/metadata/etd-ms-v1.00-rev2.html (accessed Nov. 16, 2011); Open Archives Initiative, The Open Archives Initiative Protocol for Metadata Harvesting, Protocol version 2.0 of 2002-06014, www.openarchives.org/OAI/openarchivesprotocol.html (accessed Nov. 16, 2011).
 22. EuroCRIS, CERIF 2008—Final Release 1.2.
 23. Dušlan Surla et al., "Overview of Implementation of the Networked Digital Library of Theses and Dissertations," *Infoteka* 5, no. 1–2 (2004): 75–86.
 24. International Standards Organization, *International Standard: ISO 2709, Information and Documentation—Format for Information Exchange = Information et Documentation—Format por l'échange d'information*, 4th ed. (Geneva, Switzerland: ISO Copyright Office, 2008).

Book Reviews

Norm Medeiros

University Libraries and Digital Learning Environments. Edited by Penny Dale, Jill Beard, and Matt Holland. Surrey, UK; Burlington, Vt.: Ashgate, 2011. 278p. \$114.95 hardcover (ISBN 978-0-7546-7957-8).

The purpose of this book is to discuss and evaluate the development of digital learning environments within the university library. Rather than perform a retrospective analysis of academic libraries that have implemented the digital learning environment (DLE), this book aims to provide a current snapshot and to provoke debate about the present state and immediate future of the DLE in higher education. The editors and chapter authors are current or former residents of the United Kingdom, and therefore bring a distinctive perspective to a global subject.

The editors, Penny Dale, Jill Beard, and Matt Holland, provide an insightful introduction, outlining the purpose, structure, and content of each of the sixteen chapters that constitute the book. Sue McKnight's initial chapter, "Here Today and Here Tomorrow," aptly sets the stage; each successive chapter explores the complexity of this multifaceted subject, illustrating the flexibility librarians must adopt to succeed while working within virtual learning environments. Peter Godwin's chapter on social media presents a broad and balanced look at the changes in computing habits of society at large, how college students are using Web 2.0 tools, and how librarians are implementing such tools into public services and workflows. Not surprisingly, a chapter on information literacy and the digital environment's effect on

teaching strategies and learning habits also is included. Other topics include scholarly communication, balancing collections containing both digital and analog materials, and the methods and challenges of performance assessment within the digital environment. Institutional repositories are afforded two chapters. In the first, Alma Swan asserts institutional repositories (IR) are poised to take a significant role in the support and expansion of e-research. Swan discusses how IRs are being used, the benefits of their use, and where they are headed in the future. Within this context she touches on some technical matters, such as digital preservation. The second chapter on IR addresses its role within the larger organization, how to market its use, and how to increase community buy-in for the long term.

This book also probes less commonly addressed issues, such as services for the Further Education Institution (FEI) versus the Higher Education Institution (HEI) in the Virtual Learning Environment (VLE), and online support for English as a second language (ESL) for study abroad students. For readers outside the United Kingdom not familiar with the terms FEI and HEI, they may be compared to the community college or vocational school, and the university or college, respectively. Under a variety of circumstances, the FEI often serves the nontraditional student by allowing them to pursue both short-term vocational diplomas and the traditional college degree. In this chapter, the author explores the challenges small FEI libraries face in filling information needs of their population, and how they are using consortium efforts with

HEIs, and experimentation in VLEs, to overcome the shortfall in budget and material resources. Though readers may not recognize these terms, the theme of this chapter rings familiar. In his chapter, Frank Trew investigates another important but often-neglected topic: how U.K. libraries are providing support and resources in VLE for their ESL study abroad students. Strengthening such services should be of utmost importance to librarians who provide public service and administrators wishing to attract a larger international applicant pool.

This book achieves its goal of providing a broad snapshot of university libraries as they exist today, and their adaptations soon to the ubiquitous digital learning environment—a remarkable achievement given the speed at which the subject is evolving and expanding, and the length of time it takes to publish a book. The book's strength and weakness lay within the British-centric view from which this volume is approached. Some international readers may be alienated by the discussion of HEIs and FEIs; however, many readers will find the discussion valuable because it provides opportunities to contemplate matters from a perspective that previously may not have been considered. Nevertheless, a chapter that covered digital preservation in more depth, and some examination of cloud computing, would have been useful additions. This book will be of interest to those involved at all stages of creating and supporting service points in the virtual learning environment.—*Natalie Bulick (natalie.bulick@indstate.edu), Indiana State University, Terre Haute, Indiana*

No Shelf Required: E-Books in Libraries. Edited by Sue Polanka. Chicago: American Library Association, 2010. 182p. \$65.00 paper (ISBN 978-0-8389-1054-2).

Sue Polanka (Wright State University, and author of a blog by the same name as this book) has collated nine chapters on current aspects of e-books. That each chapter has a different author is at once disconcerting and helpful; read straight through, the reader first learns the history of e-books, then is introduced to e-book issues in school media centers, public libraries, and academic libraries. The final chapters synthesize e-book issues touched on in previous chapters. Each chapter in the book varies in tone, depth of coverage, and complexity. Although the first chapter on the history of e-books provides an overview for the general reader as well as library worker, the final chapters on acquisition, preservation, and e-book standards will appeal only to those with experience managing electronic resources. Each chapter can stand alone. The downside to this construct is that important issues are not fully discussed in one place. Digital rights management (DRM), for example, is mentioned in five chapters, licensing in six. E-book readers are covered in five chapters, although the Sony Reader and Amazon Kindle also are mentioned separately in additional chapters.

The first chapter, "E-books on the Internet" by James Galbraith, offers a thorough and highly readable history of electronic books. Galbraith reminds readers that despite the recent "so-called crass commercialization of e-books," (2) prevalence of e-book readers, and availability of e-book acquisition models, e-book collections have been with us for decades due to "a relatively small but influential e-book community" (2). Galbraith is referring to Project Gutenberg, which created the first digitized document in 1971 (*The Declaration of*

Independence). Project Gutenberg was followed a decade later by Tufts University's Perseus Digital Library, one of the first single-subject e-book collections. This chapter describes the creation of other collections, recounts the technical challenges in making digitized documents widely available in the early days of the Internet, and provides an even and succinct overview of the Google Books controversy.

The next two chapters address e-books in school libraries, with discussion of e-books for student learning and e-books in the school media center. With the premise that e-books as multimedia tools will engage students in reading and are thus advantageous, both chapters recommend e-book subscription services such as Big Universe, BookFlix, International Children's Digital Library, Tumblebooks, and Tumblereadables for new readers, and free e-book collections such as Google Books, Project Gutenberg, Internet Public Library, and Bibliomania for K-12 readers. The chapters touch on use of e-reference books in the media center and adoption of e-textbooks, which have the potential to enhance learning through provision of dictionaries, pronunciation guides, read-aloud capability, links to relevant sites, embedded multimedia, and built-in learning assessments. Both chapters acknowledge issues with adding e-books to school media collections, and despite the overlap in content and loose platitudes, these chapters are helpful for school media specialists who wish to add e-books to their collections.

Chapter 4 covers e-books in public libraries and considers audiobooks as public libraries' first foray into downloadable monographs. One wonders why the author felt the need to include a paragraph on the advantages of e-books (no shelf required, for one) since these are fairly well established. Three of the primary book vendors for the public library market are profiled: Overdrive, Ingram Digital, and

Netlibrary (EBSCO's recent purchase of the last is noted), with e-reader compatibility charts, delivery methods, licensing options, and collection tools for each. Formats beyond PDF such as EPUB, Open eBook, and Mobipocket are explained succinctly. Another strength of this chapter is the contributor's exploration of how public libraries have collected and applied usage statistics for e-books. New York Public Library, for example, measures e-book use as a "virtual branch" (70) in terms of circulation reports. Acquisition for public libraries focuses on title-by-title selection, not package purchases that may be more common in academic libraries.

Chapter 5, "The Academic Library E-Book?" is a handy and highly practical chapter of interest to academic librarians, whether they are involved in e-book collections or not. It provides a thorough and much-needed discussion on the flexibility and options available in acquiring e-books, including subscription models, title-by-title selection, patron (or demand-driven) acquisition, short-term loans, subject and publisher packages, and pay-per-view. Negotiating points and issues inherent with each acquisition model are included. The chapter acknowledges the range of e-book genres of interest to academic libraries: popular, scholarly, audio, and e-reference books, as well as monographic series, scholarly bibliographies, and freely available classic texts. This chapter also describes University of Texas at Austin's experience with e-books and provides a case study of Penn State's e-reader project. A range of marketing and discovery advice is included, such as adding e-books to the catalog, utilizing vendor-sponsored webinars, embedding chapter links in course management software, creating search widgets, and highlighting e-books in information literacy courses. This chapter also acknowledges the issue of librarian buy-in, which is not addressed elsewhere in the book, and

perhaps most helpfully, identifies common issues in e-book use in academic libraries. Some of these issues include what to do when a patron requests a print version of an e-book, Americans with Disabilities Act compliance, DRM issues, and interlibrary loan. Two areas not discussed that would be appropriate here are e-book weeding and the entry of university presses into the e-book market.

Chapters 2 through 5 focus on e-books by type of library; chapters 7–9 address the nuts and bolts of e-book issues, such as acquisition, use, preservation, and standards. These chapters expand on content mentioned in less detail in the early chapters, and those staff involved in any aspect of e-book acquisition will benefit. Vendors make e-book acquisition easy; much more complicated are the myriad platforms, restrictions, access models, and device compatibility issues that librarians must understand to make e-books accessible to patrons. Polanka and contributor Emilie Delquíe cite Petway's barrier of thirty: "There are nearly thirty devices on the market (and counting), and there are thirty formats for e-book content, many of which are proprietary" (136). They follow with an alphabet soup of e-book-related acronyms (XML, ILL, DOI, ISBN EBUB, DRM, and SERU) that should be required study for all librarians.

Highly readable, this book is primer for libraries entering the e-book market, a cautionary tale for those who are wading in, and a bird's eye view for those whom e-books are business as usual.—Cathy Goodwin (*cgoodwin@coastal.edu*), *Coastal Carolina University, Conway, South Carolina*

Metadata for Digital Collections: A How-to-Do-It Manual. By Steven J. Miller. New York: Neal-Schuman, 2011. 343p. \$78.00 softcover (ISBN 978-1-5557-0746-0). How-to-Do-It Manuals.

The rapidly developing digital library environment continues to

present many challenges, not only to those who are just beginning to dabble in digital library initiatives, but also to those with experience. *Metadata for Digital Collections* is an excellent addition to the growing literature addressing this topic. The author, Steven Miller, is an experienced cataloger and cataloging department manager. This experience, combined with his position teaching courses in metadata, cataloging, and information architecture at the University of Wisconsin-Milwaukee School of Information Studies, makes him ideally suited to address the development and application of metadata to digital collections.

Metadata for Digital Collections is organized into eleven chapters that cover all aspects of creating metadata in a digital library setting. The first chapter begins with the basics: defining metadata, describing types of metadata applied to digital collections, and introducing the reader to metadata standards. Several definitions of metadata are provided and, taken together, they illustrate for the reader how diverse our understanding of metadata can be. Chapter 2 discusses the foundations of resource description, and because of its ubiquity, introduces the Dublin Core (DC) metadata element set. Although prior cataloging knowledge would help the reader put resource description into the context of library databases, prior cataloging experience is not necessary to understand the introductory concepts presented in this chapter.

Chapters 3 and 4 continue the approach of addressing metadata concepts through the application of the DC standard. Chapter 3 explores how resources are identified and how responsibility for creation, contribution, and publication is assigned. Each topic, such as title, identifier, dates, etc., is addressed in a general section, followed by a section devoted to the same topic as defined by DC. Chapter 4 addresses how subject, form, and genre are handled for digital

materials. Again, each topic, such as type, genre, and format, is discussed broadly, followed by a description of how that topic is handled in DC.

Chapter 5 makes the case that controlled vocabularies are a crucial aspect of resource description. The role of controlled vocabulary in disambiguation and establishing hierarchical relationships is explained. Many types of controlled vocabulary are discussed, including lists, synonym rings, authority files, taxonomies, and thesauri. The concept of creating a specialized vocabulary is not neglected—references to the American National Standards Institute/National Information Standards Organization (ANSI/NISO) guidelines on creating controlled vocabularies are included.

Metadata created according to a variety of standards, such as DC, Visual Resource Association (VRA) Core, or Metadata Object Description Schema (MODS), can be encoded for storage and transmission using XML. Chapter 6 is devoted to a basic introduction of how the XML encoding standard can be used effectively to store and transmit data. Chapters 7 and 8 address the MODS and VRA Core categories. These chapters are filled with useful examples of MODS and VRA Core records encoded in XML.

Chapter 9 addresses metadata interoperability, sharing, and quality—critical issues in ensuring the long-term viability of metadata created for digital resources. The chapter concludes with suggestions for ways to improve metadata interoperability and quality, such as using DC or another standard element set; including an appropriate amount of contextual information and access points; entering data values that are machine-readable and linkable; distinguishing administrative metadata from descriptive; and documenting local practices. These suggestions are standard practices in traditional cataloging policies and procedures, but have yet to be

perhaps most helpfully, identifies common issues in e-book use in academic libraries. Some of these issues include what to do when a patron requests a print version of an e-book, Americans with Disabilities Act compliance, DRM issues, and interlibrary loan. Two areas not discussed that would be appropriate here are e-book weeding and the entry of university presses into the e-book market.

Chapters 2 through 5 focus on e-books by type of library; chapters 7–9 address the nuts and bolts of e-book issues, such as acquisition, use, preservation, and standards. These chapters expand on content mentioned in less detail in the early chapters, and those staff involved in any aspect of e-book acquisition will benefit. Vendors make e-book acquisition easy; much more complicated are the myriad platforms, restrictions, access models, and device compatibility issues that librarians must understand to make e-books accessible to patrons. Polanka and contributor Emilie Delquíé cite Petway's barrier of thirty: "There are nearly thirty devices on the market (and counting), and there are thirty formats for e-book content, many of which are proprietary" (136). They follow with an alphabet soup of e-book-related acronyms (XML, ILL, DOI, ISBN EBUB, DRM, and SERU) that should be required study for all librarians.

Highly readable, this book is primer for libraries entering the e-book market, a cautionary tale for those who are wading in, and a bird's eye view for those whom e-books are business as usual.—Cathy Goodwin (*cgoodwin@coastal.edu*), *Coastal Carolina University, Conway, South Carolina*

Metadata for Digital Collections: A How-to-Do-It Manual. By Steven J. Miller. New York: Neal-Schuman, 2011. 343p. \$78.00 softcover (ISBN 978-1-5557-0746-0). How-to-Do-It Manuals.

The rapidly developing digital library environment continues to

present many challenges, not only to those who are just beginning to dabble in digital library initiatives, but also to those with experience. *Metadata for Digital Collections* is an excellent addition to the growing literature addressing this topic. The author, Steven Miller, is an experienced cataloger and cataloging department manager. This experience, combined with his position teaching courses in metadata, cataloging, and information architecture at the University of Wisconsin-Milwaukee School of Information Studies, makes him ideally suited to address the development and application of metadata to digital collections.

Metadata for Digital Collections is organized into eleven chapters that cover all aspects of creating metadata in a digital library setting. The first chapter begins with the basics: defining metadata, describing types of metadata applied to digital collections, and introducing the reader to metadata standards. Several definitions of metadata are provided and, taken together, they illustrate for the reader how diverse our understanding of metadata can be. Chapter 2 discusses the foundations of resource description, and because of its ubiquity, introduces the Dublin Core (DC) metadata element set. Although prior cataloging knowledge would help the reader put resource description into the context of library databases, prior cataloging experience is not necessary to understand the introductory concepts presented in this chapter.

Chapters 3 and 4 continue the approach of addressing metadata concepts through the application of the DC standard. Chapter 3 explores how resources are identified and how responsibility for creation, contribution, and publication is assigned. Each topic, such as title, identifier, dates, etc., is addressed in a general section, followed by a section devoted to the same topic as defined by DC. Chapter 4 addresses how subject, form, and genre are handled for digital

materials. Again, each topic, such as type, genre, and format, is discussed broadly, followed by a description of how that topic is handled in DC.

Chapter 5 makes the case that controlled vocabularies are a crucial aspect of resource description. The role of controlled vocabulary in disambiguation and establishing hierarchical relationships is explained. Many types of controlled vocabulary are discussed, including lists, synonym rings, authority files, taxonomies, and thesauri. The concept of creating a specialized vocabulary is not neglected—references to the American National Standards Institute/National Information Standards Organization (ANSI/NISO) guidelines on creating controlled vocabularies are included.

Metadata created according to a variety of standards, such as DC, Visual Resource Association (VRA) Core, or Metadata Object Description Schema (MODS), can be encoded for storage and transmission using XML. Chapter 6 is devoted to a basic introduction of how the XML encoding standard can be used effectively to store and transmit data. Chapters 7 and 8 address the MODS and VRA Core categories. These chapters are filled with useful examples of MODS and VRA Core records encoded in XML.

Chapter 9 addresses metadata interoperability, sharing, and quality—critical issues in ensuring the long-term viability of metadata created for digital resources. The chapter concludes with suggestions for ways to improve metadata interoperability and quality, such as using DC or another standard element set; including an appropriate amount of contextual information and access points; entering data values that are machine-readable and linkable; distinguishing administrative metadata from descriptive; and documenting local practices. These suggestions are standard practices in traditional cataloging policies and procedures, but have yet to be

carried over to the digital realm in a consistent way.

Despite the many metadata schemes already developed and standardized, there are still instances when it is desirable to create a specialized metadata scheme. The steps in developing a metadata scheme are thoroughly outlined in chapter 10. Examples of such schemes also are provided, including DC metadata documentation from the Collaborative Digitization Program, OhioLINK, and Indiana Memory, and MODS documentation from the Digital Library Federation's Aquifer project. These examples are invaluable resources for anyone trying to develop their own documentation.

The final chapter is devoted to a discussion of linked data and the Semantic Web. Although as yet one sees few practical applications of linked data, monitoring concepts and development in this emerging field is important.

In addition to the extensive references at the end of each chapter, a robust bibliography and an index appear at the end of the book. The book is generously illustrated with more than one hundred figures and tables. Sidebars illustrating concepts, clarifying definitions, and providing examples are present throughout. The book is clearly written and accessible to students learning about metadata for the first time, but also rich enough to be useful for the experienced practitioner.

Metadata for Digital Collections is well suited for both practicing professionals and students. It provides an excellent grounding in all aspects of applying metadata in a digital library setting and would be a useful addition to any professional library. It also would be appropriate for use in a library or information science course for students who are learning about the organization of information. —Rebecca L. Mugridge (rlm31@psu.edu), Pennsylvania State University, University Park, Pennsylvania

Graphic Novels and Comics in Libraries and Archives: Essays on Readers, Research, History, and Cataloging. Edited by Robert G. Weiner. Jefferson, N.C.: McFarland, 2010. 276p. \$45.00 paper (ISBN 978-0-7864-4302-4).

This new collection, which details the current state of comics in libraries, deserves attention. Its chapters cover diverse ground, its writers exude earnestness and enthusiasm, and its research is seminal yet exploratory. Amid the proliferation of introductory and reader's advisory guides and works on comics for literacy and instruction, the present volume is one of just a handful to address comics and librarianship more broadly, with twenty-nine individually authored chapters covering most facets of library work. Editor Robert G. Weiner (Texas Tech University) is no newcomer to the subject matter, having written considerably on it in addition to having worked with comics in both public and academic library settings.

Readers will wish to explore these essays selectively, choosing those that align with their own contexts and interests. *Graphic Novels and Comics in Libraries and Archives* is certainly not a book to read cover to cover, as its redundancies will appear overwhelming (including nearly two dozen only slightly different takes on the terminology of "graphic novels" versus "comics"). In addition, some chapters seem aimed at novices whereas others at librarians with considerable background knowledge.

Part I offers chapters on the history of comics in libraries. Although the first chapter's basic primer and idiosyncratic annotated list offers nothing new, the second chapter provides a full history of Manga in Japanese libraries, an overview for which English readers will be thankful. Amy Kiste Nyberg's "How Librarians Learned to Love the Graphic Novel" employs a literature review construct to neatly summarize the history of U.S. librarian attitudes

and activities with regard to comics.

The next four parts discuss comics and graphic novels in particular types of libraries, with seven chapters focused on academic libraries, three on public libraries, two on school libraries (elementary and high school levels), and one on Pennsylvania-related comics in the archival collection of the State Library of Pennsylvania. Some common themes resonate throughout many of these essays, especially issues related to selection, cataloging, and methods of physical placement of graphic novels. Readers often will find differing solutions to similar problems in this volume, affirming the value of local context in one's own decision-making. Yet readers also will note the emergence of some best practices, a profitable thread for subsequent research in this field.

The much larger section on academic library contexts opens with an overview of graphic novels as popular culture collections, offering tips for faculty buy-in, selection, funding, cataloging, and preservation. More focused chapters discuss course reserves, selection, public relations, and special collections. Especially in these chapters, the reader finds the common apologetic tone regarding comics in libraries alongside numerous real-life examples of the use of comics within the disciplines. Gwen Evans's contribution, "The Library after Dark," gives extended treatment of Bowling Green State University's student-librarian partnership to create a comic book to promote library collections and services, one of several such innovative productions that has excited the library science blogosphere recently. Many readers will turn first to the two chapters on the comic art collection at the Michigan State University (MSU) Libraries, which feature its bibliographer, Randy Scott. In the first article, Scott provides an overview of the collection; the second article is an interview with Scott. MSU's collection is arguably the most significant in

carried over to the digital realm in a consistent way.

Despite the many metadata schemes already developed and standardized, there are still instances when it is desirable to create a specialized metadata scheme. The steps in developing a metadata scheme are thoroughly outlined in chapter 10. Examples of such schemes also are provided, including DC metadata documentation from the Collaborative Digitization Program, OhioLINK, and Indiana Memory, and MODS documentation from the Digital Library Federation's Aquifer project. These examples are invaluable resources for anyone trying to develop their own documentation.

The final chapter is devoted to a discussion of linked data and the Semantic Web. Although as yet one sees few practical applications of linked data, monitoring concepts and development in this emerging field is important.

In addition to the extensive references at the end of each chapter, a robust bibliography and an index appear at the end of the book. The book is generously illustrated with more than one hundred figures and tables. Sidebars illustrating concepts, clarifying definitions, and providing examples are present throughout. The book is clearly written and accessible to students learning about metadata for the first time, but also rich enough to be useful for the experienced practitioner.

Metadata for Digital Collections is well suited for both practicing professionals and students. It provides an excellent grounding in all aspects of applying metadata in a digital library setting and would be a useful addition to any professional library. It also would be appropriate for use in a library or information science course for students who are learning about the organization of information. —Rebecca L. Mugridge (rlm31@psu.edu), Pennsylvania State University, University Park, Pennsylvania

Graphic Novels and Comics in Libraries and Archives: Essays on Readers, Research, History, and Cataloging. Edited by Robert G. Weiner. Jefferson, N.C.: McFarland, 2010. 276p. \$45.00 paper (ISBN 978-0-7864-4302-4).

This new collection, which details the current state of comics in libraries, deserves attention. Its chapters cover diverse ground, its writers exude earnestness and enthusiasm, and its research is seminal yet exploratory. Amid the proliferation of introductory and reader's advisory guides and works on comics for literacy and instruction, the present volume is one of just a handful to address comics and librarianship more broadly, with twenty-nine individually authored chapters covering most facets of library work. Editor Robert G. Weiner (Texas Tech University) is no newcomer to the subject matter, having written considerably on it in addition to having worked with comics in both public and academic library settings.

Readers will wish to explore these essays selectively, choosing those that align with their own contexts and interests. *Graphic Novels and Comics in Libraries and Archives* is certainly not a book to read cover to cover, as its redundancies will appear overwhelming (including nearly two dozen only slightly different takes on the terminology of "graphic novels" versus "comics"). In addition, some chapters seem aimed at novices whereas others at librarians with considerable background knowledge.

Part I offers chapters on the history of comics in libraries. Although the first chapter's basic primer and idiosyncratic annotated list offers nothing new, the second chapter provides a full history of Manga in Japanese libraries, an overview for which English readers will be thankful. Amy Kiste Nyberg's "How Librarians Learned to Love the Graphic Novel" employs a literature review construct to neatly summarize the history of U.S. librarian attitudes

and activities with regard to comics.

The next four parts discuss comics and graphic novels in particular types of libraries, with seven chapters focused on academic libraries, three on public libraries, two on school libraries (elementary and high school levels), and one on Pennsylvania-related comics in the archival collection of the State Library of Pennsylvania. Some common themes resonate throughout many of these essays, especially issues related to selection, cataloging, and methods of physical placement of graphic novels. Readers often will find differing solutions to similar problems in this volume, affirming the value of local context in one's own decision-making. Yet readers also will note the emergence of some best practices, a profitable thread for subsequent research in this field.

The much larger section on academic library contexts opens with an overview of graphic novels as popular culture collections, offering tips for faculty buy-in, selection, funding, cataloging, and preservation. More focused chapters discuss course reserves, selection, public relations, and special collections. Especially in these chapters, the reader finds the common apologetic tone regarding comics in libraries alongside numerous real-life examples of the use of comics within the disciplines. Gwen Evans's contribution, "The Library after Dark," gives extended treatment of Bowling Green State University's student-librarian partnership to create a comic book to promote library collections and services, one of several such innovative productions that has excited the library science blogosphere recently. Many readers will turn first to the two chapters on the comic art collection at the Michigan State University (MSU) Libraries, which feature its bibliographer, Randy Scott. In the first article, Scott provides an overview of the collection; the second article is an interview with Scott. MSU's collection is arguably the most significant in

academic libraries, and Scott adroitly situates it within the broader national context of comic research collections.

The weakest section of this volume is part 7, "Nomenclature and Aesthetics." As already mentioned, the nomenclature discussions herein quickly wear thin, offering neither fresh theoretical insights nor strong support for particular practices. Randy Scott, in fact, smartly dismisses the issue thus: "Graphic novel" is the new pretentious word for "comic book" (128). The final chapter of this section should be ignored by all; its authors, a pair of philosophy graduate students, seem not only unversed in library practice but unaware of relevant and comics-related work occurring within their own discipline.

Part 8 collects two brief articles seemingly as an afterthought. The first article on meta-comics, despite its straw-man approach, explains a potentially confusing aspect of comics culture to the uninitiated. The second article covers the highlights of free online comics ("webcomics") and basic library approaches to managing such content.

A section on cataloging describes the issues and provides practical potential solutions. The last section of the book presents basic studies of Canadian and Association of Research Libraries holdings, the latter perhaps providing a starting point for future, longitudinal study of the penetration of graphic novels into research library collections. Finally, Weiner's three-page afterword recounts his personal history with comics in libraries. This autobiographical account is reflective of broader trends in the field and would be most beneficial if read first.

Full indexing will aid librarians in finding chapters relevant to their interests. Chapters include references, but not suggested readings. Those seeking to track down particular graphic novels discussed in the book may run into difficulties because of frequently misspelled or otherwise

incorrect titles. The present volume, read judiciously, will prove quite useful.—*Darby Orcutt* (*darby_orcutt@ncsu.edu*), *North Carolina State University, Raleigh, North Carolina*

Conversations with Catalogers in the 21st Century. Edited by Elaine R. Sanchez. Santa Barbara, Calif.: Libraries Unlimited, 2011. 283p. \$50.00 (ISBN: 978-1-5988-4702-4; eISBN: 978-1-5988-4703-1). Libraries Unlimited Library Management Collection.

This collection of "conversations" is initiated by Michael Gorman's foreword in which he affirms the importance of catalogs and catalogers. "High levels of precision and recall, the two ways in which we judge any information retrieval system, are dependent on controlled vocabularies and national and international standards—they cannot be obtained by other systems not involving human intervention" (viii).

The book is divided into four sections. The first section is on the *Anglo-American Cataloguing Rules, 2nd edition (AACR2)* and *Resource Description and Access (RDA)*.¹ The authors of the three chapters in this section raise substantive concerns about the practicality of *RDA*. The authors wrote their chapters, however, before *RDA* was published and before the announcement of the U.S. *RDA* test results. Nevertheless, some of the authors' concerns also were expressed by the testing librarians, resulting in the decision by the U.S. national libraries not to implement *RDA* until its instructions are rewritten in "clear, unambiguous, plain English."² Also in this section Elaine Sanchez, editor of the book, reports on her extensive survey of 459 respondents (91 percent from U.S. libraries) about their views of the new cataloging code, the training that would be needed to implement it, its cost and cost effectiveness, and whether *AACR2* should be maintained in parallel with *RDA*. In this chapter I first noticed my one criticism

of the book: some of the figures have print so small they are difficult to read.

Although I found useful ideas in all five chapters of the next section, "Visions: New Ideas for Bibliographic Control and Catalogs," I will limit my discussion to just three contributions. Ed Jones makes a strong case for the importance of identifiers in library catalogs. He uses the definition of an identifier found in the 2009 *Statement of International Cataloguing Principles*: "A number, code, word, phrase, logo, device, etc., that is uniquely associated with an entity, and serves to differentiate that entity from other entities within the domain in which the identifier is assigned."³ The earliest identifiers and the ones all catalogers will be familiar with are Library of Congress Control Numbers and International Standard Book Numbers. Jones states that while identifiers are now used to satisfy two of the user tasks outlined in the Functional Requirements of Bibliographic Records—that is, find and identify—"with the growing integration of OPACs into the World Wide Web, they would soon satisfy all four (find, identify, select, and obtain)" (99).

Also in the second section is a chapter by Bernhard Eversberg in which he outlines a new format to replace MARC for both bibliographic and authority data. He has implemented the format using the Allegro software package, and he provides a link to sample records in a demonstration database. I particularly noticed his use of single quote marks around initial articles in titles to remove the article from indexing. With MARC, initial articles can be coded to be skipped only when they begin a field; Eversberg's method allows initial articles to be skipped when they begin subfields as well.

Martha Yee contributes to this section by looking into the future and seeing the benefits of a single shared catalog. Neither WorldCat nor the Semantic Web meets the eight specifications that she outlines. Her

academic libraries, and Scott adroitly situates it within the broader national context of comic research collections.

The weakest section of this volume is part 7, "Nomenclature and Aesthetics." As already mentioned, the nomenclature discussions herein quickly wear thin, offering neither fresh theoretical insights nor strong support for particular practices. Randy Scott, in fact, smartly dismisses the issue thus: "Graphic novel" is the new pretentious word for "comic book" (128). The final chapter of this section should be ignored by all; its authors, a pair of philosophy graduate students, seem not only unversed in library practice but unaware of relevant and comics-related work occurring within their own discipline.

Part 8 collects two brief articles seemingly as an afterthought. The first article on meta-comics, despite its straw-man approach, explains a potentially confusing aspect of comics culture to the uninitiated. The second article covers the highlights of free online comics ("webcomics") and basic library approaches to managing such content.

A section on cataloging describes the issues and provides practical potential solutions. The last section of the book presents basic studies of Canadian and Association of Research Libraries holdings, the latter perhaps providing a starting point for future, longitudinal study of the penetration of graphic novels into research library collections. Finally, Weiner's three-page afterword recounts his personal history with comics in libraries. This autobiographical account is reflective of broader trends in the field and would be most beneficial if read first.

Full indexing will aid librarians in finding chapters relevant to their interests. Chapters include references, but not suggested readings. Those seeking to track down particular graphic novels discussed in the book may run into difficulties because of frequently misspelled or otherwise

incorrect titles. The present volume, read judiciously, will prove quite useful.—*Darby Orcutt* (*darby_orcutt@ncsu.edu*), *North Carolina State University, Raleigh, North Carolina*

Conversations with Catalogers in the 21st Century. Edited by Elaine R. Sanchez. Santa Barbara, Calif.: Libraries Unlimited, 2011. 283p. \$50.00 (ISBN: 978-1-5988-4702-4; eISBN: 978-1-5988-4703-1). Libraries Unlimited Library Management Collection.

This collection of "conversations" is initiated by Michael Gorman's foreword in which he affirms the importance of catalogs and catalogers. "High levels of precision and recall, the two ways in which we judge any information retrieval system, are dependent on controlled vocabularies and national and international standards—they cannot be obtained by other systems not involving human intervention" (viii).

The book is divided into four sections. The first section is on the *Anglo-American Cataloguing Rules, 2nd edition (AACR2)* and *Resource Description and Access (RDA)*.¹ The authors of the three chapters in this section raise substantive concerns about the practicality of *RDA*. The authors wrote their chapters, however, before *RDA* was published and before the announcement of the U.S. *RDA* test results. Nevertheless, some of the authors' concerns also were expressed by the testing librarians, resulting in the decision by the U.S. national libraries not to implement *RDA* until its instructions are rewritten in "clear, unambiguous, plain English."² Also in this section Elaine Sanchez, editor of the book, reports on her extensive survey of 459 respondents (91 percent from U.S. libraries) about their views of the new cataloging code, the training that would be needed to implement it, its cost and cost effectiveness, and whether *AACR2* should be maintained in parallel with *RDA*. In this chapter I first noticed my one criticism

of the book: some of the figures have print so small they are difficult to read.

Although I found useful ideas in all five chapters of the next section, "Visions: New Ideas for Bibliographic Control and Catalogs," I will limit my discussion to just three contributions. Ed Jones makes a strong case for the importance of identifiers in library catalogs. He uses the definition of an identifier found in the 2009 *Statement of International Cataloguing Principles*: "A number, code, word, phrase, logo, device, etc., that is uniquely associated with an entity, and serves to differentiate that entity from other entities within the domain in which the identifier is assigned."³ The earliest identifiers and the ones all catalogers will be familiar with are Library of Congress Control Numbers and International Standard Book Numbers. Jones states that while identifiers are now used to satisfy two of the user tasks outlined in the Functional Requirements of Bibliographic Records—that is, find and identify—"with the growing integration of OPACs into the World Wide Web, they would soon satisfy all four (find, identify, select, and obtain)" (99).

Also in the second section is a chapter by Bernhard Eversberg in which he outlines a new format to replace MARC for both bibliographic and authority data. He has implemented the format using the Allegro software package, and he provides a link to sample records in a demonstration database. I particularly noticed his use of single quote marks around initial articles in titles to remove the article from indexing. With MARC, initial articles can be coded to be skipped only when they begin a field; Eversberg's method allows initial articles to be skipped when they begin subfields as well.

Martha Yee contributes to this section by looking into the future and seeing the benefits of a single shared catalog. Neither WorldCat nor the Semantic Web meets the eight specifications that she outlines. Her

specifications are complex, and I will not attempt to summarize them here, but I hope that the designers of next-generation catalogs will consult this chapter. The programmers for OCLC WorldCat also could find new directions for development.

It is inevitable in a collection like this that the various authors do not always agree with each other. For instance, Eversberg states that “display ought to be a matter of programming; logically, it ought not to be mixed up with internal formatting” (117). Yee on the other hand worries that “everything we call cataloging (effective indexing and effective displays) is pushed out of RDA and into ‘application’ or ‘implementation’” (129). Yee cautions the authors of *RDA* not to forget about the user “since, from the catalog user’s point of view, cataloging is display design” (129). Both authors do, however, agree on the importance of indexing. While Eversberg does not see indexing as part of the format definition, he does criticize *RDA* for continuing “in the AACR tradition of not bothering with filing” (112).

“The Cataloging World in Transition” is the title of the third section of the book. I thoroughly enjoyed the insights presented by all six chapter authors. I especially liked John Myers’s observation about standards. They are not “masters to intimidate us, but are instead our servants in the pursuit of our larger ideals” (179). I also found the chapter by Christine Schwartz to be particularly relevant to the stage of transition experienced by my own department. Over the past five years we have taken on the new technically challenging roles of designing and implementing batch load processes, and doing quality control by creating data sets and manipulating them in batches. I found her “Metadata Skill Set” very apt. We need in particular staff who “have traditional cataloging skills as well as database . . . skills” (184).

The final section of the book is “Cataloging and Metadata Librarians: Research, Education, Training and Recruitment.” Janet Swan Hill wonders how the profession can attract new librarians to the specialization of cataloging, and two other authors discuss their ideas for education and training. The book editor closes this section with two bibliographies. An afterword by Sheila Intner and Susan Lazinger does an excellent job of summing up the contents, although I dislike their metaphorical characterization of the book’s mood as “fear of flying” (269). I did not perceive fear of the future or fear of change in these chapters, but rather a reminder that we must stay focused on providing accurate and standardized metadata so our users can find, identify, select, and obtain the information they want. I found many stimulating ideas in this book, and I heartily recommend it to other twenty-first-century catalogers.—*Sue Wartzok (wartzoks@fiu.edu), Florida International University, Miami*

References

1. *Anglo-American Cataloguing Rules*, 2nd ed., 2002 rev., 2005 update (Chicago: American Library Association; Ottawa: Canadian Library Association; London: Chartered Institute of Library and Information Professionals, 2002); *RDA: Resource Description & Access* (Chicago: ALA, 2010).
2. U.S. RDA Test Coordinating Committee, *Report and Recommendations of the U.S. RDA Test Coordinating Committee: Executive Summary* (Washington, D.C.: Library of Congress, 2011), www.loc.gov/bibliographic-future/rda/rda-execsummary-public-13june11.pdf (accessed Nov. 3, 2011): 3.
3. International Federation of Library Associations, *Statement of International Cataloguing Principles* (The Hague, Netherlands: IFLA, 2009), www.ifla.org/files/cataloguing/icp/icp_2009-en.pdf (accessed Nov. 3, 2011): 11.

Open Access: What You Need to Know. By Walt Crawford. Chicago: American Library Association, 2011. 76p. \$45.00 softcover, \$36.00 e-book, \$53.00 print/e-book bundle (ISBN 978-0-8389-1106-8).

Crawford begins this ALA Special Report by defining open access (OA) literature as “available online to be read for free by anyone, anytime, anywhere” (1). His goals are to outline some of the issues surrounding this seemingly simple concept, suggest ways librarians can advocate for open access, and guide librarians to resources for learning more and staying up-to-date on OA.

In the first chapter, Crawford discusses why all librarians should care about OA. OA focuses on research literature, which Crawford acknowledges is of particular concern to academic and some special librarians, who serve the researchers and practitioners who use this literature. He points out that not all researchers and scholars are affiliated with institutions, so public libraries also may be called on to supply patrons with research literature. Public libraries also receive requests for research literature from patrons with special interests. While school children are unlikely to seek research literature, school librarians as citizens should care about OA. Crawford outlines ethical and pragmatic arguments for OA, including eliminating wealth as a barrier to information, providing citizens with access to research funded by taxes, improving communication between researchers, and disseminating research findings broadly for testing and validation. OA also has the potential to alleviate the problems caused by rapid price increases for science, technology, engineering, and medical (STEM) journals, which consume a disproportionate and growing share of academic and special library budgets, although Crawford acknowledges that alleviating cost pressures requires bargaining by an influential

specifications are complex, and I will not attempt to summarize them here, but I hope that the designers of next-generation catalogs will consult this chapter. The programmers for OCLC WorldCat also could find new directions for development.

It is inevitable in a collection like this that the various authors do not always agree with each other. For instance, Eversberg states that “display ought to be a matter of programming; logically, it ought not to be mixed up with internal formatting” (117). Yee on the other hand worries that “everything we call cataloging (effective indexing and effective displays) is pushed out of RDA and into ‘application’ or ‘implementation’” (129). Yee cautions the authors of *RDA* not to forget about the user “since, from the catalog user’s point of view, cataloging is display design” (129). Both authors do, however, agree on the importance of indexing. While Eversberg does not see indexing as part of the format definition, he does criticize *RDA* for continuing “in the AACR tradition of not bothering with filing” (112).

“The Cataloging World in Transition” is the title of the third section of the book. I thoroughly enjoyed the insights presented by all six chapter authors. I especially liked John Myers’s observation about standards. They are not “masters to intimidate us, but are instead our servants in the pursuit of our larger ideals” (179). I also found the chapter by Christine Schwartz to be particularly relevant to the stage of transition experienced by my own department. Over the past five years we have taken on the new technically challenging roles of designing and implementing batch load processes, and doing quality control by creating data sets and manipulating them in batches. I found her “Metadata Skill Set” very apt. We need in particular staff who “have traditional cataloging skills as well as database . . . skills” (184).

The final section of the book is “Cataloging and Metadata Librarians: Research, Education, Training and Recruitment.” Janet Swan Hill wonders how the profession can attract new librarians to the specialization of cataloging, and two other authors discuss their ideas for education and training. The book editor closes this section with two bibliographies. An afterword by Sheila Intner and Susan Lazinger does an excellent job of summing up the contents, although I dislike their metaphorical characterization of the book’s mood as “fear of flying” (269). I did not perceive fear of the future or fear of change in these chapters, but rather a reminder that we must stay focused on providing accurate and standardized metadata so our users can find, identify, select, and obtain the information they want. I found many stimulating ideas in this book, and I heartily recommend it to other twenty-first-century catalogers.—*Sue Wartzok (wartzoks@fiu.edu), Florida International University, Miami*

References

1. *Anglo-American Cataloguing Rules*, 2nd ed., 2002 rev., 2005 update (Chicago: American Library Association; Ottawa: Canadian Library Association; London: Chartered Institute of Library and Information Professionals, 2002); *RDA: Resource Description & Access* (Chicago: ALA, 2010).
2. U.S. RDA Test Coordinating Committee, *Report and Recommendations of the U.S. RDA Test Coordinating Committee: Executive Summary* (Washington, D.C.: Library of Congress, 2011), www.loc.gov/bibliographic-future/rda/rda-execsummary-public-13june11.pdf (accessed Nov. 3, 2011): 3.
3. International Federation of Library Associations, *Statement of International Cataloguing Principles* (The Hague, Netherlands: IFLA, 2009), www.ifla.org/files/cataloguing/icp/icp_2009-en.pdf (accessed Nov. 3, 2011): 11.

Open Access: What You Need to Know. By Walt Crawford. Chicago: American Library Association, 2011. 76p. \$45.00 softcover, \$36.00 e-book, \$53.00 print/e-book bundle (ISBN 978-0-8389-1106-8).

Crawford begins this ALA Special Report by defining open access (OA) literature as “available online to be read for free by anyone, anytime, anywhere” (1). His goals are to outline some of the issues surrounding this seemingly simple concept, suggest ways librarians can advocate for open access, and guide librarians to resources for learning more and staying up-to-date on OA.

In the first chapter, Crawford discusses why all librarians should care about OA. OA focuses on research literature, which Crawford acknowledges is of particular concern to academic and some special librarians, who serve the researchers and practitioners who use this literature. He points out that not all researchers and scholars are affiliated with institutions, so public libraries also may be called on to supply patrons with research literature. Public libraries also receive requests for research literature from patrons with special interests. While school children are unlikely to seek research literature, school librarians as citizens should care about OA. Crawford outlines ethical and pragmatic arguments for OA, including eliminating wealth as a barrier to information, providing citizens with access to research funded by taxes, improving communication between researchers, and disseminating research findings broadly for testing and validation. OA also has the potential to alleviate the problems caused by rapid price increases for science, technology, engineering, and medical (STEM) journals, which consume a disproportionate and growing share of academic and special library budgets, although Crawford acknowledges that alleviating cost pressures requires bargaining by an influential

and determined library system or the availability of a critical mass of OA journals. Crawford's brief description of the 2010 negotiation between the University of California and Nature Publishing Group is a powerful example of the difficulties libraries have in bargaining with STEM publishers. I have one minor quibble with this chapter: I wish Crawford's excellent explanation of how copyright is transferred in scholarly journal publishing had appeared much earlier in the chapter.

In the second chapter, Crawford discusses key documents, defines terms, and provides some history to help readers understand how the concept of OA has changed and may continue to change. He describes and quotes from three documents: the Budapest Open Access Initiative, the Bethesda Statement on Open Access Publishing, and the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities, which established the term "open access" in 2002. Crawford points out that OA, as defined in these documents, includes more than simply reading research literature: the authors of these documents envision additional distribution, data mining, and other uses as components of open access. Crawford also introduces the terms Green OA, Gold OA, Gratis OA, and Libre OA, provides brief definitions of each, then discusses the state of OA in 2010 in relation to each of these terms. Throughout the discussion, he deftly explains related terms, such as preprint, which may be unfamiliar to some readers. He also introduces some of the major figures in the OA movement.

In the third chapter, Crawford discusses issues in need of further study. Issues related to OA journals include what qualifies as an OA journal, the extent to which OA journals can compete with subscription journals for the best articles, viable funding methods for stable OA journals,

journal transitions from subscription to OA, problems with OA publishing scams, and even whether alternatives to journals, such as blogs, are feasible. Other OA issues include determining the costs of running an OA repository, the question of whether institutional or subject repositories are preferable, concerns with preservation and long-term access, the extent to which libraries provide access points for OA journals and articles in repositories, and the point at which libraries will begin seeing savings from OA. Crawford points out that as of 2010 about 20 percent of recent research articles are OA—he suggests that 75 percent OA may be the tipping point where libraries either abandon Big Deals or can use the availability of OA journals as a bargaining point for lower prices. He comments that the gradual movement to OA may seem quite slow, but will continue to grow.

The fourth chapter discusses controversies and pseudo-controversies. Crawford discusses the significant differences of opinion surrounding questions, such as the value that publishers add to scholarly articles, what it should cost to produce an online journal, whether or to what extent complete OA would save money, whether "delayed OA" makes sense or simply prolongs subscription journals, whether green or gold OA is preferable, whether OA journals that do not charge author-side fees are sustainable, and whether institutional mandates for depositing articles in repositories are effective. He also discusses some pseudo-controversies, myths, and misunderstandings that are cited as arguments against OA, such as OA undermines peer review, author-side fees taint the reliability of peer review, author-side fees prevent some researchers from publishing, and author-side fees disadvantage scholars who live in developing nations. Crawford also notes in this chapter inaccurate claims that online publishing is free; concerns that funding allocated

for OA will greatly lessen funding for research; and arguments such as inter-library loan makes research articles available to everyone, laymen should not have access to research because it might confuse them, and OA weakens copyright. The succinct but clear discussion of controversies alone makes this book valuable to librarians as a quick reference for questions from faculty and arguments from opponents to OA.

The fifth chapter suggests some actions librarians can take to support OA. Crawford lists five areas of concern for all librarians: understanding OA, communicating with community, encouraging discovery of OA articles, considering OA options when writing for publication, and keeping up-to-date as OA changes. Next, he discusses eight items that librarians should discuss with faculty and other researchers. He mentions institutional repositories, OA mandates, OA publishing, and OA funds for author-side fees as activities that some libraries may be involved with or may consider supporting. The chapter closes by encouraging librarians to consider researching OA issues and offers some questions that need to be addressed.

The final chapter recommends resources for learning more about OA and for remaining up-to-date on OA issues. In what may be the most valuable chapter of the book, Crawford annotates twenty-eight articles from the hundreds of newsletters, blogs, books, and other resources on OA. He provides lengthy annotations for five key sources, describing the content, authorship, update frequency, and extent of each. Most of the remaining items are annotated in four or five lines, with a focus on why the resource is useful. Crawford also indicates which resources espouse particular viewpoints.

Crawford's book fully meets its goals. It is a readable introduction to OA, clarifying major terms and introducing important people and issues in

the OA movement. While it is directed at librarians and library science students, both laypersons and scholars in other fields could read and understand it. A detailed index, logical layout,

and frequent headings throughout the text make it valuable as a reference for librarians discussing OA with faculty and administrators. Highly recommended for academic and large

public libraries; recommended for all libraries and for library science students.—*Ginger Williams (ginger.williams@wichita.edu), Wichita State University, Wichita, Kansas*