

# Library Resources & Technical Services

ISSN 0024-2527  
April 2010  
Volume 54, No. 2

## Approval Plan Profile Assessment in Two Large ARL Libraries

*Robert Alan, Tina E. Chrzastowski,  
Lisa German, and Lynn Wiley*

## Google Books as a General Research Collection

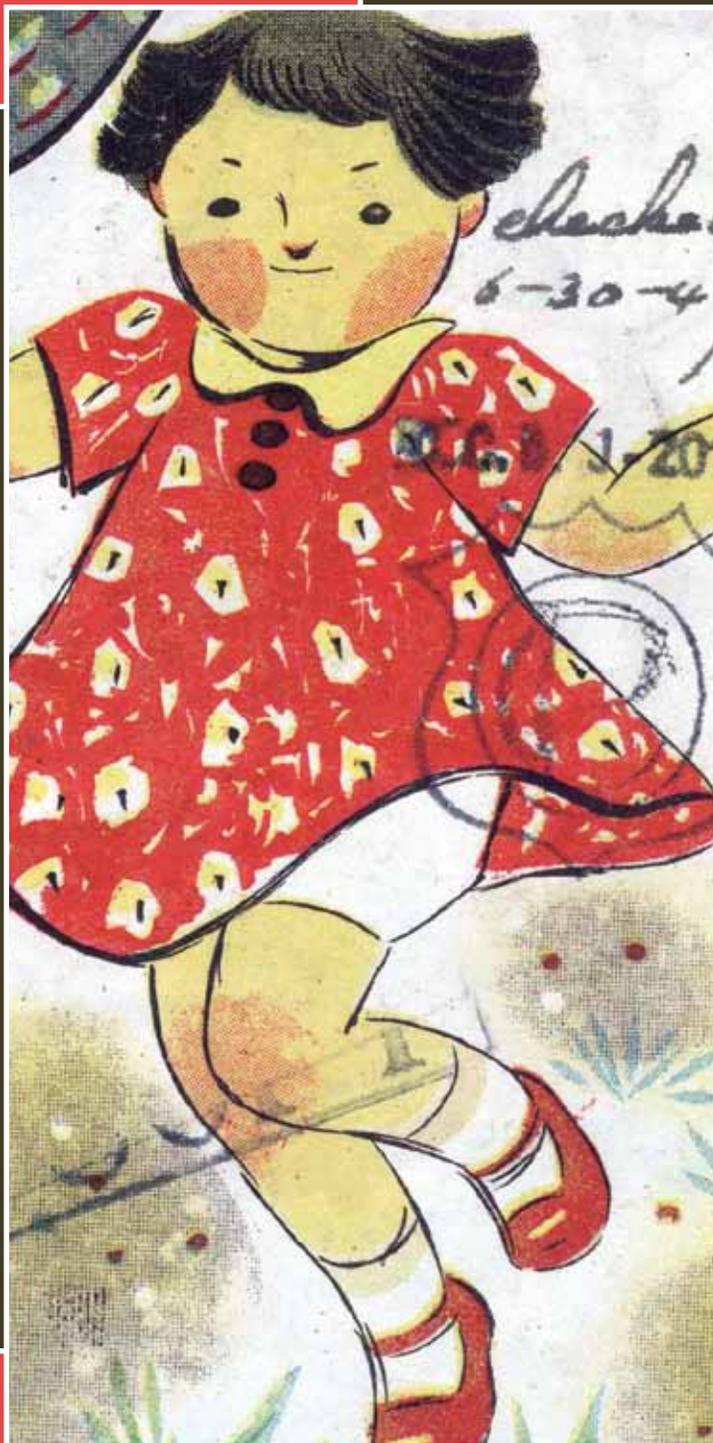
*Edgar Jones*

## Cataloging and Classification

*Sydney Chambers and Carolynne Mycall*

## Rethinking Research Library Collections

*Dan Hazen*



FROM THE LIBRARY OF CONGRESS

With Spanish,  
French, and  
German  
interfaces!

Your  
gateway to  
online cataloging,  
classification, and  
metadata tools.

Now with  
much quicker  
Class Schedule  
navigation!

## CATALOGER'S DESKTOP

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

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

**Free trial accounts & annual subscription prices:**  
Visit [www.loc.gov/cds/desktop](http://www.loc.gov/cds/desktop)

**For free trial, complete the order form at**  
[www.loc.gov/cds/desktop/OrderForm.html](http://www.loc.gov/cds/desktop/OrderForm.html)

AACR2 is the joint property of the American Library Association, the Canadian Library Association, and the Chartered Institute of Library and Information Professionals. ©2005 by ALA, CLA, CILIP. All rights reserved.



101 Independence Avenue, S.E. Washington, D.C. 20541-4912 U.S.A.  
Toll-free phone in U.S. 1-800-255-3666 | Outside U.S. call +1-202-707-6100  
Fax +1-202-707-1334 | [www.loc.gov/cds](http://www.loc.gov/cds) | [cgsinfo@loc.gov](mailto:cdsinfo@loc.gov)

## CLASSIFICATION WEB

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

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

**Free trial accounts & annual subscription prices:**  
Visit [www.loc.gov/cds/classweb](http://www.loc.gov/cds/classweb)

**For free trial, complete the order form at**  
[www.loc.gov/cds/classweb/application.html](http://www.loc.gov/cds/classweb/application.html)

Dewey and WebDewey are registered trademarks of OCLC, Inc.

**Free 30-Day Trials  
for Both Products!**

# Library Resources & Technical Services

ISSN 0024-2527

April 2010

Volume 54, No. 2

## Guest Editorial

New Areas for Cataloging Research  
*Carlen Ruschoff*

62

## ARTICLES

### Approval Plan Profile Assessment in Two Large ARL Libraries 64

University of Illinois at Urbana-Champaign and Pennsylvania State University  
*Robert Alan, Tina E. Chrastowski, Lisa German, and Lynn Wiley*

### Google Books as a General Research Collection 77

*Edgar Jones*

### Cataloging and Classification 90

Review of the Literature 2007–8  
*Sydney Chambers and Carolynne Myall*

### Rethinking Research Library Collections 115

A Policy Framework for Straitened Times, and Beyond  
*Dan Hazen*

### Book Reviews 122

### Index to Advertisers 121

Image taken from the cover of *Tanoshii asobi* by Arai Goro (Tokyo: Kōgen Shoin, 1947), a children's book held in the Gordon W. Prange Collection, University of Maryland Libraries, and reprinted with permission. The Prange Collection consists of materials published in Japan during the early U.S. Occupation from 1945 to 1949, when all printed materials were submitted to the Civil Censorship Detachment (CCD), an operating unit of Supreme Commander of the Allied Powers (SCAP).

*Library Resources & Technical Services* (ISSN 0024-2527) is published quarterly by the American Library Association, 50 E. Huron St., Chicago, IL 60611. It is the official publication of the Association for Library Collections & Technical Services, a division of the American Library Association, and provided as a benefit to members. Subscription price to nonmembers, \$85 per year in the U.S., Canada, and Mexico, and \$95 per year in other foreign countries. Single copies, \$25. Periodical postage paid at Chicago, IL, and at additional mailing offices. POSTMASTER: Send address changes to *Library Resources & Technical Services*, 50 E. Huron St., Chicago, IL 60611. Business Manager: Charles Wilt, Executive Director, Association for Library Collections & Technical Services, a division of the American Library Association. Send manuscripts to the Editorial Office: Peggy Johnson, Editor, *Library Resources & Technical Services*, University of Minnesota Libraries, 499 Wilson Library, 309 19th Ave. So., Minneapolis, MN 55455; (612) 624-2312; fax: (612) 626-9353; e-mail: m-john@umn.edu. Advertising: ALCTS, 50 E. Huron St., Chicago, IL 60611; 312-280-5038; fax: 312-280-5032. ALA Production Services: Troy D. Linker, Chris Keech, Tim Clifford, and Justine Wells. Members: Address changes and inquiries should be sent to Membership Department—*Library Resources & Technical Services*, 50 E. Huron St., Chicago, IL 60611. Nonmember subscribers: Subscriptions, orders, changes of address, and inquiries should be sent to *Library Resources & Technical Services*, Subscription Department, American Library Association, 50 E. Huron St., Chicago, IL 60611; 1-800-545-2433; fax: (312) 944-2641; subscription@ala.org.

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

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

©2010 American Library Association

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

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials, ANSI Z39.48-1992. ∞

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

Association for Library Collections & Technical Services

Visit *LRTS* online at [www.ala.org/alcts/lrts](http://www.ala.org/alcts/lrts).

For current news and reports on ALCTS activities, see the *ALCTS Newsletter Online* at [www.ala.org/alcts/alcts\\_news](http://www.ala.org/alcts/alcts_news).

# Year of Cataloging Research

## EDITORIAL BOARD

Editor and Chair  
Peggy Johnson  
*University of Minnesota*

### Members

Kristen Antelman, *North Carolina State University*

Stephen Bosch, *University of Arizona*

Elise Calvi, *Indiana Historical Society*

Allyson Carlyle, *University of Washington*

Mary Casserly, *University at Albany*

Elisa Coghlan, *University of Washington*

Leslie Czechowski, *University of Pittsburg Health Sciences Library*

Lewis Brian Day, *Harvard University*

Dawn Hale, *Johns Hopkins University*

October Ivins, *Ivins eContent Solutions*

Edgar Jones, *National University*

Birdie McLennon, *University of Vermont Libraries*

Randy Roeder, *University of Iowa*

Christine Marie Ross, *University of Illinois at Springfield*

Carlen Ruschoff, *University of Maryland*

Sarah Simpson, *Tulsa City-County Library System*

Elaine L. Westbooks, *University of Nebraska-Lincoln Libraries*

### Ex-Officio Members

Charles Wilt, Executive Director,  
ALCTS

Mary Beth Weber, *Rutgers University*, Editor, *ALCTS Newsletter Online*

Edward Swanson, *Minitex Library Information Network*,  
Book Review Editor, *LRTS*

## Guest Editorial

### New Areas for Cataloging Research

**Carlen Ruschoff**

**Carlen Ruschoff** (ruschoff@umd.edu) is Director of Information Technology and Technical Services, University of Maryland Libraries, College Park.

*LRTS is celebrating 2010 as the Year of Cataloging Research by publishing guest editorials and highlighting papers that advance this important initiative.*

The ALCTS-sponsored Year of Cataloging Research gives us the opportunity to take stock of where cataloging research has been and to consider a vision for future research initiatives. During the last few centuries, libraries and library catalogs were the only game in town. With no Internet, a person who wanted information either had to go to a library to find information, or to the source (e.g., the researcher, archives) to find information. We librarians used our time well in researching user behavior, developing subject-access schema, and honing cataloging codes to deal with an ever-growing number of publication types with an increasing array of complex publication patterns. Yes, in the good old days, we had the corner on the information discovery market, and there appeared to be little need to go beyond our boundaries to investigate how our work might intersect with seeming disparate disciplines.

What has changed? Of course, the Internet hit us. At first the search engines were clunky. Their imprecision and the large numbers of items returned for each search made users feel like they were drinking from a fire hose. Librarians could be arrogant and claim that the filters just were not robust enough to compete with the strict organization that only a cataloger can bring. As Web content multiplied at an exponential rate, search engines began to improve and users became better at using them. At the very least, users now feel that they are successful in finding information on the Web. Librarians began to say that users need us to help them decide what information is reliable and what is not. More recently, we have recognized the draw of the Web and now, in addition to creating finding aids, we provide tips to researchers on anything from how to choose a search engine to how to compose a dependable search query.

During the last twenty years, search engines have become more vigorous and dynamic. Other entities can afford to put a lot more resources into building search engines, including relevance ranking, full-text data mining, and filters. While no single search engine is the silver bullet in the world of discovery, we do know that the preferred finding tool is not the online public access catalog (OPAC). This news should be a wake-up call to all librarians—not just catalogers. It should be the inspiration to embark upon new research that will motivate all of us to find ways we can contribute our organizational skills. As Roeder said in the January *LRTS* guest editorial, the last thing we need to do is perform another user assessment and then change the OPAC.<sup>1</sup> More lipstick on our catalogs is not going to make our OPACs the search engine of tomorrow.

Are libraries and librarians the only groups to feel the shift? I think not. A few weeks ago, I was having a conversation with one of my colleagues on the University of Maryland teaching faculty. He told me that the Web has forever changed the way professors teach. He was not bemoaning the change, but simply remarking on it. He said that the former way of teaching was professor-centric and based on a limited sphere of knowledge. The professor created a series of lectures by pulling together the content, shaping it, and presenting it. The professor was the go-to person for all questions about the content. The library collection augmented the course content and provided more in-depth information to allow the student to expand on the specific course topics, but the Web has changed the entire framework of teaching. It provides access to more information, with varying levels of accuracy. My faculty colleague went on to say that to be an effective professor today and tomorrow, the professor must be more like a spiritual guide through the morass of a “universe of unstructured knowledge.” The professor must be able to teach students not only how to search but how to analyze what they find and transform the collective knowledge into thoughts, ideas, and conclusions. He added that it is important to recognize that the nature of the ideas and conclusions also may be changing. The student, using the infinite associative power of the Web, has a very different sense of a conclusion or an idea than those of his or her counterparts of twenty years ago. Both librarians and professors will have to learn what an idea means in this new era. Librarians, whether they work in reference or cataloging, will play more active roles in both the searching and teaching arena.

What does all of this have to do with cataloging research? I think the changes in our environment indicate that we must move our investigations away from the traditional cataloging codes and subject schemas. Instead, we need to conduct research that helps us better comprehend the information environment of today to build a greater understanding of how we can integrate our library skills into the discovery and learning milieu. We have a great deal to learn about how traditional cataloging principles fit (or do

not fit) with the architecture of the current search engines, the language of searching and tagging, and the organization of search results.

We can also learn a lot about the role of social software in discovery and learning. How do users employ folksonomic services? How do they apply and arrange tags? What do we know about how these arrangements affect how the researcher thinks about a work or rethinks about a work? And how do users respond and learn from the ways others have deployed tags? The historical knowledge regarding subject access and organization of information that catalogers can bring to this research will help the community reshape its thinking in this area.

What about the Maryland professor who sees a new pedagogical framework in the unstructured universe of knowledge and his desire for librarians to play a complementary role in the teaching environment? Knowing the landscape of search tools, how they are designed, and how to maximize their use can help us help our faculty.

Our research into these new areas (that is, new for us) will help us find fresh avenues through which to apply our organizational skills. It will put us in a stronger position to amplify search engine design, make the most of the social software services available, and remain true partners in the learning and teaching missions of our institutions.

To accomplish these research goals, we need library leaders who are willing to encourage and support librarians as they delve into these new areas of study. We also need catalogers who understand the Web environment well enough to be able to formulate research inquiries that will move the profession forward. In this Year of Cataloging Research, we should celebrate the research our profession has carried out in the past and, at the same time, roll up our sleeves and write some new chapters.

### Reference

1. Randy Roeder, “Guest Editorial: A Year of Cataloging Research,” *Library Resources & Technical Services* 54, no.1 (Jan. 2010): 2–3.

# Approval Plan Profile Assessment in Two Large ARL Libraries

## University of Illinois at Urbana- Champaign and Pennsylvania State University

By Robert Alan, Tina E. Chrzastowski, Lisa German,  
and Lynn Wiley

*Two Association of Research Libraries member libraries, the University of Illinois at Urbana-Champaign (UIUC) and Pennsylvania State University (Penn State), evaluated their monograph acquisition approval plan profiles to answer basic questions concerning use, cost effectiveness, and coverage. Data were collected in tandem from vendors and local online systems to track book receipt, item circulation, and overlap between plans. The study period was fiscal year 2005 (July 1, 2004–June 30, 2005) for the approval plan purchasing data, and circulation use data were collected from July 1, 2004, through March 31, 2007, for both UIUC and Penn State. Multiple data points were collected for each title, including author, title, ISBN, publisher, Library of Congress classification number, purchase price, and circulation data. Results of the study measured the cost-effectiveness of each plan by subject and publisher, analyzed similarities and differences in use, and examined the overlap between the two approval plans. The goals were to establish a benchmark for consistently evaluating approval plan profile effectiveness and to provide a reproducible method with baseline data that will allow other libraries to collect comparable data and conduct their own studies.*

Submitted June 29, 2009; tentatively accepted September 15, pending modest revision; revised and resubmitted August 27, and accepted for publication.

**Robert Alan** (roa1@psu.edu) is Head of Serials and Acquisitions Services, Pennsylvania State University Libraries, University Park, Pennsylvania; **Tina E. Chrzastowski** (chrz@illinois.edu) is Chemistry Librarian, Chemistry Library, University of Illinois at Urbana-Champaign; **Lisa German** (lgerman@psu.edu) is Assistant Dean for Technical and Collection Services, Pennsylvania State University Libraries; and **Lynn Wiley** (lnwiley@illinois.edu) is Head of Acquisitions, University Library, University of Illinois at Urbana-Champaign.

Penn State thanks Heather Benner, Sydney Bennington, Betty Nirberger, and Deb Richner for their data expertise and administrative support. UIUC thanks Jon Gorman for his data expertise and to the Consortium of Academic and Research Libraries in Illinois for travel support.

Approval plans have been considered an efficient and cost-effective way for libraries to acquire books in large quantities across many disciplines. Through approval plans, vendors supply current imprints as well as notification slips or forms to libraries on the basis of selected publisher output, subject profiles, and nonsubject categories such as readership level, country of origin, and format. When combined, these factors determine the parameters for selecting titles within the approval plan. Approval plan profiles can be limited by any number of factors, including price, scope, format, audience, language, and publisher. Each approval plan's profile is carefully established by library subject specialists to meet the research, curricular, and learning needs of the library's users.

If a library commits to purchase large quantities of books on approval, vendors may offer substantial discounts off the list price. Libraries also may have the option to return titles that they consider outside of the approval profile. Additional vendor services include shelf-ready services, such as cataloging, bar

coding, and labeling, at an added cost. However, shelf-ready titles cannot be returned unless they are received damaged or clearly outside of the approval profile (e.g., item exceeds price limit).

Approval plan profiles can take considerable time to formulate and, once implemented, may not always be subject to regular review and revision. However, libraries should regularly consider a number of questions concerning their approval plan profiles, including the following:

- How frequently should profiles be evaluated and revised?
- What criteria should be used when assessing the effectiveness of approval plan profiles?
- Can cost-effectiveness be measured, and if so, do the results point to reevaluation of local profiles?

To answer these and other questions surrounding the use of approval plans in large libraries, especially within the Association of Research Libraries (ARL), the authors conducted an assessment of domestic approval plans at the University of Illinois at Urbana-Champaign (UIUC) and Pennsylvania State University (Penn State). The study examined receipts from two book vendors: Blackwell Book Services at UIUC and YBP (formerly known as Yankee Book Peddler) at Penn State. These two institutions, both of which are members of the Committee on Institutional Cooperation, planned to undertake major reviews of their approval plans and decided that developing a study comparing results from similar institutions would be advantageous.

Although these university libraries vary in size and use different vendors, they share the mission of all libraries: to acquire the materials needed by their clientele. In large academic libraries, such as UIUC and Penn State, this usually entails the use of approval plan profiles. Differences and similarities between the two libraries and their approval plans became apparent during the research process. The authors conducted the study at each library using the same method in order to compare results, and they logged the discrepancies through the data collection and analysis process.

### Research Questions

The most critical area examined relates to the use and cost of material acquired by libraries through approval plan profiles. The primary research question focused on examining the current method of providing large quantities of books to support the research, teaching, and learning needs of the students and faculty of each university studied and asked the following: Can a cost/use ratio be derived that indicates the point at which an approval plan profile is effective or

ineffective? Beyond the basic analyses of cost and use, other questions were framed to guide the data analysis:

- How does circulation and cost/use compare between UIUC and Penn State?
- How does cost/use vary by subject discipline at UIUC and Penn State?
- Do the two approval vendors (Blackwell and YBP), in combination with the different profiles, overlap? Are the two libraries buying a high percentage of the same titles?
- What publishers represent the highest use at each library? Is there a correlation between the highest volume publishers and the highest average use?
- Is Trueswell's 80/20 rule applicable to approval book purchases; that is, do 20 percent of the approval books account for 80 percent of their circulation?<sup>1</sup> Is Kent's hypothesis in *Use of Library Materials: The University of Pittsburgh Study*, "A very small portion (perhaps 10 percent) of the library collection of book titles accounts for major portion (80 percent or more) of circulation and in-house use," a more likely outcome?<sup>2</sup>

### Literature Review

Numerous publications have broadly examined the use of library materials, and several important studies have examined the use of books over a period of time. Research also has been conducted on the use and cost of books acquired through approval plan profiles.

Studies that have addressed the effectiveness of approval plans include those by Kingsley and Brush.<sup>3</sup> In 1996, Kingsley found that 50 percent of approval plan books circulated within the first five months after receipt, and 67 percent circulated within the first sixteen months after receipt at Western Michigan University. In her subsequent (2000) study, Kingsley advocated the use of management reports to closely monitor circulation patterns of approval materials, asserting that "the likelihood that an approval plan will continue on automatic pilot, adding books in some very low-use areas and perpetually short-changing some heavy-use topics offers the risk of particularly ineffective spending if management information about approval plan book use is not monitored."<sup>4</sup> Brush compared the circulation of engineering titles received on approval with the circulation of all materials in the Library of Congress "T" call number classification at Rowan University in the 2005 fiscal year (FY05), with both acquisition and circulation taking place in FY05. The results showed that books received from the approval plan profile did circulate at a rate much higher

than the collection as a whole. The overall circulation rate for approval plan books was 23 percent, versus 6 percent of the engineering collection as a whole. Brush concluded, "Our approval plan books (the most recent ones) circulated at a much higher rate than the engineering collection as a whole, indicating that they are filling a real need."<sup>5</sup>

A few studies have taken the next step and examined the cost/use ratio of monographs. Crofts looked at cost and circulation of monographs by subject to develop a funding formula. Over a five-year period (1990–95), "values expended per book range from less than one dollar (recreation) to almost twenty-five dollars (accounting)."<sup>6</sup> Rodriquez studied the cost and use of monographs at an academic health sciences library over a three-year period: July 1, 2004, to June 30, 2007. He found, using a ratio of expenditure (cost of book) to circulation, that health science subjects varied in Actual Cost of Use (ACU) from \$8.04 to \$191.31 with a mean of \$39.03.<sup>7</sup> A University of Texas study calculated the cost per use of printed books at between \$3.24 and \$28.57; no time frame was given for these data, but they include the ongoing costs of heating, ventilation, air conditioning, shelving, and maintenance.<sup>8</sup>

These studies show that both cost/use ratios and circulation rates for books can vary widely. These variances can be attributed to the subject matter and the size, scope, and type of library, as well as the size of the approval plan and the nature of the profile. The different results confirm the need to compare similar libraries with similar plans or to conduct multiple year studies at a single library with an approval plan profile that is consistent over time. Previous studies helped establish a baseline for comparison with this research study. Most circulation studies look at longitudinal data over a series of years to demonstrate use; in this study, titles had between twenty-one and thirty-three months to circulate.

Juran initially proposed the law of the vital few (20 percent) and trivial many (80 percent) in the context of business operations.<sup>9</sup> Trueswell later applied the 80/20 rule to library collection development by suggesting that 20 percent of the collection accounts for 80 percent of the circulation.<sup>10</sup> He used the 80/20 rule to support the development of core collections centered on the 20 percent of the collection that generated the most use with the understanding that the remaining 80 percent would circulate less frequently or not at all.

Kent examined for five years the circulation of monographs acquired in 1969 at the University of Pittsburg.<sup>11</sup> He proposed three hypotheses: that 10 percent of the library collection would account for 80 percent of circulation and in-house use; that 25 percent of the collection would not be used in ten years; and that 50 percent of the collection would circulate once or not at all in a ten-year period. The study confirmed these hypotheses and determined that the window for a book to circulate was

limited and the first two years of availability determined future circulation.

The 80/20 rule has been tested over time with varying results and has been one benchmark used to assess the effectiveness of collection development. Hardesty's study of the circulation of books acquired in a six-month period indicated that only 63 percent of books acquired at DePauw University circulated within five years, and 30 percent of books generated 80 percent of the total circulation.<sup>12</sup> Hardesty later replicated the DePauw study at Eckerd College with similar results, finding that 34 percent of books received accounted for 80 percent of circulation.<sup>13</sup>

Results of other studies do not strongly support the 80/20 rule. Hamaker studied recently acquired monographs to determine circulation patterns within a very limited timeframe.<sup>14</sup> Of the newly acquired monographs cataloged in September 1990, 43 percent had circulated by mid-February 1991. Treadwell's study examined the use of titles at Texas A&M coded "select" by vendor Baker and Taylor (titles most likely to be reviewed in a scholarly journal), testing the hypothesis that these materials were more likely to circulate than those that were not coded "select."<sup>15</sup> She also hypothesized that titles at the undergraduate level were more likely to circulate than those at the graduate level and that books covering broad subject areas were more likely to have circulated at least once in the first eighteen months of receipt than books covering specific disciplines. The results of Treadwell's study showed that 95 percent of all "select" books circulated in the first year, except for undergraduate humanities titles, which circulated at a 76 percent rate. This study also showed that social science and science materials circulated at almost the same high circulation rates—95 to 99 percent.

Studies on monograph use in health science libraries do not support either the 80/20 rule or Kent's 80/10 hypothesis. Eldredge found that of 1,306 monographs added to the collection of the Health Sciences Center Library at the University of New Mexico in 1993, 84 percent circulated at least once by November 1997.<sup>16</sup> Eldredge also found that 19.45 percent of monographs accounted for 57.80 percent of circulations and 36.29 percent of monographs for 79.76 percent of circulations. In another study of 1,600 monographs at the University of Illinois at Chicago Health Sciences Library, Bleic found that monograph use did not decline sharply over the three-year period of her study, with percentages of use at 38.69 percent in year 1, 32.37 percent in year 2, and 29.85 percent in year 3, for a total of 7,659 circulations of 1,674 monographs.<sup>17</sup> The results did not strongly support the 80/20 rule because 38 percent of monographs accounted for 80 percent of circulation and 2.21 percent of monographs accounted for 21.84 percent of circulation. The higher use may be attributed to the difference in user populations, that is, health sciences students

versus undergraduates, who were the focus of the Kent and Hardesty studies.

### **Penn State's YBP Library Services Approval Plan**

In FY05 the Penn State system included twenty-three campuses located across the Commonwealth of Pennsylvania. Penn State is organized as a single university geographically dispersed, and therefore all campus libraries are part of the University Libraries. Campuses range in size from less than 800 students at smaller campuses to more than 42,000 students at the largest campus, University Park. In FY05 Penn State student enrollment totaled approximately 81,000 students (70,000 undergraduate and 11,000 graduate students). With more than 5,000 faculty (tenure track and fixed term) and 12,000 staff (non-tenure track positions in all job classifications), the total university-affiliated user population was 98,000. Each campus library is allocated a materials budget and is responsible for selecting information resources that support the teaching and research needs of its own campus faculty and students. With few exceptions, materials acquired at any campus library are available to Penn State users regardless of location within the university. Acquisition and cataloging operations for most campus locations are centralized at University Park.

The University Libraries maintained nine approval plans in FY05. One small awards plan was at a non-University Park campus; the remaining eight plans supported collections at University Park. Seven of the eight plans were relatively small and focused on specific subject content (e.g., music scores) and foreign language materials. The largest approval plan was a comprehensive English language plan with YBP for automatic delivery of books and notification slips covering most subject areas. The YBP approval plan was augmented by a small plan with YBP's British subsidiary, Lindsay and Croft.

Penn State's YBP approval plan was first established in 1992 and has been refined over the years. The plan is divided into multiple subprofiles that are based on broad subject areas that mirror the University Park subject library orientation. Subprofiles include the following subject areas: arts and humanities (including architecture), business, education, engineering, earth and mineral sciences, life sciences, physical sciences, mathematical sciences, and social sciences. Each subprofile has a fund allocation based on historical publication data and projected inflation for the fiscal year.

The YBP plan is detailed and granular to ensure appropriate coverage across subprofiles and eliminate any overlap. For example, the arts and humanities subprofile supplied books for LC class P (philology and linguistics

(general)) with the exception of P88-96 (communication, mass media), P301-302 (style rhetoric, composition), and P304 (vocabulary), which were included in the social science subprofile. A price limitation requiring notification, as opposed to automatic book delivery, varied from \$175 to \$200. Preference was given to receipt of cloth-bound over paperback books in FY05. The YBP plan is a unified plan that includes university presses, trade publishers, and other publishers within each subprofile. University press coverage was and still is a priority, and often subprofiles exclude books in some subject areas with the exception of automatic delivery of books published by university presses.

YBP approval books are received shelf-ready on a weekly basis. Shipments are reviewed by selectors to monitor the quality of the collection. The review also allows selectors to monitor the publishing output and discover emerging trends across subject areas. Selectors flag approval receipts for additional processing (e.g., binding, location changes, etc.). Because the books are received shelf-ready, returns are limited to defective volumes and obvious vendor errors (e.g., book exceeds \$200 price limit).

### **Penn State's YBP FY05 Receipts**

The YBP universe of titles profiled in FY05 was 52,794. Penn State's YBP approval plan profile resulted in the automatic delivery of 15,520 (29 percent) of YBP's profiled titles. Furthermore, notification slips were profiled for an additional 23,339 titles, leading to 3,119 firm orders placed with YBP. The combination of automatic book delivery and firm orders generated from notification slips resulted in the receipt of 18,639 (35 percent) of YBP profiled titles.

### **UIUC's Blackwell Library Services Approval Plan**

UIUC is the largest of three University of Illinois campuses, with other locations in Chicago and Springfield. In FY05, the UIUC campus population consisted of approximately 39,000 students (29,000 undergraduate and 10,000 graduate students) and almost 6,000 academic staff (including faculty, academic professionals, and postdoctoral students), for a total of approximately 45,000 potential library users on campus. The UIUC library system is composed of a main library, an undergraduate library, and thirty-eight departmental libraries, many of which are dispersed throughout the campus in departmental buildings. The UIUC Library has a centralized acquisitions department that processes materials for all but the Asian and Slavic Libraries (the Law Library is under a separate administration). The largest approval plan covers English language monographs published in the United States and the United Kingdom and is vended to

Blackwell Book Services. The UIUC Library central acquisitions unit also maintains twenty other plans: four European blanket order plans, twelve Latin American profiles, several African order plans, and two blanket order plans for music (one for books and one for scores). The plans are built to serve the users of the campus at UIUC.

The domestic publisher approval plan at the UIUC is serviced by Blackwell's Book Services and is the largest and most comprehensive approval plan at the library. UIUC awarded a contract to Blackwell Book Services in 2003 for domestic firm and standing orders as well as the approval plan. The domestic plan for English language books is for a comprehensive subject range. The approval plan in FY05 served every departmental library with books received in all disciplines. Notification slips for higher priced titles and legal and medical titles were sent. Legal and medical notification slips were reviewed by library subject specialists for approval selection to meet the focused needs of the law school research interests and clinical veterinary medicine program. The approval plan was funded centrally in FY05 and was not allocated into subaccounts. Support was provided for all materials covered by the plan at that time, although records were kept by selection location for reporting purposes. In FY05, 11,037 books were received on the Blackwell domestic approval plan. The library used the same vendor to purchase UK imprints. Although the U.S. publication was always preferred, the UK title was shipped if the U.S. title was not published simultaneously. Several thousand additional titles were purchased on the UK side, but only the domestic imprints were analyzed for this study.

The Blackwell approval plan contained two publisher plans. One covered 90 mainstream trade press publishers with many of their imprints. The second plan covered 74 university presses for 175 imprints. Liberal price caps were in place at that time, allowing any title under \$500 to be shipped. UIUC uses the Dewey Decimal Classification system with locally applied exceptions, therefore complete shelf-ready processing was not possible. However, the Blackwell supplied PromptCat catalog records (an automated service that delivers copy cataloging records for materials purchased from vendors) and bar coded each book. UIUC's Blackwell approval plan was a paperback-preferred plan. Books were shipped weekly by courier service and were available for review by selectors for two weeks. Because liberal centralized funding was available and the plan was well focused, very few titles were returned in FY05.

The UIUC approval plan specifically excluded many important standing orders. These continuations were purchased on separate orders and were budgeted for within specific subject funds. Book series that selectors wished to receive directly did not come via the publisher-based approval plan and therefore were not included in the data

used for this study. Also, as noted earlier, the approval plan was a U.S.–, and then UK–, preferred match plan. This meant that when a title was only available from the UK for a set time interval, the UK title was sent rather than waiting for the U.S. imprint in order to best serve the goal of obtaining the content as soon as possible. The UK titles were not analyzed by UIUC or Penn State for this study.

### UIUC's Blackwell FY05 Receipts

In FY05 the Blackwell universe of profiled titles was 56,489. UIUC's Blackwell approval plan profile resulted in the delivery of 11,037 domestic titles, or 20 percent of Blackwell's profiled titles. Notification slip receipts were part of that total. Of the 11,037 titles received on approval, 6,030 were trade press titles and 5,007 were university press titles. Approximately \$500,000 was spent in FY05 on the domestic titles received on approval from Blackwell. The library selectors also ordered and received another 4,882 titles as firm orders from the Blackwell Collection Manager online interface to select titles that were then batch ordered. That brought the number of monographs ordered to 15,919, or 28 percent of the Blackwell output. In the same year, UIUC received on standing order another 2,635 volumes; more than 1,600 of these were yearbooks, directories, and other annuals that traditionally have been excluded from most approval plans. However, approximately 1,000 titles were separately classed monographs received as part of an ongoing book series, such as the Springer book series, and these volumes brought the final total to approximately 17,000 titles, or 31 percent of the Blackwell profiled titles.

FY05 was not a normal year for UIUC's Blackwell approval plan. In 2004, Blackwell implemented a new distribution system that did not deliver on its promise of more efficient and timely selection and delivery of books. Approval matches could not be fulfilled, and consequently hundreds of orders were redirected from Blackwell to other vendors during the second half of the 2004–5 academic year. This resulted in a significant reduction in books acquired on approval from Blackwell in FY05. Nevertheless, the authors decided to use FY05 approval data for this study, knowing that the method and data would be sufficient for comparison purposes and provide a base year for future comparisons.

### Research Method

The study examined the use of domestic monograph titles received on approval at UIUC and Penn State for FY05 (July 1, 2004–June 30, 2005). The authors studied circulation data for these approval receipts to determine use patterns by publisher and subject. Additionally, overlap

between the two approval plans was examined. Data were extracted from each library's integrated library management system (Sirsi Unicorn at Penn State and Endeavor Voyager at UIUC) for titles acquired on approval from Blackwell (UIUC) and YBP (Penn State) for FY05. Circulation data were extracted for these titles from the time of receipt in FY05 through March 31, 2007. Therefore approval books received in FY05 had between twenty-one and thirty-three months from the time of receipt to circulate at Penn State and UIUC. The extracted data were then moved into Microsoft Access databases, which were queried to answer specific research questions.

As previously noted, Penn State received YBP approval books shelf-ready and PromptCat catalog records for loading into Sirsi Unicorn. Because of system limitations and workload constraints, Penn State could not create order records for YBP approval books. The YBP approval titles for FY05 were identified by the presence of a single MARC 980 field (PromptCat acquisition data) in the catalog record. The MARC 980 field included fund codes and the YBP invoice date, which facilitated identification of approval titles received in the various broad subject areas. Penn State identified and extracted the data in May/June 2007 for records loaded into Sirsi Unicorn for FY05, and therefore some catalog records had been updated and either lacked a 980 field or included multiple 980 fields. Of the 15,520 YBP approval books received in FY05, this study analyzed the 13,660 titles (88 percent) that contained a single MARC 980 field. Data fields extracted from Sirsi Unicorn were the following:

- Date Record Created
- OCLC Record Control Number
- 020—ISBN
- 050—Library of Congress Call Number
- 090—Local Call Number
- 1xx—Main Entries
- 245—Title Statement
- 246—Varying Form of Title
- 260—Publication, Distribution, etc. (Imprint)
- 980—YBP PromptCat acquisition information
- Item record—Location information and circulation counts

UIUC extracted MARC bibliographic records and associated acquisitions data from the Voyager system using the Blackwell approval plan ledger reports established to account for approval receipts title by title. Brief bibliographic records provided by Blackwell were automatically loaded into Voyager on a weekly basis. Staff successfully matched these files to the full bibliographic record, thereby allowing for alignment of UIUC's records with those of Penn State. Data fields extracted from Voyager were the following:

- Bib ID
- Fund Code
- Location Name
- Price
- Brief Title (order record)
- Author (order record)
- ISBN
- OCLC Number (taken from bib)
- Line item create date (order record = Blackwell)
- Line item status (recvd)
- Status date (receive date so date item itself available)
- Publisher
- Publisher Date
- Item ID
- Barcode
- Display Call Number (local Dewey Call)
- Place Code (used to eliminate UK items)
- Action Date (first time record changes after record imported)
- Count Charges
- Count Renewal
- LC Call Number (050)

Books from the approval plan profiles that were selected for reference collections (noncirculating location) or reserve (potential for high circulation) were noted in the database. Noncirculating and high-circulation locations can skew circulation patterns, which prompted a review to decide whether to exclude these categories from the study. However, reference and reserve titles accounted for approximately 1 percent of the overall titles listed, and the authors decided to retain these categories in the study.

Penn State analyzed 88 percent of YBP approval receipts, whereas UIUC successfully matched and analyzed all Blackwell approval receipts for FY05. At both libraries, data included the full fiscal year of approval receipts. Circulation data were extracted to match the fiscal year for the FY05 approval titles extending through March 31, 2007. Circulation policies were compared at the two libraries and were found to be nearly identical, with books circulating to faculty and graduate students for a semester (sixteen weeks) and to undergraduates for four weeks.

## Findings and Analysis

The authors examined patterns of use, nonuse, cost per use, and overlap between the Penn State and UIUC approval plan profiles for FY05. Table 1 shows the number of approval titles received from trade and university presses and summarizes the circulation and cost per use for approval plans in the aggregate. Both libraries received a much higher proportion of receipts from trade publishers.

**Table 1.** Summary of Receipts, Circulation, and Cost per Use of Approval Plan Books

Institution	Trade Publisher Titles		Univ. Press Titles		All Approval Plan Titles		Circ.	Ave. Cost/ Title	Ave. Cost/ Use	Ave. Circ./ Title
	Titles	% of Total	No. of Titles	% of Total	Titles	Total Cost				
Penn State	8,963	66	4,695	34	13,658	\$772,610	38,942	\$56.56	\$19.83	2.85
UIUC	6,455	58	4,582	42	11,037	\$425,876	19,112	\$38.59	\$22.28	1.73

**Table 2.** Circulation and Noncirculation of Approval Plan Books

Institution	% of All Titles that Circulated	Frequency of Circulation Activity (%)				
		0	1	2-4	5-9	> 10
Penn State	69	31	15	29	20	4
UIUC	60	40	18.6	32.5	8	1
<b>Circulation of Trade Books</b>						
Penn State		30.5	15.1	29	20.5	5
UIUC		39.5	18.6	33	8.1	1
<b>Circulation of University Press Books</b>						
Penn State		32.8	15	29.6	20.1	2.4
UIUC		40	18.6	32	7.5	1.3

This result was attributed to the higher publication output of trade versus university presses. The total number of Penn State approval books included in the study (13,660 titles out of 15,520 approval titles acquired) was 19 percent more than at UIUC. Penn State's approval books were acquired at a substantially higher total cost to serve a much larger number of users because of the purchase of higher-priced clothbound books at a lower vendor discount. As previously noted, UIUC did not acquire as many approval books as would have been expected in FY05 because of the implementation of a new materials distribution system at Blackwell. Nevertheless, in terms of averages, the cost-per-use ratio (total cost from the vendor divided by total use) and the average number of charges per book were in the same range, although slightly higher at Penn State. Penn State's higher cost-per-use ratio and average charges per title were most likely attributable to their higher acquisition costs and larger user population.

As shown in table 2, Penn State's percentage split between approval titles that circulated and titles that did not circulate was approximately 70/30, whereas UIUC's was 60/40. Each of Penn State's circulation categories was slightly higher than UIUC's, again most likely because the Penn State user population was more than twice the size of UIUC (98,000 compared to 45,000). Proportionally, the numbers followed similar trends. As expected, the number of books circulating more than ten times was low and most likely represents course-reserve book use. The large percentage of books not circulating within one to two years of acquisition (31 percent at Penn State, 40 percent at UIUC)

was disappointing but not unexpected and is addressed later in this paper.

Table 2 also shows the circulation data for each university by publisher group, comparing circulation frequency of the trade publishers to that of the university press publishers. The circulation frequency for trade and university presses was comparable at both institutions. The 0, 1, and 2-4 circulation frequencies correlate highly between the two libraries. Higher circulation at Penn State is again most likely because of its larger user population.

Table 3 shows circulation by broad subject discipline: humanities, social sciences, and sciences, based on LC class numbers. The humanities included LC classes B, C, M, N, and P; social sciences included LC classes D, E, F, G, H, J, K, and L; and sciences included LC classes Q, R, S, T, and any National Library of Medicine call numbers. The sciences had the highest average number of circulations per book: 3.74 at Penn State and 1.87 at UIUC. These results dispel the notion that scientists use only journal literature and not books, or use books less frequently than the humanities or social sciences. The highest number of circulations when disciplines are compared was in the humanities at Penn State and the social sciences at UIUC. These also are the subject areas where the most books were acquired. The question remains whether the profiles for the sciences are more effective, resulting in the highest number of circulations per book, or whether fewer science books are acquired on approval, meaning fewer choices for scientists looking to borrow a book. Perhaps the higher circulation is a consequence of both of these reasons. The lower number of

**Table 3.** Number of Titles, Circulation, and Average Cost and Circulation per Book by Broad Subject Discipline

		No. Titles	Total Circ.	Cost/Title	Cost/Use	Circ./Title
<b>Humanities</b>	Penn State	6,664	15,788	\$46.14	\$19.48	2.37
	UIUC	3,909	6,147	\$31.59	\$20.12	1.57
<b>Social Sciences</b>	Penn State	4,338	13,215	\$51.79	\$17.00	3.05
	UIUC	4,803	8,620	\$34.46	\$19.25	1.79
<b>Sciences</b>	Penn State	2,658	9,945	\$82.88	\$22.15	3.74
	UIUC	2,321	4,345	\$58.98	\$31.54	1.87

**Table 4.** Top Ten Subjects by Number of Titles Purchased, with Use and Cost/Use Data

Penn State	Titles	Circ./Title	Cost/Use	UIUC	Titles	Circ./Title	Cost/Use
Electric./Engineering	1483	4.45	\$18.90	American Literature	1321	1.51	\$14.17
Theory/Pract. Educ.	1456	3.6	\$11.88	History: America	1192	1.80	\$16.92
American Literature	1406	2.07	\$13.66	Economics	790	1.73	\$23.84
History	1402	2.71	\$14.22	Math/ Comp. Science	778	1.97	\$31.30
English Literature	1353	2.39	\$18.18	Literary History	615	1.71	\$15.90
Economics	1310	2.78	\$18.24	History Americas	583	1.48	\$17.39
Math/Comp. Science	1306	3.07	\$22.69	Social Pathology	553	1.79	\$19.21
Sociology	1133	3.85	\$12.71	English Literature	485	1.64	\$20.40
Literary History	1117	2.63	\$18.90	History of Asia	459	1.88	\$17.00
Architecture	985	3.97	\$11.88	Theory/Pract. Educ.	454	2.04	\$17.44

science books received on approval is most likely the result of the sciences moving available funding into serials over time, leaving less monograph funding available for books, including approval books. Additionally, important science-related series with an expected and relatively high use (e.g., the many “Springer Lecture Notes in . . .”) were maintained on standing order at UIUC, and their circulation numbers were not included in this study. Another factor may be that the significantly lower price caps at Penn State resulted in fewer selections of the higher-priced science books arriving as approval titles, resulting in a smaller pool of those titles for a larger group of users.

Table 4 shows the top ten subject disciplines by number of titles acquired at Penn State and UIUC. Although major disciplines have considerable overlap in titles collected, two programmatic differences likely affected the extent of overlap. Penn State’s profiles focused on engineering and architecture, both important academic programs at

Penn State, whereas UIUC received a higher percentage of approval receipts in the humanities, an area of higher publication output. These data show a correlation between relatively high circulation and number of titles purchased in the discipline, meaning that users are finding and using materials in fields where the libraries are purchasing larger quantities of titles.

#### Overlap between Penn State and UIUC

An earlier study by Nardini, Getchell, and Cheever examined overlap in YBP approval plan receipts at two larger academic libraries (Penn State and the University of Southern California (USC)) and two medium-size academic libraries (Occidental and Wake Forest) in FY95.<sup>18</sup> That study found a 51 percent overlap between Penn State and USC. The authors of the present study expected that the overlap between Penn State and UIUC would be similar to those

**Table 5.** Top Trade Publishers by Number of Titles Acquired on Approval

<b>Penn State</b>	<b>Titles</b>	<b>Circ./Title</b>	<b>Cost/Use</b>	<b>UIUC</b>	<b>Titles</b>	<b>Circ./Title</b>	<b>Cost/Use</b>
Routledge <sup>°</sup>	772	2.97	\$26.87	Wiley <sup>°</sup>	665	1.88	\$27.44
Springer <sup>°</sup>	646	2.87	\$31.72	Harcourt	562	1.69	\$31.46
Palgrave/MacMillan <sup>°</sup>	556	2.11	\$25.67	Palgrave/MacMillan <sup>°</sup>	404	1.51	\$29.78
Wiley <sup>°</sup>	490	5.05	\$9.95	Random House <sup>°</sup>	333	1.73	\$10.81
Ashgate	285	2.01	\$45.40	Routledge <sup>°</sup>	301	1.83	\$22.60
Elsevier <sup>°</sup>	256	3.45	\$38.00	CRC <sup>°</sup>	234	2.28	\$40.55
Kluwer <sup>°</sup>	197	2.63	\$41.79	Kluwer <sup>°</sup>	200	1.97	\$52.68
CRC <sup>°</sup>	194	3.31	\$40.68	Springer <sup>°</sup>	195	1.61	\$48.22
Pearson	181	5.27	\$11.29	Simon&Schuster	178	1.92	\$10.13
Praeger <sup>°</sup>	176	2.19	\$34.48	McGraw Hill <sup>°</sup>	149	2.15	\$23.92

<sup>°</sup> Denotes top ten trade publishers in common between Penn State and UIUC

**Table 6.** Top University Press Publishers by Number of Titles Received

<b>Penn State</b>	<b>Titles</b>	<b>Circ./Title</b>	<b>Cost/Use</b>	<b>UIUC</b>	<b>Titles</b>	<b>Circ./Title</b>	<b>Cost/Use</b>
Cambridge	661	3.22	\$20.27	Oxford <sup>°</sup>	334	2.01	\$18.09
Oxford <sup>°</sup>	661	2.57	\$20.80	Yale <sup>°</sup>	221	1.53	\$24.94
Yale <sup>°</sup>	164	2.89	\$12.60	SUNY <sup>°</sup>	207	1.61	\$17.26
SUNY <sup>°</sup>	145	3.30	\$15.55	Princeton <sup>°</sup>	205	1.73	\$16.23
Princeton <sup>°</sup>	132	2.98	\$12.21	U. California <sup>°</sup>	201	1.59	\$16.71
U. California <sup>°</sup>	124	4.04	\$9.71	Harvard <sup>°</sup>	196	2.21	\$12.05
U. Chicago <sup>°</sup>	113	2.96	\$13.30	Columbia	148	1.68	\$18.28
Manchester	95	1.77	\$10.35	MIT <sup>°</sup>	133	2.11	\$14.97
MIT <sup>°</sup>	95	3.69	\$33.49	U. Chicago <sup>°</sup>	132	2.01	\$18.09
Indiana <sup>°</sup>	87	2.52	\$16.37	Indiana <sup>°</sup>	127	1.94	\$13.99

<sup>°</sup>Denotes top ten university presses in common between Penn State and UIUC

found by Nardini, Getchell, and Cheever, given the similarities in academic programs and user populations. Because of limitations in matching by ISBN or title from the available data, the authors conducted a manual comparison of titles. The sort and comparison yielded a match of 5,593 approval titles that were selected and received by both Penn State and UIUC. This corresponds to a 41 percent overlap for Penn State (5,593 of 13,660 titles) and a 51 percent overlap for UIUC (5,593 of 11,037 titles). The results for UIUC are exactly those found by Nardini, Getchell, and Cheever. The lower overlap for Penn State (41 percent versus UIUC's 51 percent) was due in part to Penn State's higher number of approval receipts, which increased the probability of a unique title. The overlap rates for the receipts from university presses was higher (74 percent for Penn State and 76 percent for UIUC), which is attributed to both libraries' commitment to select broadly from this publisher type.

The authors conducted an additional analysis of the

overlap group to identify the publishers supplying these popular titles. Tables 5 and 6 compare receipts, circulation counts, and cost/use data for Penn State and UIUC for the top ten trade and university press publishers.

At both Penn State and UIUC, the top ten trade and university press publishers accounted for nearly half of the total approval plan profile books received (44.1 percent for Penn State and 46.4 percent for UIUC). The circulation counts indicate a similar pattern, showing that those high-volume publishers had the highest circulation counts. Circulation data show that the content distributed by these publishers is in high demand by the library users at Penn State and UIUC. Both libraries are acquiring these titles on approval, lending support to the argument that approval plan profiles provide for the easy selection of a core group of titles from a core group of publishers. These data also help inform planning for the move to acquiring more electronic books. Acquiring e-books from publishers that already

**Table 7.** Top Ten Trade and University Press Publishers by Circulation for Penn State and UIUC

Penn State Top Trade Publishers			UIUC Top Trade Publishers		
	Circulations	% of Titles		Circulations	% of Titles
Wiley	2,472	82	Wiley	1,234	64
Routledge	2,146	74	Harcourt	947	63
Springer	1,854	68	Plagrave Macmillan	611	56
Plagrave Macmillan	1,194	63	Random House	557	33
Pearson	953	87	Routledge	552	61
Elsevier	884	67	CRC Press	444	72
CRC Press	643	76	Kluwer	393	65
McGraw-Hill	639	87	Springer	376	59
Blackwell	640	84	Simon & Schuster	342	66
Roman Littlefield	606	80	McGraw Hill	321	68
Penn State Top University Publishers			UIUC Top University Publishers		
Cambridge	2,129	77	Oxford	672	62
Oxford	1,697	68	Harvard	433	65
U. California	501	84	Princeton	354	64
SUNY	478	76	Yale	339	58
Yale	474	71	SUNY	333	53
Princeton	393	83	U. California	319	58
MIT	351	82	MIT	281	71
U. Chicago	335	71	Columbia	249	61
Teachers College	303	100	U. Chicago	247	70
Johns Hopkins	250	79	Indiana	247	67

provide libraries with the highest number of heavily used print titles makes sense.

As shown in table 7, Penn State and UIUC both have a high use of approval plan profile materials from different publishers. For example, 100 percent of the titles acquired from Teachers College Publishing by Penn State in FY05 circulated. The highest circulation rate (71 percent) for any one publisher at UIUC was for MIT titles. Determining the value of an approval plan profile by publisher is made easier by these data, showing selectors where they cannot go wrong by using a publisher-centric profile to automatically purchase books in demand by users. Conversely, the low end of this list, where none of the titles from a given publisher were used, can be considered for elimination from the profile. Figure 1 graphically represents the data from table 7, showing the number of books purchased by discipline and the correlating number of circulations that took place in FY05 and FY06.

### Unused Books

The primary goal of this study was to determine how well the two approval plan profiles were serving users at separate but comparable ARL libraries. The books that did not serve users well, or at all, also are a focus of this study. Varying percentages of noncirculations have been reported in previous studies in the literature. Hamaker reported no use for 54 percent of materials within the first four months following receipt.<sup>19</sup> Treadwell's study yielded very low noncirculation results (5 percent) that were based on circulation counts gathered eighteen months after receipt.<sup>20</sup> Hardesty at Eckerd College found that 33 percent of books acquired in a fiscal year had not circulated in a two-and-one-half- to three-year study period following acquisition.<sup>21</sup> The widely varying results of these earlier studies did not provide a true benchmark given the different methods and sizes of approval plans, institutions, and user populations.

This study found that 31 percent of Penn State's approval plan receipts did not circulate during the study period, resulting in \$217,382.70 spent on as-yet-unused books. At UIUC, 40 percent of approval plan receipts did not circulate during the study period, resulting in \$164,339.50 spent on as-yet-unused books.

The authors attributed the lower noncirculating title percentage at Penn State to the much larger user population. Books made available to larger audiences are more likely to garner higher circulation. However, universities often do not have the option to quickly expand, and thus the answer might be consortial purchasing and sharing, thereby gaining users along with a more diverse pool of books. This option will be the subject of future research at UIUC and the Consortium of Academic and Research Libraries in Illinois.

The cost of noncirculation was approximately 24 percent higher at Penn State largely because of the cost of cloth versus the paper-preferred option used at UIUC. However, the percentage of UIUC's approval budget used to acquire noncirculating titles was higher because of the higher percentage of titles that did not circulate (40 percent at UIUC versus 31 percent at Penn State). Both libraries view the relatively high noncirculation rate to be a sign that approval plan profiles need to be reviewed and adjusted. It is no longer economically feasible for large research libraries to acquire a certain percentage of books "just in case" a title might be needed in the future.

The circulation of the FY05 approval plan receipts for both Penn State and UIUC did not conform to Trueswell's 80/20 rule or to Kent's 80/10 hypothesis. Results from this study did not approach 80 percent use from 10 or 20 percent of the titles acquired; more than half of the circulation resulted from 20 percent of the approval receipts at both Penn State and UIUC. Specifically, 20 percent of approval plan books acquired by Penn State accounted for 59 percent of circulation of all approval plan books acquired during the study period. At UIUC, 20 percent of approval plan books accounted for 64 percent of circulation.

## Implications

### Penn State

Penn State's results indicated that 69 percent of approval receipts circulated within one to two years of receipt. The average number of circulations per book (2.85) and cost

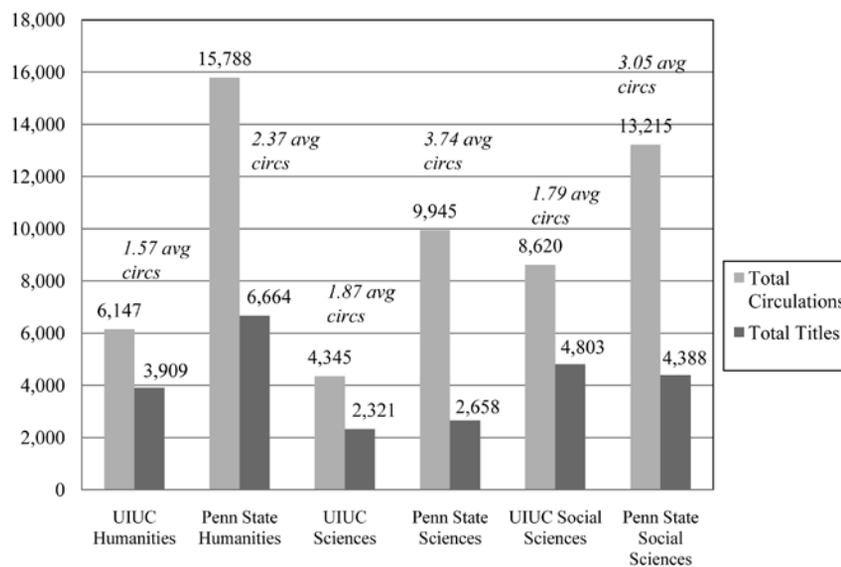


Figure 1. Graph Showing the Ratio of Books Circulating by Discipline for both Penn State and UIUC

per circulation (\$19.84) compares favorably to other studies, such as those conducted by Crotts and Rodriguez.<sup>22</sup> Because of the need to reallocate collection funds from print to support electronic resources and other collection development priorities, analyzing nonuse is as important as tracking use. It is crucial that the materials Penn State purchases match the needs of its users. This study determined which publishers and subjects supply significant percentages of books that either do not circulate or receive low circulation. These data reveal the effectiveness of approval plan profiles, which need to be regularly reviewed and updated to move automatic delivery of books to notification slips in some subject areas and for some publishers. Although automatic delivery of books is a time-saving method of acquiring large quantities of books, automatic delivery of books with a higher probability of not circulating is not cost-effective. This study provided the basis for future assessment of the approval plan profiles to contain costs while continuing to supply access to the books needed by the Penn State user community.

### UIUC

This research provided data on current approval plan profiles that showed cost/use and identified a core group of materials defined by subject and publisher. The results have already resulted in plans by UIUC to redesign profiles, track use, and modify the library's publisher list. More than 60 percent of approval selections were used within one to two years of availability, and science titles showed the highest circulation rate, with nearly 64 percent circulating and 1.87 uses per book. Science materials also had the highest use when compared

to other disciplines. However, the cost of books in all disciplines that did not circulate was high (more than \$164,000); 38 percent of the total approval budget was expended on unused books. UIUC will continue to address and research this issue, using these and similar data to create approval plan profiles that are more balanced toward materials that garner immediate interest and use. Subject selectors appreciated the data generated from this study and welcomed the information detailing the wants and needs of patrons. UIUC plans to conduct future assessments and hopes to follow up on the circulation frequencies of the 11,037 books purchased through the approval plan profiles in FY05 to determine if use declines, plateaus, or increases over time.

### Conclusion

Research questions posed for this study focused on measuring cost effectiveness and establishing a time frame and a method for conducting an approval plan profile assessment. The primary goal of this study was to determine what approval plan profile maintenance routines could be recommended to ensure that books being acquired meet users' current research and teaching needs. The results of this study clearly point to the need for regular assessment of the approval plan profiles and necessary adjustments based on user needs and fiscal constraints. Making informed decisions requires reliable use and cost/use data as well as benchmarks for comparing cost/use data. This study presents a cost/use per title range of \$19.83 to \$22.28 and a circulation per title range of 1.73 to 2.85. Although these data compare favorably to those found in the literature, the most important data concern titles that did not circulate in the period of twenty-one to thirty-three months from purchase. For research libraries, even one use of a single book can be considered a worthy purchase. But unused books can signal a disconnect with users, especially at the rate of 30 to 40 percent of an approval plan profile collection. Further studies need to be conducted on the profiles for these unwanted books, but for now the answer is to focus on use, specifically on the top publishers whose books are in demand and show high levels of circulation soon after purchase. High-circulation subject areas also must be supported because these areas show selectors where users are expressing their need. Data showed that for both Penn State and UIUC, the top ten publishers accounted for nearly 50 percent of books received as well as the highest levels of circulation. These data point to where support should be increased, perhaps diverted from that spent on low- or zero-use publishers and subjects.

The limitation of this study at both Penn State and UIUC was the difficulty in efficiently extracting the data needed to conduct an analysis. Even with the aid of online catalogs and vendor databases, finding, extracting, verifying,

collating, and analyzing these data is very labor intensive. The lack of order records for Penn State approval books led to the use of the 980 PromptCat data housed in the bibliographic record, which proved to be an imperfect but reasonable solution. The results, however, can be critical to the ability of a library to serve as a vital and current source of information for users. One answer to the problem of data collection and analysis is to partner with vendors. Vendors are in the unique position of having access to all the data needed to conduct these studies (other than local circulation information). Having the data delivered in a clear, consistent, and standardized format would streamline the entire process, leaving institutions to gather circulation data. The authors encourage vendors to create systems and databases to support the continued analysis of mutual and considerable investments. The goal is to develop mechanisms for regularly providing collection development librarians with the tools needed to make more informed decisions regarding the management of approval plans.

### Future Studies

Approval plan profile collection assessment must be ongoing, or at the very least regularly conducted, to monitor the large investments libraries make in approval plans. For this study in particular, continued analysis of circulation levels of the FY05 approval books will be conducted to determine if those titles that did not circulate were eventually used. If not, why did those books not circulate? Can patterns be identified? For those books that did circulate, did those books continue with relatively high use, or did use decrease over time? Which subjects garnered the most use over five, ten, or fifteen years?

Further research also will address the effect of user status and loan periods on circulation use. These data were collected during this study but have not yet been analyzed. Although this study provided important insight into circulation use across subject disciplines and publishers, additional research of specific publishers and titles is needed to determine patterns of use and nonuse. Is there an appreciable level of overlap of nonuse between UIUC and Penn State based on publisher? For example, what titles in common are not circulating? If approval plans between comparable academic libraries consist of high levels of duplication and percentages of nonuse, are large approval plan profile programs still viable given the need to support other collection priorities? Or are they an outmoded collection strategy given the changes in the economic climate? Further study is needed to determine the level of overlap between approval plans at academic libraries over time and the effect of virtual approval plans and consortial approval plans on collection development.

Finally, further exploration should be done to determine how libraries can better work with approval vendors to conduct assessment efficiently. Librarians must communicate their needs and state their expectations, knowing that a partnership in assessment will benefit both parties, resulting in a high use of books that serve the needs of our users.

### References

1. Richard W. Trueswell, "Some Behavioral Patterns of Library Users: The 80/20 Rule," *Wilson Library Bulletin* 43, no. 5 (1969): 458–61.
2. Allen Kent et al., *Use of Library Materials: The University of Pittsburgh Study* (New York: Marcel Dekker, 1979): 6.
3. Marcie Kingsley, "Circulation Statistics for Measuring Approval Plan Effectiveness," *Against the Grain* 8, no. 4 (1996): 1, 16–17; Marcie Kingsley, "Evaluating Approval Plans and Other Methods of Book Selection Through System Management Reports," *Acquisitions Librarian* 12, no. 24 (2000): 57–64; Denise Brush, "Circulation Analysis of an Engineering Monograph Approval Plan," *Collection Building* 26, no. 2 (2007): 59–62.
4. Kingsley, "Evaluating Approval Plans," 58.
5. Brush, "Circulation Analysis," 62.
6. Joe Crofts, "Subject Usage and Funding of Library Monographs," *College & Research Libraries* 60, no. 3 (1999): 265.
7. Marcos A. Rodriguez, "Knowledge Discovery in a Review of Monograph Acquisitions at an Academic Health Sciences Library" (master's thesis, University of North Carolina, 2008).
8. Dennis Dillon, "University of Texas EBook Study," (unpublished study, University of Texas, Austin, 2007).
9. J. M. Juran, "Universals in Management Planning and Controlling," *Management Review* 43, no. 11 (1954): 748–61.
10. Trueswell, "Some Behavioral Patterns of Library Users."
11. Kent et al., *Use of Library Materials*.
12. Larry Hardesty, "Use of Library Materials at a Small Liberal Arts College," *Library Research* 3, no. 3 (Fall 1981): 261–82.
13. Larry Hardesty, "Use of Library Materials at a Small Liberal Arts College: A Replication," *Collection Management* 10, no. 3–4 (1988): 61–80.
14. Charles A. Hamaker, "Some Measures of Cost Effectiveness in Library Collections," *Journal of Library Administration* 16, no. 3 (1992): 57–69.
15. Jane B. Treadwell, "Circulation Patterns of Recent Imprints: Implications for the Materials Budget" in *Building on the First Century: Proceedings of the Fifth National Conference of the Association of College and Research Libraries, Cincinnati, Ohio, April 5–8, 1989*, ed. Janice C. Fennell (Chicago: Association of College and Research Libraries, 1989): 230–34.
16. Jonathan D. Eldredge, "The Vital Few Meet the Trivial Many: Unexpected Use Patterns in a Monographs Collection," *Bulletin of the Medical Library Association* 86, no. 4 (Oct. 1998): 496–503.
17. Deborah Blecic, "Monograph Use at an Academic Health Sciences Library: The First Three Years of Shelf Life," *Bulletin of the Medical Library Association* 88, no. 2 (Apr. 2000): 145–51.
18. Robert F. Nardini, Charles M. Getchell, and Thomas E. Cheever, "Approval Plan Overlap: A Study of Four Libraries," *Acquisitions Librarian* 16, no. 8 (1996): 75–97.
19. Hamaker, "Some Measures of Cost Effectiveness in Library Collections."
20. Treadwell, "Circulation Patterns of Recent Imprints," 2311.
21. Hardesty, "Use of Library Materials at a Small Liberal Arts College," 65.
22. Crofts, "Subject Usage and Funding of Library Monographs," 265; Rodriguez, "Knowledge Discovery in a Review of Monograph Acquisitions at an Academic Health Sciences Library," 23.

# Google Books as a General Research Collection

By Edgar Jones

*The current study attempts to measure the extent to which “full view” volumes contained in Google Books constitute a viable generic research collection for works in the public domain, using as a reference collection the catalog of a major nineteenth-century research library and using as control collections—against which the reference catalog also would be searched—the online catalogs of two other major research libraries: one that was actively collecting during the same period and one that began actively collecting at a later date. A random sample of 398 entries was drawn from the Catalogue of the Library of the Boston Athenæum, 1807–1871, and searched against Google Books and the online catalogs of the two control collections to determine whether Google Books constituted such a viable general research collection.*

“There’s an east wind coming, Watson.”

“I think not, Holmes. It is very warm.”

“Good old Watson! You are the one fixed point in a changing age. There’s an east wind coming all the same, such a wind as never blew on England yet. It will be cold and bitter, Watson, and a good many of us may wither before its blast. But it’s God’s own wind none the less, and a cleaner, better, stronger land will lie in the sunshine when the storm has cleared.”

—Arthur Conan Doyle, *His Last Bow*

On December 14, 2004, Google announced that it had concluded agreements with five major research libraries to begin what is now known as the Google Books Library Project.<sup>1</sup> The libraries—the so-called Google 5—were the New York Public Library and the libraries of Harvard, Michigan, Oxford, and Stanford universities. These libraries agreed to let Google digitize volumes from their printed book and serial collections in exchange for institutional copies of the digitized volumes.<sup>2</sup> While the agreements set broad parameters for cooperation, Google gave the libraries sole discretion in determining the volumes to be digitized.

The Library Project—and the discretion given the libraries in determining which volumes would be digitized—raises an interesting question: To what extent is Google creating a research collection? Coyle has suggested that the manner in which collections are being selected for inclusion in the Library Project—many being taken *en bloc* from low-use remote storage facilities—makes it difficult to characterize Google Books as a “collection” in the accepted sense, though for better or worse “it will become a de facto collection because people will begin using it for research.”<sup>3</sup> Is this true? Is this testable? Can sheer volume, in fact, render

**Edgar Jones** (ejones@nu.edu) is Bibliographic and Metadata Services Coordinator, National University Library, San Diego, California.

Submitted July 7, 2009; tentatively accepted September 17, pending modest revision; revised and resubmitted September 22 and accepted for publication.

moot the role of selection in this case.<sup>3</sup> The current study attempted to answer these questions.

While the focus of this study was on content digitized by Google through 2008, one should keep in mind that the volume of available digitized content continues to grow. Since the initial Google 5 cooperative agreements at the end of 2004, Google has entered into agreements with an increasing number of research libraries, both in the United States and abroad, while the European Union has begun funding a digitization program of its own centered on the collections of European cultural heritage institutions (libraries, archives, and museums).<sup>4</sup> Initially, there also was competition from elsewhere in the commercial arena, but this proved to be comparatively short-lived. Within a year of the Google announcement, Microsoft, in cooperation with the Internet Archive, began to digitize print content from several libraries under the rubric of Live Search Books. In May 2008 this effort was abandoned, though content already digitized under that program—some 750,000 volumes—remained available via the Internet Archive.<sup>5</sup>

In terms of scope, several of the Library Project partnerships cover both older public domain materials and more recent publications still subject to copyright protection. To this extent they complement Google's partnerships with publishers to provide access to a continuity of content across time periods.

This continuity of content is important from Google's perspective. In Google's December 2004 press release, cofounder Larry Page set the Library Project in the context of his firm's stated mission "to organize the world's information and make it universally accessible and useful."<sup>6</sup> As a search engine, Google's principle interest in digitizing printed materials is in indexing the content, both structured and unstructured, to enhance search results. In its business model, Google uses search terms and results as triggers for the online display of related advertising. By providing additional indexed content from Google Books (and the Library Project), Google both increases the usefulness of its flagship search engine (by incorporating results from Google Books as well as other sources) and makes it more appealing to advertisers (by increasing the potential customer base to include researchers and other interested parties).

As has been noted frequently, Google is digitizing on an industrial scale, indeed on a scale unlike anything seen before.<sup>7</sup> The process is easy to describe. Books are removed from the shelves, barcodes are scanned—to change the volume's circulation status and to extract the related metadata from the catalog—and the volumes are removed to a Google facility for digitization. Google digitizes the individual page, subjects the digitized images to sophisticated (if not foolproof) optical character recognition (OCR), and finally indexes the OCR-extracted text. The digitized page images may be freely available for public viewing (if determined to be in the public domain), or viewing may be restricted in

some way, depending on the copyright status of the digitized work (or one or more of its components) and the nature of Google's agreements with the publisher.<sup>8</sup>

### What Is a Research Collection?

While other digitization programs on various scales also are under way (as noted above), none approaches the scale or ambition (or potential for market dominance) of Google Books. For this reason the volumes digitized by Google seemed the most appropriate objects of which to ask: Do these digitized volumes in themselves now constitute a viable general research collection? This may seem a fairly straightforward question, but it raises an antecedent question: What is a research collection?

In the abstract, a research collection is a collection of materials used primarily to support research (as opposed to one that supports teaching and learning or one that is used primarily for recreational purposes). Unfortunately, this definition does not lend itself to objective measurement, and it says nothing about the content of such a collection, since, in theory, *any* collection can support research of *some* kind.

Indeed, most research collections are developed to address the needs of a particular research community, a community that will typically reflect a variety of research interests and intensities. This variety will itself change over time. Research collections are by their nature complex. Such complexity underpins the design of the Conspectus model developed in the early 1980s by the Research Libraries Group for cooperative collection development. The Conspectus asked participants to characterize their collections according to a variety of parameters, including research area (defined by ranges within a bibliographic classification scheme or by subject descriptors), language, geographical scope, chronological periods, formats, and collection depth (this on a five-point scale, with 4 indicating "research level").<sup>9</sup> This produced a nice matrix for describing the variety of possible research collections, but it also made clear that the idea of a "generic" research collection was an oxymoron.

Ideally, in addition to being designed for a particular research community, a research collection also satisfies the needs of that community. But while research collections can come in a variety of shapes and sizes, data suggest that whatever their shape and size, local researchers will always feel that their own library's collection falls short—this despite years of earnest collection development by librarians. At member institutions of the Association of Research Libraries, for example, respondents to successive LibQual+ library service quality surveys routinely report that their libraries provide inadequate support for their research needs. On three LibQual+ items measuring collection support for research, the "adequacy gap"—the degree to which

an item exceeds (or not) a user's minimum requirements—has typically been a negative number.<sup>10</sup> To put as generous a spin as possible on the meaning of these LibQual+ responses, one could say that research collections always must be works in progress.

Although there may be no such thing as a typical research collection, at major research universities—where the research communities are larger and more multidisciplinary—a certain amount of homogeneity can be expected to develop across the associated research collections. It is not unreasonable then to treat one of these collections as approximating a “generic” large university research collection.

Having posited that the collections of large research libraries approximate to a generic research collection, a question remains about how much unique content is found in a typical research library. No one knows for sure, but overlap studies suggest it is more than one might expect.<sup>11</sup> In a 2005 study (shortly after the announcement of the Google Books Library Project) Lavoie, Connaway, and Dempsey determined that the Google 5 libraries collectively held about one-third of the resources cataloged as books in the OCLC WorldCat database, and most of these resources—61 percent—were unique to just one of the five. This percentage of uniquely held resources increased with the age of the resources involved, with 74 percent of resources published between 1801 and 1825 being held by just one Google 5 library.<sup>12</sup> However, there is unique and then there is *unique*. In subsequent research, Lavoie and Schonfeld examined a random sample of one hundred WorldCat records for such “uniquely held” resources and found that “many of the English language materials appear to be locally-produced ephemera rather than traditional published books.”<sup>13</sup> This suggests that unique holdings may be less of a problem for the idea of a “generic” research collection than might otherwise appear.

Given its existing digitization agreements, it seems reasonable to expect that Google Books will eventually become a generic research collection in this sense. It will at the very least become the University of Michigan Library, since Michigan has agreed to digitize its entire library (aside from Special Collections).<sup>14</sup>

### Scope of the Study: The Public Domain

This study is limited to materials that would have been viewable in full in Google Books in 2008 under current copyright law, that is, materials then in the public domain. This was a pragmatic decision in that such materials are the only ones that can be demonstrably shown to have been digitized. But it also attempted to address the idea of the research collection as implied in Coyle's original comment—something that would be used for research.<sup>15</sup>

The author also felt that restricting the study to public

domain materials would produce a more conservative estimate of overlap than would be true for more recent imprints, given the participation of publishers as well as libraries for this material in Google Books, the growth and expansion of American research libraries over the last century, and Lavoie, Connaway, and Dempsey's findings of increasing overlap within WorldCat as imprints become more recent.<sup>16</sup> This suggested that whatever conclusions the study reached about materials in the public domain would apply with even more force to materials still covered by copyright should those materials be exposed to viewing via Google Books at some future point.

How big is the public domain? Nobody knows. The online English Short Title Catalogue, covering pre-1801 materials published chiefly in the British Empire and the United States, had accumulated nearly half a million entries by mid-2009.<sup>17</sup> More generally, Buringh and Van Zanden estimated that more than 1,750,000 books (defined as fifty pages or more) were published in Western Europe prior to 1801.<sup>18</sup> To bring this closer to the present, in a 2006 examination of books in OCLC WorldCat, Lavoie and Schonfeld found that roughly six million books (18 percent of the books then in WorldCat) were published prior to 1923 (a rough indicator of public domain status in 2008).<sup>19</sup> The chart accompanying the Lavoie and Schonfeld article suggested a steady rate of increase in publishing output throughout the nineteenth century.

But as anyone who has tried to establish the copyright status of a book knows, public domain status is not simply a question of date of publication, and Google's caution increases as the date of publication approaches 1923 (with digitized volumes more likely to offer “no preview,” “snippet,” or “limited preview” access rather than “full view”). For that reason, the current study adopted a more conservative interpretation of public domain, looking for a cutoff farther back in the nineteenth century.

Given this working definition of the public domain, one question that still needs to be addressed is that of page image quality. Scholars have not been silent on this question, and anecdotal evidence suggests the frequency of poor page image quality is not insignificant, though the solution is by no means clear.<sup>20</sup> Quality control of scanned texts, unlike the initial scanning, cannot be done on an industrial scale and is labor-intensive. In theory, however, volumes to be rescanned could be prioritized on the basis of user complaints. A “flag this page” feature, presumably for this purpose, was present in earlier versions of Google Books but had been discontinued at the time of writing. Similarly, the problem of nontextual content (illustrations, maps, etc.) in the digitized volumes, while not significant for indexing, is significant for research, especially in the case of folded material, which Google scanners typically digitize in its folded form.<sup>21</sup>

Given both this more restricted definition of the public domain and the caveats regarding image quality and folded

materials, the current study attempted to measure what proportion of the volumes in a generic research collection (as described above) would be present in digitized form in Google Books, using as a reference collection the catalog of a major research library that was actively collecting during the period, and using as control collections—against which the reference catalog also would be searched—the online catalogs of two other major research libraries, one actively collecting during the same time and one that began actively collecting at a later date.

The general research question can be stated formally as follows: Given that A and B were major American research libraries during the nineteenth century, while C became one during the same period, is there a greater probability of manifestation-level overlap between A and Google Books, between A and B, or between A and C? The question can be answered by testing two hypotheses:

1. A larger proportion of a random sample of entries drawn from A's catalog will be found in Google Books than will be found in the online catalog of B.
2. A larger proportion of a random sample of entries drawn from A's catalog will be found in Google Books than will be found in the online catalog of C.

## Experimental Design

The experiment involved six steps:

1. identifying an appropriate reference collection
2. extracting a random sample of entries from the reference collection
3. searching the metadata from those entries against Google Books for matching manifestations
4. identifying appropriate control collections
5. searching the metadata from the reference collection entries against the OPACs of the control collections
6. recording the results and performing appropriate statistical tests

### Identifying the Reference Collection

What would serve as an acceptable reference collection? More precisely, what contemporary collection would meet the criteria for a major American research collection? Only the largest libraries of the nineteenth century might lay claim to such a distinction, and of these it would be necessary to select one with a reliable printed catalog to efficiently select a random sample of entries for testing.

Identifying the major American research libraries of the nineteenth century is not difficult, but reliable comparative statistics are rare. As noted in the landmark *Public Libraries in the United States*, for much of the nineteenth century

library statistics were collected and published only occasionally, and reports were often of dubious reliability.<sup>22</sup> In his 1851 survey, Charles Jewett, then librarian of the Smithsonian Institution, listed five libraries as having collections of at least 50,000 volumes: Harvard University, Yale College, the Philadelphia Library (including the Loganian Library), the Library of Congress, and the Boston Athenæum. Harvard's libraries collectively held the largest number of books at the time: 84,200 volumes.<sup>23</sup> Less than two decades later the largest libraries had grown—Justin Winsor's survey in 1868–69 reported that there were now six libraries containing at least 100,000 volumes—but the ranks had changed. Half of the six largest were newcomers (Winsor's own Boston Public Library and the Astor and Mercantile libraries in New York City). Only the Library of Congress, Harvard University, and the Boston Athenæum continued from Jewett's earlier list.<sup>24</sup> And of these three, only the Boston Athenæum had a contemporary published catalog that could serve as a reliable data source for the current study: the five-volume catalog compiled under the direction of Charles Ammi Cutter and published between 1874 and 1882.<sup>25</sup>

The use of any library catalog as a surrogate for a generic authoritative research library will be biased to the extent that it necessarily reflects the collection interests of the library involved. In the case of the Boston Athenæum, this will be a bias toward topics of interest to its members and toward printed books available since the library was founded in 1807 (leading to an underrepresentation both of non-American imprints and of older books), as well as a slight regional bias. However, against this unavoidable bias must be set the clear intention of the founders to create an exceptional general research library “containing the great works of learning and science in all languages, particularly such rare and expensive publications as are not generally to be obtained in this country.”<sup>26</sup> If one accepts the judgment of contemporaries, then the founders achieved their purpose. According to Jewett, the Athenæum library was “hardly surpassed either in size or in value by any other in the country.”<sup>27</sup> According to Edwards, it “[stood] saliently out from amongst its compeers alike for its extent, its liberality of access, [and] its richness in departments not usually well-filled in American Libraries.”<sup>28</sup> So while taking the Boston Athenæum as representative of a generic major research library in some absolute sense may not be possible, it can be described with confidence as one of the major American research libraries of that time.

Cutter's catalog was said to include 92,000 volumes and 36,000 pamphlets.<sup>29</sup> It does not include works from the fifty public domain years that followed its publication, and given the increase in research collection size and overlap during this period observed by Lavoie et al., the use of Cutter's catalog in the current study will tend to underestimate the actual degree of overlap between the reference collection and Google Books, i.e., the reported (pre-1872) hit rate

in Google Books will likely be less than it would be for the entire (pre-1924) “public domain” period.

### Extracting a Random Sample from the Reference Collection

The objects of interest were manifestations represented by main entries in the *Catalogue of the Library of the Boston Athenæum, 1807–1871*.<sup>30</sup> While Cutter did not go so far as Panizzi, who included his cataloging rules in the first volume of his aborted 1841 *Catalogue of Printed Books in the British Museum*, he did include one page of detailed explanations of the organization and structure of entries in the catalog, including the determination of the main entry.<sup>31</sup> Cutter entered works of personal authorship under the name of the author, collections under the name of the editor, and works of corporate responsibility (interpreted broadly) either under the name of the body, territorial authority, or—for local institutions—under the name of the place where it was located.

The five volumes of Cutter’s catalog are numbered consecutively, and the author of this study used a random number generator to create the sample of pages from which main entries would be extracted. For each page in the sample, the first main entry appearing on the page was selected. In cases where no main entry appeared on a page (for example, the page consisted entirely of subject entries), the first main entry appearing on a subsequent page was selected. While Cutter made entries for the component parts of aggregate works, these were ignored in selecting the sample unless they represented physically distinct volumes (rather than contributions to a single volume) to restrict the sample to physically substantial works. Similarly, physically separate volumes of fewer than fifty pages were omitted from the sample.

### Searching the Sample in Google Books

The author used the metadata from the sample entries in Cutter’s catalog—specifically the author (when present) and title—to search Google Books for corresponding entries. In cases where a corresponding entry was not found in Google Books, the Boston Athenæum’s online public access catalog Athena (<http://catalog.bostonathenaeum.org>) was searched on the chance that more complete metadata might be located. If no corresponding entry was found in Athena, the lists of corrections at the beginning of each of the first four volumes of Cutter’s catalog (and the end of the fifth) were searched for unsuspected typos and similar errors.

In cases of aggregate works—multivolume monographs and serials—the author deemed a manifestation to be a match if digitized images of at least 50 percent of its component physical volumes were present in Google Books; otherwise it was deemed to be absent from Google Books.

In terms of Functional Requirements for Bibliographic Records (FRBR) Group 1 entities, Google Books is technically not a collection of manifestations but rather of images of items (specific exemplars) that have been digitized from participating library collections.<sup>30</sup> This anomaly is recognized in the FRBR discussion of reproductions, but it may have practical consequences in that a given alteration to a specific item may add or subtract value from a given manifestation, depending on whether the alterations are positive in form (e.g., scholarly annotation) or negative (e.g., vandalism).<sup>33</sup>

While in many cases a given manifestation will not be represented in Google Books in digitized form, in other cases it will be represented multiple times by exemplars from different institutions (and sometimes from the same institution). This is an unavoidable consequence of digitization undertaken on such a scale and on such an accelerated timetable. The current study did not attempt to count such instances of multiple digitized items, though instances were not uncommon.

The author considered a digitized volume in Google Books a match if it included all elements of the manifestation as described in Cutter’s catalog (augmented when necessary by elements from the description of the same manifestation in the Boston Athenæum’s Athena online catalog or other sources). Digitized volumes that appeared to represent the same text (what FRBR calls an expression) were rejected if they differed on any element of bibliographic description; that is, identity of container took precedence over identity of content. The author felt this was necessary to ensure rigorous matching criteria, since differences in text from one manifestation to another—signaling a different expression—are not always readily apparent from bibliographic data. One consequence was that legislative documents represented singly or in small groups in Cutter’s catalog could not be linked to the bound volumes that superseded them in the collections of the Google Books partner libraries.

### Identifying the Control Collections

The author also used the metadata from the sample entries in Cutter’s catalog to search entries in the OPACs of two major research libraries (defined as Association of Research Libraries member institutions with reported holdings greater than eight million volumes).<sup>34</sup> One of these selected institutions was actively collecting during the same period as the Boston Athenæum, while the other began collecting at a later date.

### Searching the Sample in the Control Collections

In searching the OPACs of the control collections, the author defined a match as a record representing the same manifestation or a reproduction of that manifestation as defined in *FRBR*.<sup>35</sup> The decision to include reproductions as

matches was made on the basis that Google Books manifestations are de facto reproductions, and reproductions in the control collections should therefore be treated as equivalent. In some cases, fidelity to the original might not be perfect, particularly in the case of originals that included color versus a reproduction made in monochrome (typically a microform copy or a digitized copy made from a microform copy).

In many cases with older books, research libraries have purchased collections of books listed in a given printed bibliography (e.g., Charles Evans' *American Bibliography* (1639–1820), originally supplied by Readex Microprint Corporation on micro-opaque, then on microfiches, and now online).<sup>36</sup> In the current study, items in such collections were “matches” only if they were represented by records in the library's OPAC. This inevitably resulted in undercounting, especially in the case of the control library that began its collecting during a later period than the Boston Athenæum. The author made the assumption in this case that a resource that was not represented by a record in the OPAC would be invisible to most library users seeking it and perceived as not being held by the library concerned. For a user to track down a desired resource in such cases would require a detailed knowledge of bibliography, the identity of major microform and digital collections, and the highways and byways of the website of the library concerned—a knowledge far beyond that of most library users.

As with Google Books, the author considered a manifestation of aggregate works a match if digitized images of at least 50 percent of its component bibliographic volumes were present in a control collection as determined from its OPAC; otherwise it was deemed to be absent from the control collection. In those rare instances where the extent of holdings of an aggregate work could not be determined from data in the library's OPAC, it was deemed to be a match for the purposes of the study. This decision was based on an extrapolation of the pattern for other aggregate works in the sample where the control libraries typically owned more than half of the bibliographic volumes concerned. In cases where sample entries represented component parts of aggregate works (e.g., an analyzable component of a series), the author searched the OPAC under indexed elements (creator, title, and so on) for both the component work and the aggregate work.

### Performing the Statistical Tests

Once the author searched Google Books and the OPACs of the control institutions for the same sample entries, estimates were made of the proportion of manifestations represented by entries in Cutter's catalog that were likely to be found in (a) Google Books, (b) the collection of the older control institution, and (c) the collection of the later control institution. Z-tests were then performed to determine

whether the proportion of entries found in Google Books was significantly different from the proportion found in either of the control collections.<sup>37</sup>

## The Process

Problems encountered during the current study can be subdivided into problems determining the extent of the manifestation represented by an entry in Cutter's catalog, identifying a manifestation in Cutter's catalog, locating a manifestation in Google Books, and identifying manifestations in the catalog of the control collections. These are discussed below.

### Extent of an Item

Cutter's catalog includes entries for both books and pamphlets. The inclusion of pamphlets makes it an unusually thorough catalog. Unfortunately, it also makes it too unusual to justify including pamphlets in the current study. Because of their brevity and their often ephemeral nature, pamphlets are seldom cataloged at the level of the individual pamphlet, so identifying their presence or absence unequivocally in Google Books or the control collections would have been problematic.

To exclude pamphlets on a consistent basis, the study adopted a modified version of the UNESCO definition of “book”: “a non-periodical printed publication of at least 49 pages, exclusive of cover pages . . . made available to the public.”<sup>38</sup> For purposes of the study, the author modified the UNESCO definition to include periodicals, even when shorter than forty-nine pages. Application of the definition in practice was complicated by the fact that Cutter seldom recorded the extent of an item in his catalog (other than for multipart items). Overcoming this obstacle required determining an item's extent by searching Athena, the online catalog of the Boston Athenæum, or occasionally the online catalog of the Countway Library of Medicine at Harvard University (where the library of the Boston Medical Society—formerly in the Boston Athenæum—currently resides) for a fuller bibliographic record. Metadata in Google Books turned out to be unreliable in this regard, as different Google Books metadata purporting to describe the same manifestation occasionally presented contradictory data as to the manifestation's extent.

### Identifying a Manifestation in Cutter's Catalog

One difficulty encountered when examining entries in Cutter's catalog resulted from the extensive manipulation of title data by the cataloger preparing the entry. This manipulation was necessary to take full advantage of the different

environment constituted by a closed system such as Cutter's book catalog (compared to the open systems of today). In a book catalog, entries are interpreted in the context of the surrounding entries, and in the best such catalogs they are crafted to exploit that context to the utmost. Cutter's catalog reduced titles to the extent that they conveyed the maximum amount of information with a minimum amount of text, eliminating text that was redundant with the introductory heading and also text that added nothing of substance to the title. For example, Cutter reduced "Catalogue de la riche bibliothèque de D. José Maria Andrade" to "Andrade, José Maria. Catalogue de [sa] bibliothèque," and he reduced "Report of the debates in the Convention of California on the Formation of the State Constitution, in September and October, 1849" to "California. Constitutional Convention, 1849. Report of debates."

Cutter assumed the users of books in foreign languages would be familiar with those languages, an assumption that gave him greater latitude in his manipulation of the entries but that presents some problems to the modern researcher. This is especially the case with titles in Greek or Latin, where the elimination of introductory words or phrases can often alter the grammatical cases of the remaining title words. Being able to reconstruct the title as it appears on the item facilitates searching in Google Books (always providing the title has not been similarly, or differently, abridged in the metadata supplied to Google).

In rare instances, Cutter's catalog altered a title to make it more descriptive of a book's content. For example, Edward T. Channing's *Lectures Read to the Seniors in Harvard College* was rendered by Cutter as "Lectures on rhetoric and oratory; biographical notice by R. H. Dana." Such cases presented particular challenges when searching Google Books or the OPACs of the control collections.

Like all catalogs, Cutter's catalog was of course subject to error. Cutter's staff were not immune to the odd typographical or other error, and Cutter himself did not catch all of these (as witness the multiple pages of corrections appended to each volume after the first). Whenever possible, the author verified entries in Cutter's catalog against bibliographic records in Athena. In rare cases, the volumes listed were no longer owned by the Athenæum.

### Locating a Manifestation in Google Books

There was no problem in identifying the sample manifestations once they were located because Google Books contains digitized page images, including the title pages. There were, however, problems *locating* the sample manifestations.

If the title was not distinctive, the search result might include both volumes from the target publication and volumes from other publications with the same title mixed together randomly. For example, clicking on the "other

editions" link from "Report of the Treasurer" in Google Books produced a list that included reports from the treasurers of Alabama, Connecticut, Maine, and several other states. This was somewhat ameliorated by the fact that Google Books initially sorts "other editions" by date of publication.

If the title was subject to varying treatment ("cataloger judgment") by the originating catalogers or by the cataloging rules in force at the time the item was cataloged locally, the same manifestation might be scattered across the pages of a Google Books search result. For example, of the four digitized sets of Cutter's Boston Athenæum catalog in Google Books at the time of the current study, the metadata supplied by Harvard and New York Public Library used the full title from the title page (*Catalogue of the Library of the Boston Athenæum: 1807–1871*) while the Michigan metadata omitted the terminal dates and the Oxford metadata (taken from the Bodleian pre-1920 catalog) followed older cataloging practice in omitting the text that was redundant with the principal access point by reducing the title to "*Catalogue . . . 1807–1871*." Had the Oxford copy been the only one digitized, locating the manifestation in Google Books would have been problematic. When searched in early November 2008 using the "Advanced Book Search" page to limit by viewability and dates of publication, the Oxford copy appeared on page seven of ten in the result set.

To be weighed against this confusion is the serendipity resulting from the acts of digitization and indexing. In one case, a particular manifestation was discovered at the end of a digitized volume, bound with a different manifestation: Oxford University's copy of Henri Storch's *Considérations sur la nature du revenu national* appears after page 428 of volume 4 of his *Exposition des principes qui déterminent la prospérité des nations*. It was located not by searching against the Google Books metadata but by searching against the full text. It was not listed in Oxford's OPAC, suggesting that in this case the Google digitization may have turned up a hidden work in Oxford's collection.

Similarly, Google Books may serendipitously discover works that are significant at the item level. In one case from the sample, an item's provenance could be traced from the author to then-Vice President John Adams to the American Academy of Arts and Sciences and finally to Harvard University, with some minor annotations in Adams' hand. The provenance was unremarked in HOLLIS (<http://discovery.lib.harvard.edu>), the Harvard online catalog, where the item was in the general circulating collection.

The author used the metadata in Google Books sparingly, since its reliability was open to question. Limiting a search by certain metadata elements occasionally produced anomalous results. For example, a search for "full view" books published in German between 1807 and 1871 also returned some books in English. Likewise, a search for

full-text books published in 1749 produced an initial result of 2,311 books, but after paging through the results, this ultimately resolved to 227 volumes, presumably through the progressive elimination of duplicate entries. Curiously, Latin was not available as a choice for limiting a search by language, though this will presumably be changed as increasing numbers of pre-1801 imprints are added to Google Books.

Result sets from Google Books must be examined with care. For example, a search for A. O. Abbott's *Prison Life in the South* with the result set restricted to full-view manifestations does not retrieve a matching manifestation in the first page of results but rather in the second. Google Books can likewise be unforgiving of faulty metadata. For example, a search for E. S. Abdy's *Journal of a Residence and Tour in the United States, 1833-34* produces a matching full-view manifestation only if one removes the "-34"—reduced from "1834" in the Boston Athenæum catalog—from the search argument: a case where "less is more" in a search argument.

#### Identifying a Manifestation in the Catalog of the Control Collections

Very few library catalogs have not undergone retrospective conversion to machine-readable form at some point over the last several decades. Such retrospective conversions often carry two very large caveats. They may represent the conversion of files other than the official catalog, and they may have been outsourced to private firms on terms that would convert the largest number of records at the lowest cost.

The records resulting from such conversion are often incomplete, sometimes to such an extent that the resource represented by the record cannot be identified with certainty. Perhaps the most famous (or notorious) example of this is the PREMARC records in the OPAC of the Library of Congress (LC), the products of a conversion of records from the LC's old (pre-AACR2) shelflist. PREMARC records have a high transcription error rate (which the LC estimates at 15 percent for call numbers), and the contractor instructions allowed for the routine omission of subtitles (interpreted broadly by the contractor), contents notes, and series, with results that were often less than helpful.<sup>39</sup> For example, browsing the LC online catalog under "Fourteenth Census of the United States" returns dozens of records with this title—and only this title—and on which the only distinguishing features are varying paginations and hints from the variety of subject headings assigned.

Similar conversions have taken place at most research libraries, including the ones that served as the control collection for this study. Fortunately, incomplete records in the online catalog of the control collections were only an occasional problem. Nevertheless, in cases where there was

insufficient bibliographic data, the author needed to come up with a rule of thumb to determine whether a particular catalog record represented the manifestation being sought. Given that the object of the current study was to determine whether the number of matches in Google Books equaled or exceeded the number in the control collections, the benefit of the doubt was given to the control collections in these cases. The manifestation being described could not be determined in just three of these cases.

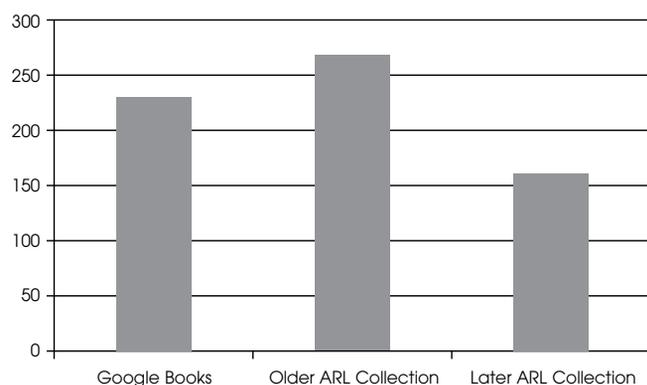
Finally, authority control remains imperfect among retrospectively converted records in online library catalogs, where records from different files—some authority controlled, some not—may have been merged. It was not uncommon when searching the catalog of a control collection to find the same person or corporate body represented by three or more headings, often differing from one another only very slightly. Unfortunately, while the differences were often barely noticeable to a human reader, a miss was as good as a mile to a machine, which duly segregated them under discrete headings.

#### Results

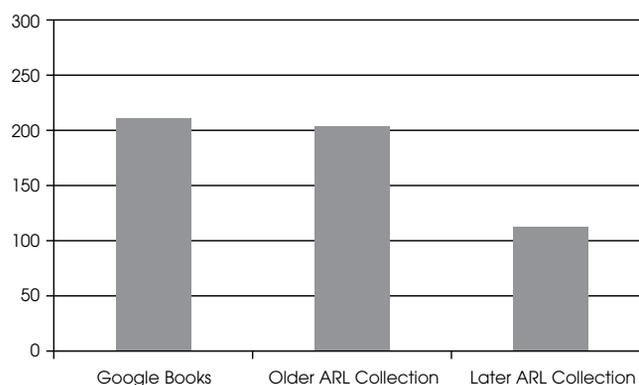
The author found digitized items (matching "full view" books) representing 235 entries in Google Books out of 398 in the sample from Cutter's catalog. The match rate for Google Books was 59.05 percent  $\pm$  4.83 percent. The corresponding match rate for the older research library (268 matches) was 67.34 percent  $\pm$  4.61 percent and for the later research library (162 matches) 40.7 percent  $\pm$  4.83 percent. All rates were significant at the 95 percent confidence level. The Z-tests comparing the Google Books results with those of each of the control libraries were significant in both cases (Google Books versus the older research library:  $Z = 2.351$ ; confidence level  $> 95$  percent; versus the later research library:  $Z = 5.106$ ; confidence level  $> 99$  percent). These data are shown in figure 1.

The sample included 98 entries representing pre-1801 imprints. Google Books and the two control collections differed markedly in their match rates for these materials. Only 27 were found in Google Books, while 66 were found in the "contemporary" research library and 48 were found in the "later" research library.

For post-1800 imprints, Google Books had a slightly higher match rate than the contemporary research library and a significantly higher rate than the later library (see figure 2). Out of 300 entries in the sample representing post-1800 imprints, 208 were found in Google Books while 202 were found in the contemporary library's OPAC ( $Z = -0.439$ ; confidence level  $> 95$  percent) and 114 in the later library's OPAC ( $Z = 7.613$ ; confidence level  $> 99$  percent). (See figure 2.)



**Figure 1.** Number of Matches with Boston Athenæum Catalog Sample of 300 Titles



**Figure 2.** Number of Matches with 398 Post-1800 Sample Entries from Boston Athenæum Catalog

## Discussion

The study employed a very strict definition of manifestation, requiring matching resources to conform in all bibliographic details with the entries in Cutter's Boston Athenæum catalog. This meant that in several cases a seemingly identical text—an expression in FRBR terms—was rejected as a match. Consequently, the match rates reported in this study should be treated as floors in terms of matching expressions.

In a similar vein, the author makes no judgments as to the continued value of the expression embodied in a given manifestation vis-à-vis a later expression, one that might have superseded that earlier expression in the Athenæum catalog. A library that owned a manifestation of that later expression but not the manifestation represented by the Athenæum catalog entry was deemed not to own a match. As an example of this phenomenon, the Athenæum catalog contained a four-volume collection of the works of William H. Seward published during his lifetime, while another library had a more complete set published posthumously. Presumably the later set was both more complete and more accurate from a scholarly perspective, but with no way to determine this objectively, it was rejected as a match in this study. Again, the result is an undercount of what scholars might consider to be matches.

Beyond these effects of the experimental design on match rates, some differences were the product of the design of the digitization projects themselves. The observed differences in Google Books match rates between pre-1801 imprints and post-1800 imprints is largely explained by the collections that are being selected for mass digitization. Initially, these have tended to be low-use items in remote storage facilities. American libraries often treat 1801 as a heuristic cutoff point for rare books, with books published before that year routinely segregated to rare book libraries or similar facilities not subject to mass digitization.<sup>40</sup> When

pre-1801 imprints have been digitized, the volumes have occurred in the general collections, especially in the general collections of European partner libraries. Of the twenty-seven pre-1801 sample imprints found in Google Books, fourteen were digitized from European collections.

While the digitization of pre-1801 imprints would be attractive to Google in terms of increasing the comprehensiveness of Google Books, research libraries have less incentive. Many of the titles involved are already accessible online to these institutions on a subscription basis from vendors that have digitized preexisting proprietary microform collections of old books (Eighteenth Century Collections Online, Early American Imprints, etc.), so bringing them into the mass digitization stream would not necessarily increase the digital content available to their researchers. Of course, this does not rule out digitization of these materials at a later time.

One side effect of indiscriminately digitizing great swathes of the print collections of large research libraries is an increasing rate of duplication as more collections are digitized, and a decreasing marginal rate of return for each newly digitized collection. For example, Lavoie estimated that while 58 percent of the holdings (approximately 18 million) of the original Google 5 were unique to a single institution, this would likely be true for only 22 percent of the holdings (approximately 8 million) of an additional Google 5.<sup>41</sup>

However, increasing duplication can introduce potential benefits in some contexts. While the study encountered many volume-level duplications, the study itself was not designed to measure this duplication and made no attempt to do so. Under ordinary conditions, one would want to keep such duplication to a minimum. But in a world where the power and storage capabilities of computers continue to increase at a staggering rate, this approach may no longer be valid, and a certain amount of redundancy may actually

present advantages. In a slightly different context—that of translating a text from one language to another—Garreau observed in *The Washington Post*,

The explosion of the Web . . . has enabled a revolution. Like so many successful human approaches, it relies on brute force and ignorance. This method cares little for how any language works. It just looks—Rosetta stone fashion—at huge amounts of text translated into different languages by humans. (Dump decades of U.N. documents into the maw.) Then it lets the machine statistically express the probability that words in one language line up together in a fashion comparable to another set of words in another language.<sup>42</sup>

In the context of Google Books, item-level duplication feeds into this “brute force and ignorance” method of machine-based quality control, especially as an alternative to more labor-intensive page-by-page human quality control (now mainly the responsibility of the partner libraries). One can easily imagine, for example, a two-step probabilistic method whereby Google first identifies identical images (setting a very high bar for a match in terms of visual pattern recognition), then selects from them the image from which optical character recognition (OCR) has produced the most satisfactory rendering into plain text (taking this as evidence of a relatively clear image of the original). This would simultaneously both reduce the number of duplicate volumes in Google Books and the number of poorly scanned images that would otherwise require manual intervention. A simpler process might be used to enable the supplying of missing pages when comparing two otherwise identical sequences of page images.

From the point of view of scholarly research, poor image quality and occasional missing pages may be less of a problem pragmatically than it is absolutely (as a guarantor of textual integrity). Researchers—the intended audience of a research library—are typically interested in books not as artifacts of cultural heritage or even necessarily as integral texts, but rather as containers of certain desired content. Consequently, most of them read books—at least online—only to the extent necessary to extract the desired content, and this may not be reading as we know it. This is seen, for example, in a recent large-scale survey of British academics, where most respondents reported that they “dipped in and out of several chapters” when reading e-books rather than reading continuously.<sup>43</sup> Given this user behavior, Google’s post-scanning priorities might not necessarily be in image quality in general (except to the extent that this affects the ability of their software to recognize and index text) but rather in the quality of those images that people actually view (or try to view), since this can affect advertising revenue. Again,

one should note that post-scanning quality control is currently the responsibility of the partner libraries.

## Conclusion

This study has shown that—with some caveats—the pre-1872 digitized content now available in Google Books approximates that content available via the online catalog of a generic major American research library, and indeed is probably superior for post-1800 imprints. Google Books has reached this point in a remarkably short time—less than five years after the announcement of its initial Library Project—and given the large number of research library partners that have since been recruited, it seems likely that Google Books will eventually (perhaps very soon) become the single largest source for this content.

On the negative side of the ledger, two significant caveats must be recalled. The digitized images of individual pages are not always reliable—poor scanning can occasionally be so extensive as to render a digitized volume unusable—and folded maps and other illustrative matter are routinely scanned in their folded state, rendering them useless for research. One can reasonably expect that these flaws will be corrected over time, at least for high-demand texts: The users of the texts will insist on it and, at any rate, the libraries involved are committed to it.<sup>44</sup> Measuring the extent of this problem was not within the scope of the current study, but an extremely useful future research project would try to do so.<sup>45</sup>

On the positive side, Google Books provides full-text indexing, something of incalculable value that would have been inconceivable had these volumes not been scanned. This indexing allows one to search both within individual volumes and across the entire collection, facilitating text-based research in general, but especially historical research and the comparison of variant texts. While this indexing is dependent in individual cases on the quality of the original page scan and the fidelity of the OCR rendering, in the aggregate the amount of hidden content that is thus exposed far exceeds the amount that remains hidden (or imperfectly rendered via OCR). Additionally, Google Books is serving as a huge laboratory for what is called “document image understanding”—the increasingly sophisticated probabilistic analysis of page images to facilitate indexing, interpretation, and other activities.<sup>46</sup>

As noted above, in the past, large collections of works in the public domain—especially older English language works—were microfilmed by commercial firms in collaboration with various research libraries. The resulting microform collections have subsequently been digitized, either by the firms that did the original microfilming or by successor firms, and made available on a subscription basis. As Google

Books and other mass digitization projects continue their progress through various research library collections, the viability of these preexisting collections may increasingly come into question as subscribing institutions weigh their annual use of these materials against the annual charges they pay for access.

Currently only a small fraction of the materials in Google Books—perhaps 15 percent—is thought to be in the public domain.<sup>47</sup> The great bulk is still protected by copyright, including a large but unknown number of so-called orphan works for which it is difficult or impossible to locate the current copyright holder.

These materials, many of which have been digitized in the course of the Library Project, were the object of class-action lawsuits brought against Google in 2005 by the Association of American Publishers and the Authors Guild.<sup>48</sup> The parties proposed a settlement of these lawsuits on October 28, 2008, but at the time of this writing the fairness of its terms is still to be determined by the U.S. District Court involved. On July 2, 2009, the U.S. Justice Department informed the court that it had opened an anti-trust investigation into the settlement.<sup>49</sup>

Among librarians and researchers, the reaction to the proposed settlement was in some ways emblematic of the ambivalence felt by many who stand to benefit from Google's mass digitization. Harvard University and others have objected to the proposed settlement on the grounds that it would grant a de facto monopoly to Google. Robert Darnton, director of the Harvard University Library, summarized his misgivings in the *New York Review of Books*:

Google is not a guild, and it did not set out to create a monopoly. On the contrary, it has pursued a laudable goal: promoting access to information. But the class action character of the settlement makes Google invulnerable to competition. Most book authors and publishers who own US copyrights are automatically covered by the settlement. They can opt out of it; but whatever they do, no new digitizing enterprise can get off the ground without winning their assent one by one, a practical impossibility, or without becoming mired down in another class action suit. If approved by the court—a process that could take as much as two years—the settlement will give Google control over the digitizing of virtually all books covered by copyright in the United States.<sup>50</sup>

Given the research benefit that would accrue from providing direct integrated access to copyrighted material via Google, some mutually acceptable arrangement is likely to be reached, though its ultimate shape is hard to fathom at this point. The ramifications of any settlement are such that

a lengthy court review seems likely.

Beyond the terms of the proposed settlement lies the larger question of how Google Books will ultimately affect the world of learning. By making so much of the printed record available in digital form—and so rapidly—Google Books is both transforming how scholars use the printed texts of the past and feeding a larger fundamental reshaping of the world of scholarly research. Fortunately for the author, speculation on the ramifications of these changes lies beyond the scope of the current study. But Google Books clearly is already having a dramatic effect, both on libraries and on scholarship. Indeed, a significant number of the sources cited in this study—beyond the objects of the study itself—were consulted in Google Books rather than in a physical library. As more and more scholarly research is conducted online first—and especially if the universe of digitized copyrighted works is ultimately opened up in Google Books—libraries may find that researchers are not linking out of online catalogs to versions of works available on Google Books but are rather linking to library catalogs for those cases where a version available on Google Books is not satisfactory for their purposes. We will then be entering a brave new world for both research and libraries.

#### References and Notes

1. Google, "Google Checks Out Library Books," press release, Dec. 14, 2004, [www.google.com/press/pressrel/print\\_library.html](http://www.google.com/press/pressrel/print_library.html) (accessed Sept. 18, 2009).
2. For an example, see the agreement between the University of Michigan and Google, "Cooperative Agreement" (n.d.), [www.lib.umich.edu/mdp/um-google-cooperative-agreement.pdf](http://www.lib.umich.edu/mdp/um-google-cooperative-agreement.pdf) (accessed Sept. 18, 2009).
3. Karen Coyle, "Google Books as Library," online posting, Nov. 22, 2008, Coyle's InFormation, [http://kcoyle.blogspot.com/2008\\_11\\_01\\_archive.html](http://kcoyle.blogspot.com/2008_11_01_archive.html) (accessed Sept. 18, 2009).
4. See, for example, Committee on Institutional Cooperation and Google, "Cooperative Agreement" (n.d.), [www.cic.net/Home/Projects/Library/BookSearch/CIC-Google.aspx](http://www.cic.net/Home/Projects/Library/BookSearch/CIC-Google.aspx) (accessed Sept. 18, 2009), which commits to digitizing "not less than 10,000,000 volumes" (2); University of California and Google, "Cooperative Agreement" (n.d.), [www.cdlib.org/news/ucgoogle\\_cooperative\\_agreement.pdf](http://www.cdlib.org/news/ucgoogle_cooperative_agreement.pdf) (accessed Sept. 18, 2009), which commits to digitizing "no less than two and a half million (2,500,000) volumes" (2); University of Texas at Austin and Google, "Cooperative Agreement" (n.d.), [www.lib.utexas.edu/sites/default/files/google/utexas\\_google\\_agreement.pdf](http://www.lib.utexas.edu/sites/default/files/google/utexas_google_agreement.pdf) (accessed Sept. 18, 2009); and University of Virginia and Google, "Cooperative Agreement" (n.d.), [www2.lib.virginia.edu/press/uvagoogle/pdf/Google\\_UVA.pdf](http://www2.lib.virginia.edu/press/uvagoogle/pdf/Google_UVA.pdf) (accessed Sept. 18, 2009). The agreements typically include previously digitized content, according to Europeana, About Us: Background, [www.europeana.eu/portal/aboutus.html#background](http://www.europeana.eu/portal/aboutus.html#background) (accessed Sept. 18, 2009).
5. Miguel Helft, "Microsoft Will Shut Down Book Search Program," *New York Times*, Technology section, May 24,

- 2008, [www.nytimes.com/2008/05/24/technology/24soft.html](http://www.nytimes.com/2008/05/24/technology/24soft.html) (accessed Sept. 18, 2009).
6. Google, "Google Checks Out Library Books," press release, Dec. 14, 2004, [www.google.com/press/pressrel/print\\_library.html](http://www.google.com/press/pressrel/print_library.html) (accessed Sept. 18, 2009); Google, Corporate Information: Company Overview, [www.google.com/intl/en/corporate](http://www.google.com/intl/en/corporate) (accessed Sept. 18, 2009).
  7. Michael A. Keller, University Librarian, Stanford University, quoted on Google Book Search Library Partners, <http://books.google.com/googlebooks/partners.html> (accessed Sept. 18, 2009).
  8. The University of Michigan's workflow is illustrated in a flow chart, "Michigan Digitization Project Workflow," [www.lib.umich.edu/files/services/mdp/MDPflowchart\\_v3.pdf](http://www.lib.umich.edu/files/services/mdp/MDPflowchart_v3.pdf) (accessed Sept. 18, 2009).
  9. International Federation of Library Associations and Institutions, Section on Acquisition and Collection Development, "Guidelines for a Collection Development Policy Using the Conspectus Model," (2001), <http://archive.ifla.org/VII/s14/nd1/gcdp-e.pdf> (accessed Sept. 18, 2009).
  10. The items are IC-3, "The printed library materials I need for my work," IC-4, "The electronic information resources I need," and IC-8, "The print and/or electronic journal collections I require for my work." On the 2008 survey, the gap on IC-3 slipped back into the positive territory it had occupied in 2002, but the importance of this is difficult to determine. Standard deviations on these items are large—typically between 1.9 and 2.2—making detecting and measuring trends difficult. The ARL LibQual+ notebooks for 2002 to the present can be examined at [www.libqual.org/Publications/all.cfm?PubType=11](http://www.libqual.org/Publications/all.cfm?PubType=11) (accessed Sept. 18, 2009).
  11. Walt Crawford, "Libraries and Google/Google Book Search: No Competition!" online posting, June 21, 2006, Google Librarian Central, [www.google.com/librariancenter/articles/0606\\_03.html](http://www.google.com/librariancenter/articles/0606_03.html) (accessed Sept. 18, 2009). Thomas E. Nisonger's *Collection Evaluation in Academic Libraries* (Englewood, Colo.: Libraries Unlimited, 1992) contains annotated bibliographies on overlap in general and in the context of the Research Libraries Group Conspectus.
  12. Brian Lavoie, Lynn Silipigni Connaway, and Lorcan Dempsey, "Anatomy of Aggregate Collections: The Example of Google Print for Libraries," *D-Lib Magazine* 11, no. 9 (Sept. 2005), doi:10.1045/september2005-lavoie, [www.dlib.org/dlib/september05/lavoie/09lavoie.html](http://www.dlib.org/dlib/september05/lavoie/09lavoie.html) (accessed Sept. 18, 2009).
  13. Brian F. Lavoie and Roger C. Schonfeld, "Books without Boundaries: A Brief Tour of the System-wide Print Book Collection," *Ubiquity* 7, no. 37 (Sept./Oct. 2006), doi:10.1145/1167867.1167868, [www.acm.org/ubiquity/views/v7i37\\_books.html](http://www.acm.org/ubiquity/views/v7i37_books.html) (accessed Sept. 18, 2009).
  14. University of Michigan and Google, "Cooperative Agreement."
  15. Coyle, "Google Books as Libraries."
  16. Lavoie, Connaway, and Dempsey, "Anatomy of Aggregate Collections."
  17. British Library, Help for Library Researchers, English Short Title Catalogue—Introduction, [www.bl.uk/reshelp/findhelprestype/catblhold/estcintro/estcintro.html](http://www.bl.uk/reshelp/findhelprestype/catblhold/estcintro/estcintro.html) (accessed Sept. 18, 2009).
  18. Eltjo Buringh and Jan Luiten van Zanden, "Charting the 'Rise of the West': Manuscripts and Printed Books in Europe, a Long-Term Perspective from the Sixth through Eighteenth Centuries," *The Journal of Economic History* 69, no. 2 (June 2009): 409–45, doi:10.1017/S0022050709000837, <http://dx.doi.org/10.1017/S0022050709000837> (accessed Sept. 18, 2009).
  19. Lavoie and Schonfeld, "Books without Boundaries."
  20. Robert B. Townsend, "Google Books: What's Not to Like?" online posting, April 30, 2007, AHA Today, <http://blog.historians.org/articles/204/google-books-whats-not-to-like> (accessed Sept. 18, 2009); Ronald G. Musto, "Google Books Mutilates the Printed Past," *The Chronicle of Higher Education* 55 (June 12, 2009): B4, <http://chronicle.com/article/Google-Books-Mutilates-the/44463> (accessed Sept. 18, 2009).
  21. Karen Coyle, "Mass Digitization of Books," *Journal of Academic Librarianship* 32, no. 6 (Nov. 2006): 641–45, <http://dx.doi.org/10.1016/j.acalib.2006.08.002> (accessed Feb. 9, 2010).
  22. *Public Libraries in the United States of America* (Washington: GPO, 1876): 1:760, [www.archive.org/details/publiclibrariesi00unitrich](http://www.archive.org/details/publiclibrariesi00unitrich) (accessed Sept. 18, 2009). The term *public libraries* is used very broadly in the nineteenth century, including both academic and private research libraries. Curiously, part 1 of this work is not present in Google Books aside from the chapter on theological libraries digitized from the Andover-Harvard Theological Library. A copy identifying itself as part 1 of a University of Michigan copy in Google Books is, in fact, a digitization of part 2.
  23. Charles C. Jewett, *Notices of Public Libraries in the United States of America* (Washington, D.C.: Smithsonian Institution, 1851): 192, <http://books.google.com/books?id=paMnAAAAMAAJ> (accessed Sept. 18, 2009).
  24. Jewett, *Notices of Public Libraries*, 762–73.
  25. *Catalogue of the Library of the Boston Athenæum, 1807–1871* (Boston: The Athenæum, 1874–1882).
  26. Prospectus quoted in Josiah Quincy, *The History of the Boston Athenæum* (Cambridge, Mass.: Metcalf and Co., 1851): 12, <http://books.google.com/books?id=RHHRoEiGRdgC> (accessed Sept. 18, 2009).
  27. Jewett, *Notices of Public Libraries*, 22.
  28. Edward Edwards, *Memoirs of Libraries* (London: Triebner & Co., 1859): 2:194, <http://books.google.com/books?id=TH4NAAAQAAJ> (accessed Sept. 18, 2009).
  29. James Lyman Whitney, "Considerations as to a Printed Catalog in Book Form for the Boston Public Library," *The Library Journal* 24 (July 1899): 10, <http://books.google.com/books?id=ggcCAAAYAAJ> (accessed Sept. 18, 2009).
  30. *Catalogue of the Library of the Boston Athenæum, 1807–1871*.
  31. Anthony Panizzi, "Rules for the Compilation of the Catalogue," *Catalogue of Printed Books in the British Museum*, (London, 1841): 1:[v]–ix; Charles Ammi Cutter, "Explanations," *Catalogue of the Library of the Boston Athenæum, 1807–1871* (Boston: The Athenæum, 1874–1882): 1:page following title page, <http://books.google.com/books?id=Zbjcm5OGIsIC> (accessed Sept. 18, 2009).
  32. IFLA Study Group on the Functional Requirements for

- Bibliographic Records, *Functional Requirements for Bibliographic Records: Final Report*, UBCIM Publications-New Series, vol. 19 (Munich: K.G. Saur, 1998): 17–25, [www.ifla.org/files/cataloguing/frbr/frbr.pdf](http://www.ifla.org/files/cataloguing/frbr/frbr.pdf) (accessed Mar. 2, 2010).
33. Ibid., 74.
  34. Association of Research Libraries, ARL Library Data Tables 2007–08, worksheet “coll1,” [www.arl.org/bm~doc/08tables.xls](http://www.arl.org/bm~doc/08tables.xls) (accessed Sept. 18, 2009).
  35. IFLA Study Group, *Functional Requirements for Bibliographic Records*, 74.
  36. Charles Evans, *American Bibliography: A Chronological Dictionary of All Books, Pamphlets, and Periodical Publications Printed in the United States Of America from the Genesis of Printing in 1639 Down to and Including the Year 1820* (New York: P. Smith, 1941–67); Early American Imprints, Series I: Evans, 1639–1800, [www.readex.com/readex/product.cfm?product=247](http://www.readex.com/readex/product.cfm?product=247) (accessed Sept. 18, 2009).
  37. A Z-test is a statistical test used to calculate the probability that a given sample result lies within the range of results that one would expect if the only effect present was the operation of chance. A results outside this range suggests that something other than the operation of chance is contributing to that result.
  38. UNESCO, Revised Recommendation Concerning the International Standardization of Statistics on the Production and Distribution of Books, Newspapers and Periodicals (Nov. 1, 1985), [http://portal.unesco.org/en/ev.php-URL\\_ID=13146&URL\\_DO=DO\\_TOPIC&URL\\_SECTION=201.html](http://portal.unesco.org/en/ev.php-URL_ID=13146&URL_DO=DO_TOPIC&URL_SECTION=201.html) (accessed Sept. 18, 2009).
  39. Thomas Mann, *The Oxford Guide to Library Research* (New York: Oxford Univ. Pr., 1998): 185–87.
  40. For example, the original agreement with the University of Michigan explicitly excluded special collections materials. “Cooperative Agreement” (n.d.) 1.1, [www.lib.umich.edu/mdp/um-google-cooperative-agreement.pdf](http://www.lib.umich.edu/mdp/um-google-cooperative-agreement.pdf) (accessed Sept. 18, 2009).
  41. Brian Lavoie, “Anatomy of Aggregate Collections: Exploring Mass Digitization and the ‘Collective Collection,’” PowerPoint presentation, NELINET, Sept. 21, 2006, [www.oclc.org/research/presentations/lavoie/nelinet2006.ppt](http://www.oclc.org/research/presentations/lavoie/nelinet2006.ppt) (accessed Sept. 18, 2009).
  42. Joel Garreau, “Tongue in Check: With Translation Technology on Their Side, Humans Can Finally Lick the Language Barrier,” *Washington Post*, Style section, May 24, 2009, [www.washingtonpost.com/wp-dyn/content/article/2009/05/21/AR2009052104697.html](http://www.washingtonpost.com/wp-dyn/content/article/2009/05/21/AR2009052104697.html) (accessed Sept. 18, 2009).
  43. David Nicholas et al., “UK Scholarly E-Book Usage: A Landmark Survey,” *Aslib Proceedings* 60, no. 4 (2008): 311–34.
  44. See, for example, the HathiTrust instructions at <http://babel.hathitrust.org/cgi/mb?a=page;page=help> (accessed Sept. 18, 2009).
  45. Reports of this problem to date have been anecdotal. See, for example, Robert B. Townsend, “Google Books: Is It Good for History?” *Perspectives Online* 45, no. 6 (Sept. 2007), [www.historians.org/perspectives/issues/2007/0709/0709vie1.cfm](http://www.historians.org/perspectives/issues/2007/0709/0709vie1.cfm) (accessed Sept. 18, 2009); Musto, “Google Books Mutilates the Printed Past.”
  46. See, for example, Faisal Shafait et al., “Background Variability Modeling for Statistical Layout Analysis,” *Proceedings of the 19th International Conference on Pattern Recognition, December 8–11, 2008, Tampa, Florida, USA*, (2008) IEEE Computer Society, 2008, doi:10.1109/ICPR.2008.4760964, <http://dx.doi.org/10.1109/ICPR.2008.4760964> (accessed Sept. 18, 2009).
  47. Geoffrey Nunberg, “Google’s Book Search: A Disaster for Scholars,” *Chronicle of Higher Education* 56, no. 1 (Aug. 31, 2009): B4, <http://chronicle.com/article/Googles-Book-Search-A/48245> (accessed Sept. 18, 2009).
  48. Jonathan Band, “A Guide for the Perplexed: Libraries and the Google Library Project Settlement” (Nov. 13, 2008), <http://wo.ala.org/gbs/wp-content/uploads/2008/12/a-guide-for-the-perplexed.pdf> (accessed Sept. 18, 2009); Jonathan Band, “A Guide for the Perplexed Part II: The Amended Google–Michigan Agreement” (June 12, 2009), <http://wo.ala.org/gbs/wp-content/uploads/2009/06/google-michigan-amended.pdf> (accessed Sept. 18, 2009).
  49. William F. Cavanaugh, Deputy Assistant Attorney General, letter to Hon. Denny Chin, July 2, 2009, [http://graphics8.nytimes.com/packages/pdf/technology/20090702\\_GOOGLE\\_DOJLetter.pdf](http://graphics8.nytimes.com/packages/pdf/technology/20090702_GOOGLE_DOJLetter.pdf) (accessed Sept. 18, 2009).
  50. Robert Darnton, “Google and the Future of Books,” *The New York Review of Books* 56, no. 2 (Feb. 12, 2009), [www.nybooks.com/articles/22281](http://www.nybooks.com/articles/22281) (accessed Sept. 18, 2009). A subsequent response to Darnton is Paul N. Courant et al., “Google and Books: An Exchange,” *The New York Review of Books* 56, no. 5 (Mar. 26, 2009), [www.nybooks.com/articles/22496](http://www.nybooks.com/articles/22496) (accessed Sept. 18, 2009).

## Cataloging and Classification

### Review of the Literature 2007–8

By Sydney Chambers and Carolynne Myall

*This paper surveys library literature on cataloging and classification published in 2007–8, indicating its extent and range in terms of types of literature, major subject areas, and themes. The paper reviews pertinent literature in the following areas: the future of bibliographic control, general cataloging standards and texts, Functional Requirements for Bibliographic Records (FRBR), cataloging varied resources, metadata and cataloging in the Web world, classification and subject access, questions of diversity and diverse perspectives, additional reports of practice and research, catalogers' education and careers, keeping current through columns and blogs, and cataloging history.*

Speculations, questions, anxieties, and excitement about the roles and possible futures of cataloging and catalogers underlie much of the literature of cataloging and classification during 2007–8. While many publications focused on the future and the significant changes that emerging trends may require, other contributions addressed a variety of aspects of practice—many immediate and practical—and underlying philosophy. Topics in the history of bibliographic control and in representing diverse and global perspectives in cataloging data also were strongly in evidence. The objectives of this paper are to

- survey the extensive and varied literature of cataloging and classification during 2007–8;
- indicate the range of this literature in terms of types of publications, including scholarly works but also publications intended to aid practitioners and communicate cataloging issues to noncatalogers;
- identify major subject areas and themes; and
- recommend substantive contributions in these areas, along with more ephemeral but worthy contributions useful to catalogers struggling to keep cataloging alive and useful in a period of scrutiny, uncertainty, multiple initiatives, and change.

#### Research Method

Using a bibliographic management program, we began the project by creating a database of citations with folders for 2007 and 2008. To do this, we searched several online databases, including Library Literature and Information Science Full Text; Library, Information Science, and Technology Abstracts with Full Text; Online Computer Library Center (OCLC) WorldCat; and Dissertations and Theses Online. Search strategies included both keyword and subject heading searches, using many pertinent terms, such as bibliographic control, cataloging,

**Sydney Chambers** (chambers@gonzaga.edu) is Assistant Professor, Chair of Materials Management, Foley Center Library, Gonzaga University, Spokane, Washington, and **Carolynne Myall** (cmyall@mail.ewu.edu) is Head of Collection Services, University Libraries of Eastern Washington University, Cheney.

Submitted November 13, 2009; accepted for publication December 14, 2009.

The authors thank Julie Miller, PhD, Associate Dean of Libraries, Eastern Washington University, for her help in improving the manuscript.

classification, RDA: Resource Description and Access, and *Functional Requirements for Bibliographic Records (FRBR)*. We also searched Google Books and Google itself, although most of our Web searching followed the strategy of following links that led from one document or blog to another. We favored more stable or persistent formats of resources, since some Web drafts and documents appeared and then were removed or relocated, a problematic characteristic for long-term identification.

In an initial review, citations for news articles and reviews for each year were moved to separate folders. We used citations for reviews to identify monographic publications; citations for news articles were reviewed to identify major concerns and events. Citations for works evidently out-of-scope or in languages we could not read were moved to “holding pen” folders. Then we began a more thorough review of the remaining citations. Ultimately, we limited the scope to English-language resources, particularly those that applied to North American libraries. The approximate numbers of unique citations remaining were 444 for 2007 and 350 for 2008.<sup>1</sup> Next, we worked through the citations and abstracts, obtained print or online copies, assembled a file of printed resources, and attempted to cull further. With the shifting boundaries of cataloging, determining whether an article was in-scope was not always easy, even after an initial review. Acknowledging the difficulties posed by the indeterminate boundaries of the field at the moment, we agreed that for a contribution to be defined as in-scope, it must have a library cataloging or classification application or orientation, or it must represent an application of library cataloging methods to a problem in the broader universe of information (e.g., Yahoo!). User studies were similarly limited by this definition of scope. Since we were interested in the distribution of cataloging literature across many publications, we next arranged journal articles by publication title and cross-tabulated for subject area and types of literature using lists and colored tabs (a primitive method, but effective for the visually oriented).

Our file grew to include 468 articles from 104 periodicals, plus many new or revised cataloging standards, such as *Descriptive Cataloging of Rare Materials (Books) (DCRM (B))*, as well as monographs, reports, theses, and Web documents.<sup>2</sup> While we emphasized scholarly articles, reports, and monographs intended for specialists in cataloging and classification, we also considered literature directed toward wider audiences in librarianship. We were struck by the large number of possibly pertinent resources—surely a testimony to the high level of activity and interest in bibliographic organization—and by their appearance in many different publications.

Evaluation and selection was a time-consuming undertaking. We attempted to include some contributions of interest to all parts of the cataloging community. While

each of us concentrated on specific areas, we reviewed each other's selections and suggested works for inclusion or removal. In each area of the literature discussed in this review, inclusions are representative, not comprehensive; as any cataloger could predict, many items legitimately might have been placed in more than one category. Despite our best efforts to include all pertinent materials in our search, some works no doubt eluded us, and many valuable contributions could not be included because of space constraints, to our great regret. Finally, any selection from a large number of works will reflect the interests and tastes of the selectors. In particular, our current experience as practitioners in mid-size academic libraries rather than in research libraries or as information studies faculty, and our past experience in other types of libraries, may have influenced our perceptions and choices.

The most immediately prominent contribution to cataloging literature was, we believe, the report of the Library of Congress Working Group on the Future of Bibliographic Control (LCWG); it raised themes of change and adaptation to new conditions that appeared throughout the literature.<sup>3</sup> Also prominent in the literature were themes of cataloging tradition as well as its objectives, methods, and values, and their continuing pertinence in the emerging environment. These two strands of thought—and the dialogue between them—provide the structure for this paper, which begins with the LCWG report and the future of cataloging and closes with contributions on cataloging and classification history.

### The Future(s) of Bibliographic Control

In April 2006, the Library of Congress (LC) announced that it would no longer create or maintain authority records for series. This unanticipated announcement provoked an apparently unanticipated uproar in the library community. The LC, while postponing but not rescinding its series decision, responded to the concerns of the community by appointing the LCWG, intended to be broadly representative, to consider the long-term future of bibliographic control in the twenty-first century.

The LCWG held its first meeting in November 2006. Over the next year, the LCWG solicited opinions on its website, held regional sessions available to others via webcast, participated in a meeting hosted by the American Library Association (ALA), issued a draft report, and gathered 135 pages of written comments, among other activities, to promote discussion at national and local levels. On January 9, 2008, the LCWG delivered to the LC what was probably the most widely read and debated contribution to cataloging literature during this two-year period: the LCWG's final report, *On the Record*.<sup>4</sup>

Presenting more than one hundred specific recommendations, *On the Record* was based on three guiding principles or redefinitions: first, that bibliographic control was broader than library cataloging; second, that the bibliographic universe now extended beyond libraries and publishers to include creators, vendors, distributors, users, and other groups across various kinds of boundaries; and third, that the LC, while still playing a unique role in the U.S. library community, could not continue to perform all of its traditional functions under its current conditions and legislative mandate. Following these principles, LCWG envisioned the future of bibliographic control as

collaborative, decentralized, international in scope, and Web-based. Its realization will occur in cooperation with the private sector, and with the active collaboration of library users. Data will be gathered from multiple sources; change will happen quickly; and bibliographic control will be dynamic, not static.<sup>5</sup>

The LCWG organized its specific recommendations under five general proposals: increase the efficiency of bibliographic production and maintenance; enhance access to rare, unique, and other special hidden materials; position our technology for the future; position our community for the future; and strengthen the library and information science profession. At the macro level, some of the general recommendations seemed mostly unexceptionable. As the literature attested, many libraries were already engaged in creating a more collaborative cataloging future by increasing their contributions to cooperative efforts.<sup>6</sup> And what cataloger could disagree in principle, if not in specifics, with recommendation 4.3, “Optimize LCSH for Use and Reuse”?<sup>7</sup> For that matter, who could deny that libraries were no longer the primary element of the information universe, or that library catalogs were not the starting point for most information-seekers in the digital environment? Taken collectively, though, the LCWG’s recommendations, in Hill’s words, might require “us to take up residence in an alternate universe, with new understandings, new perspectives, and new responsibilities,” a disconcerting prospect.<sup>8</sup>

The specific recommendations of the report touched on many unresolved issues in cataloging and sparked controversies in a number of areas and from a number of points of view. For example, the report recommended that libraries invest more local effort in “exposing to more users rare and unique materials held by libraries that are currently hidden from view,” a presumed benefit for both local and global users, while it seemed to advocate less response to local needs through making adjustments to bibliographic records, an arguable point for a library with a strong local mission.<sup>9</sup> Recommendation 3.2.5 proposed a suspension of work on

RDA: Resource Description and Access, the new cataloging code to which much effort had already been devoted, an unpopular suggestion even among those who viewed RDA as imperfect.<sup>10</sup> Hillmann offered a stimulating discussion of this and other recommendations from the draft version of the LCWG report.<sup>11</sup> The number and variety of responses to the LCWG’s recommendations testified to its identification of key issues, if not to agreement about how to address them. As a snapshot of where leaders in the library community (as represented by the members of the LCWG) thought we were and where we thought we were going, *On the Record* seemed likely to remain a key document in cataloging and U.S. library history of the early twenty-first century.

Sanchez’s *Emerging Issues in Academic Library Cataloging & Technical Services* is a different kind of snapshot of the cataloging environment from the LCWG report, but obviously a picture of the same reality. *Emerging Issues* is a compilation of the results of a survey of nine cataloging and technical services departments in academic libraries of varying nature and size concerning their organization, staffing, the use of new technologies, the transition to metadata standards, and other activities.<sup>12</sup> Many of the changes and directions reported here anticipated the recommendations of the LCWG report. For example, departments were engaged in a growing number of digital projects within a more diverse metadata environment, and they reported more extensive collaboration with library and nonlibrary colleagues, in their own institutions and beyond.

Another LC publication, issued in 2007, addressed the future of the Cataloging in Publication (CIP) program. *CIP Poised for Change*, a report prepared by the CIP Review Group (CRG), presented findings from the 2006 CIP surveys of libraries, publishers, and MARC Distribution Service customers (the MARC Distribution Service supports the CIP program as an important means to increase marketing and sales for publishers, as well as to supply standardized cataloging to libraries quickly).<sup>13</sup> The CRG recommended enhancing the CIP service with enriched records, more cataloging partners, greater timeliness, the inclusion of a wider range of formats, and a user interface for libraries and publishers. One CRG recommendation, “Examine the feasibility of developing an optional ECIP [Electronic CIP] front end that would accommodate a batch ONIX feed of forthcoming books,” was close to the LCWG’s recommendation 1.1.3.1., “Develop content and format guidelines for submission of ONIX data to the CIP program and require publishers . . . to comply.”<sup>14</sup> Both reports, in different ways, supported the importance of availability of bibliographic data early in the supply chain.

The underlying assumptions of the LCWG report and the directions it proposed were not universally accepted in the cataloging literature. In a slightly revised version of a paper read before the LCWG’s second meeting, Bade, a

defender of traditional library cataloging quality standards, took issue with many of the LCWG's assumptions—for example, defining users of libraries as consumers and managers and consequently viewing information management as a technical and industrial process.<sup>15</sup> Bade proffered that we had choices with regard to these assumptions:

We can choose to understand what happens in libraries according to a theory of transportation in which all data is equal and we simply move it from one place to another, or we can understand work in libraries according to a theory of communication in which we engage readers in conversations about writers and writings, musicians and music, etc. The importance of that choice lies in the fact that we can do either—with radically different consequences.<sup>16</sup>

In “Will the Response of the Library Profession to the Internet Be Self-Immolation?” Yee advised all librarians, not just catalogers, to follow the deliberations of LCWG: “These library leaders have forgotten, or never knew, the fact that expertise in organization of information is at the core of the profession of librarianship.”<sup>17</sup> She saw threats to the profession of librarianship not in the Internet, but rather in librarians who did not understand the nature of librarianship or appreciate the value of human intervention in information organization, in library system designers who did not understand the nature of catalog records and searching, and in the failure to recognize the larger implications of relinquishing to commercial interests the intermediary role of libraries in the information provision chain. “‘On the Record’ But Off the Track,” by Mann, a defender of the roles of libraries and of the LC specifically in supporting research and scholarship, detailed many objections to the LCWG report.<sup>18</sup> Mann, a strong proponent of the value of arrays of precoordinated Library of Congress Subject Headings (LCSH) strings in scholarly research, discussed the LCWG report's failure to distinguish between the needs of scholarly and other researchers; inadequate recognition of the value of controlled vocabulary and LCSH arrays, with subdivisions and cross-references, as opposed to keyword retrieval; minimizing of the LC's contributions as a de facto national library; and assignment of higher priority to digitization of textual special collections rather than maintenance of LCSH and Library of Congress Classification, among other objections.

Some contributions to the literature assumed that significant change in cataloging practice must occur in response to changing library priorities and user expectations. In “Utility, Library Priorities, and Cataloging Policies,” Banush and LeBlanc defended the application of the philosophical concept of utility—“the pursuit of the greatest happiness for all,

the definition of happiness as the pleasure and absence of pain of sentient beings, and the principle of impartiality”—to the consideration of cataloging policies within the context of overarching library priorities.<sup>19</sup> Banush and LeBlanc quoted a question posed by Marcum: “Just how much do we need to continue to spend on carefully constructed catalogs . . . in the Age of Google?”<sup>20</sup> They observed that “doing the most good for the greatest number of items does not necessarily mean doing the most good for an individual title” and proposed that a “pragmatic, triage-based process of evaluating needs and allocating staff effort allows for a utility-based approach to solving a thorny need-satisfaction problem,” then examined the results of “cataloging pragmatism” at their institution.<sup>21</sup> The authors concluded that in terms of full access to some library resources, their institution's “triage-based strategy may be bibliographically risky, [but] continuing past practices indefinitely also entails its own potential perils.”<sup>22</sup>

“Subject Headings in Full-Text Environments: The ECCO Experiment,” by Garrett, quoted the same question by Marcum (albeit in a different manifestation) about the necessity for full catalog records in a universe of digital information and keyword searching, but arrived at a different answer.<sup>23</sup> This case study demonstrated that subject headings on catalog records for digitized historical materials enriched the keyword index. This enrichment enhanced access to the materials because terms and even proper names changed over time and current searches did not necessarily match the terms found in the full-text documents themselves. Garrett concluded, “Meta-vocabulary . . . performs an important hermeneutic and heuristic function in bibliographic search and discovery, across centuries and across languages—and . . . even across states of the same language over time.”<sup>24</sup> Bair and Carlson, in “Where Keywords Fail: Using Metadata to Facilitate Digital Humanities Scholarship,” also demonstrated the weakness of keyword searching of digitized primary source documents—in this case a set of eight U.S. Civil War diaries.<sup>25</sup> Misspellings, alternate spellings, regional variations in word usage, errors and lack of knowledge on the part of the diarists, abbreviations, and variations of names all contributed to retrieval difficulties. This project added authorized headings for personal, corporate, and place names and incorporated subject analysis at the word level in the XML markup. The authors noted that “merely scanning and providing full-text keyword searchability may not fully meet the needs of digital humanities scholars. . . . The addition of metadata in the form of normalized name headings and topics can greatly enhance the research experience and save the time of users, especially students.”<sup>26</sup>

The future of cataloging and bibliographic control was the explicit focus of many contributions. Simultaneously recognizing and utilizing change in the information environment, while preserving distinctive and valuable aspects of

the library cataloging tradition, was the underlying challenge addressed in the literature. Responses varied greatly. For example, “The Online Library Catalog: Paradise Lost and Paradise Regained?” by Markey, identified the failures of online catalogs, rejected the idea that cataloging simplification and primary sources would regain users, and proposed an alternate route to catalog improvement.<sup>27</sup> Markey suggested that in a post-mass-digitization future, probabilistic searching (rather than outdated Boolean-based retrieval), subject cataloging “to take advantage of the user’s ability to recognize what they want or do not want during the course of the search,” and metadata that enables precise qualification of searches could bring the catalog into prominent use again.<sup>28</sup> In the meantime, several specific steps would improve catalogs immediately: including data elements that users want to see in brief displays; ranking algorithms that weight summary data such as titles, subject headings, and class numbers; and “integration of online library catalog searching into the larger scenario of information seeking generally,” among other proposals.<sup>29</sup> From another perspective, “Who Moved My Pinakes? Cataloging and Change” by Gross, acknowledged the changes inevitably underway in cataloging and librarianship, but observed, “What has yet to be resolved is whether these changes will actually mean progress and improvement or the gutting of our mission.”<sup>30</sup> Gross objected to the characterization of reactions against recent developments merely as fear of change. In this context, objections might be legitimate defenses against threats to the integrity of cataloging.

### General Cataloging Standards and Texts

RDA, the new cataloging code intended to replace *Anglo-American Cataloguing Rules*, 2nd ed. (*AACR2*), was under development throughout the period; drafts and discussion papers appeared on the website of the Joint Steering Committee (JSC) for Development of RDA, culminating in a full draft in November 2008.<sup>31</sup> While not yet finalized, RDA and its background and concepts were discussed in the cataloging literature. For example, in “Designating Materials: From ‘Germane Terms’ to Element Types,” Weihs and Howarth traced the use of formalized terms following the title to designate physical format, to development of the general material designation (GMD), to RDA’s proposed media type, carrier type, and content type.<sup>32</sup> In the context of the JSC’s announcement that RDA would permit the use of family names as authors, Creider discussed the ways family names had been established in both the archival and cataloging traditions, and he also proposed changes to better address needs of users.<sup>33</sup> In “Cataloging Cultural Objects (CCO), Resource Description and Access (RDA), and the Future of Metadata Content,” Beacom identified

commonalities and differences between RDA and CCO, a metadata scheme largely used in the museum community, and presented RDA’s development as an opportunity for engagement between metadata communities, even when significant differences exist.<sup>34</sup>

“Resource Description and Access (RDA): Cataloging Rules for the 20th Century” proposed that the emerging code should focus more on moving library cataloging into the digital age.<sup>35</sup> Rather than emphasizing continuity with the Anglo-American Cataloguing Rules in any of its manifestations, RDA developers should make a more radical break with the past, examine other metadata models—such as Dublin Core and CCO—and incorporate concepts regarding bibliographic relationships. “If new cataloging rules are developed without the parallel development of new models for library catalogs,” authors Coyle and Hillmann warned, “then it will be necessary for some in the library world to set off in their own direction, rejecting what they see as insufficient change.”<sup>36</sup> Gorman, editor of *AACR2*, delivered an emphatic rebuke to RDA developers in “RDA: Imminent Debacle.” Gorman’s objections concerned the structure of the developing code, its relationship (or lack thereof) to ISBD, poor editing, and a perceived lack of compelling reasons for abandoning *AACR2*. He stated that “The RDA seeks to find a third way between standard cataloging . . . on the one hand and the metadata crowd and boogie-woogie Google boys on the other. . . . The sad thing is that betraying the former has not managed to appease the latter.”<sup>37</sup>

As noted earlier, probably the most startling comments about RDA in the cataloging literature appeared in *On the Record’s* recommendation 3.2.5: “Suspend Work on RDA.” The LCWG advised the JSC to

suspend further new development work on RDA until a) the use and business cases for moving to RDA have been satisfactorily articulated, b) the presumed benefits of RDA have been convincingly demonstrated, and c) more, large-scale, comprehensive testing of FRBR as it relates to proposed provisions of RDA has been carried out against real cataloging data.<sup>38</sup>

The LC decided to proceed with the test of RDA, but the LCWG recommendation may have reflected unspoken feelings in the cataloging community. The literature of 2007–8 appeared to show more interest in and energy toward the topic of *FRBR*, one of the theoretical underpinnings of RDA, than around the new code itself.<sup>39</sup> (*FRBR*-related literature is discussed in the next section.)

The International Federation of Library Associations (IFLA) moved closer to a new Statement of International Cataloguing Principles to replace the statement known as the Paris Principles approved by the International Conference

on Cataloguing Principles in 1961.<sup>40</sup> The Statement of International Cataloguing Principles, while not a cataloging code itself, was intended to update the Paris Principles, build on world cataloging traditions, and provide guidance for the development of future codes. Tillett provided a helpful overview of the April 2006 version of the draft statement at a 2006 IFLA Meeting of Experts, published in the meeting proceedings in 2007.<sup>41</sup> A final draft statement was available for worldwide review in 2008 (the final version was ultimately published in 2009).<sup>42</sup>

Also during this period, the IFLA issued *International Standard Bibliographic Description (ISBD)*, Preliminary Consolidated Edition. Intended to supersede the individual ISBDs, the consolidated edition incorporated many revisions made during recent years to ISBD (A), ISBD (CR), and others, while retaining the basic ISBD structure and data elements.<sup>43</sup>

The third edition of Chan and Hodge's *Cataloging and Classification: An Introduction* (a staple of many cataloging departments) retained the organization and coverage of previous editions, but it expanded several topic areas and added coverage of metadata schema and other recent developments.<sup>44</sup> Another worthy addition to the training shelf of libraries of many types (despite its title) was *Standard Cataloging for Public and School Libraries*, 4th ed., by Intner and Weihs.<sup>45</sup> While including material especially valuable for school libraries (e.g., use of *Sears List of Subject Headings*), the work also provided cogent discussion and examples related to many cataloging topics and standard tools. *Organizing Information from the Shelf to the Web*, by Chowdhury and Chowdhury, also provided context to students and practitioners via an overview of library and nonlibrary approaches to organizing information.<sup>46</sup> This work, with primarily British examples, covered cataloging rules and concepts (e.g., AACR2, the FRBR model), MARC formats, major library classification schemes (Colon Classification, LCC, Dewey Decimal Classification (DDC), Universal Decimal Classification), controlled vocabulary for subject description, and the structure of thesauri. It also discussed issues in organizing Internet resources, such as the application of vocabulary control tools.

### FRBR

FRBR, a conceptual model of the bibliographic universe, continued to be a major focus of cataloging literature and made appearances in many publications of which FRBR was not the stated topic.<sup>47</sup> Still, a decade after publication of the final report of the IFLA Study Group on Functional Requirements for Bibliographic Records, FRBR concepts remained difficult for many to grasp with confidence. *FRBR: A Guide for the Perplexed*, by Maxwell, came to their aid, demonstrating the

entity-relationship model of FRBR and explaining FRBR entities, relationships, and user tasks in clear, accessible language.<sup>48</sup> Maxwell also discussed difficulties in applying the entity-relationship model to existing databases of MARC records and the lack of progress, so far, in integrating FRBR concepts into catalogs and codes. A special section of the *Bulletin of the American Society for Information Science and Technology* presented FRBR concepts, their implications, and the challenges facing FRBR implementation in information systems. Riva, chair of the IFLA FRBR Review Group, provided a concise introduction for this section, with discussion of FRBR's influence on emerging cataloging codes and other ongoing developments involving authority records.<sup>49</sup> Another helpful contribution was Dickey's "FRBRization of a Library Catalog: Better Collocation of Records, Leading to Enhanced Search, Retrieval, and Display."<sup>50</sup> Dickey discussed the benefits of applying FRBR concepts to catalogs—showing advantages particularly for collections of resources in music and fine arts, theology, and literature—and described technical solutions for changing database structure to reflect bibliographic relationships.

*Understanding FRBR: What It Is and How It Will Affect Our Retrieval Tools*, edited by Taylor, contained essays on FRBR's principles and development, its relationship to cataloging history, and its possible applications in a number of library and other environments.<sup>51</sup> For example, Shadle's "FRBR and Serials: One Serialist's Analysis" presented a view of how serial publications might be modeled in FRBR, and Vellucci's "FRBR and Music" discussed the FRBR structure as it related to musical works and expressions.<sup>52</sup> "FRBR and Works of Art, Architecture, and Material Culture," by Baca and Clarke, demonstrated that for many cultural objects, "the conceptual model of the FRBR Group 1 entities (*work, expression, manifestation, item*) does not apply" because work is embodied in a single material object rather than existing as an abstract entity.<sup>53</sup> In "Bibliographic Families and Superworks," Smiraglia distilled much of his thinking about works, constellations of works, and instantiation networks in the context of FRBR concepts of work, expression, and manifestation.<sup>54</sup>

The IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR) was appointed in 1999 to analyze FRBR entities related to authority data (e.g., persons, families, works, and places), their attributes, the names by which they are known, and controlled access points for them. FRANAR's draft report, discussing relationships that may exist between names of entities and access points based on those names, was available for review and comment on IFLA's website during the period under discussion. (When this review was written in mid-2009, however, the draft had been replaced on IFLA's site with publication information for FRANAR's *Functional Requirements for Authority Data (FRSAD)*: A

*Conceptual Model.*<sup>55</sup> Meanwhile, the IFLA Working Group on Functional Requirements for Subject Authority Records (FRSAR) posted a draft in November 2008, addressing its charge

- to build a conceptual model of Group 3 entities within the FRBR framework as they relate to the *aboutness* of works,
- to provide a clearly defined, structured frame of reference for relating the data that are recorded in subject authority records to the needs of the users of those records, and
- to assist in an assessment of the potential for international sharing and use of subject authority data both within the library sector and beyond.<sup>56</sup>

“Functional Requirements for Bibliographic Records: An Investigation of Two Prototypes” considered how the *FRBR* conceptual model worked when applied to two prototype bibliographic databases, LibraryLabs (National Library of Australia) and FictionFinder (OCLC). Search results pointed out some difficulties in implementing *FRBR* and issues that need resolution.<sup>57</sup> “Perhaps the biggest drawback of FRBRised displays in the prototypes is that they give a sense of completeness, even though they don’t allow users to carry out all of their tasks,” observed authors Pisanski and Žumer.<sup>58</sup> Nevertheless, they concluded, “even imperfect displays of the structure of the bibliographic universe should be better than the displays usually associated with OPACs, especially once the user understands the problems related to the concept.”<sup>59</sup>

“Making the Pieces Fit: *Little Women*, Works, and the Pursuit of Quality” reported on research into the extent to which records for manifestations of an identified work set could be automatically identified, since problems in cataloging rules and practice have resulted in records that may make the automatic identification of work set members difficult.<sup>60</sup> The project selected four works of fiction (one of which—*Little Women*—presented a particularly tricky situation) and used a detailed, multiple-pass process utilizing information in authority and bibliographic records as well as LCC numbers to identify work set members. For the four works, 77 to 95 percent of the records were, ultimately, correctly identified. Authors Carlyle, Ranger, and Summerlin proposed further research into the potential of records’ LCC numbers as work identifiers, among other possibilities, and concluded that “by expanding the means by which author name and title attributes identifying a work are discovered, an automatic work identification process could work very well to improve catalog performance.”<sup>61</sup>

Petek, in “Bibliographic Families and Relationships among Family Members in COBIB,” used an entity-relationship model to examine the Slovenian national cooperative

bibliographic database, COBIB.<sup>62</sup> The project constructed bibliographic “families” by identifying the progenitor (the work or superwork level in *FRBR*) and all the related “family members” (expressions and manifestations), and then analyzed the size of bibliographic families, types of relationships presented, and how well those relationships were expressed in the bibliographic records. While the catalog managed to identify family members, the work identification was only done implicitly. Based on how poorly family relationships were expressed in bibliographic records, Petek proposed an authority database for works, to which all related bibliographic records would be linked: “Since there is a difference between the content and the carrier, i.e., the work and the item, and assuming that users are more interested in works, some changes should be undertaken to adapt the current catalogue to new user needs.”<sup>63</sup>

Université catholique de Louvain in Belgium has been experimenting with *FRBR* cataloging since 2003 and is working on a *FRBR* approach for cataloging maps. In “FRBR: An Opportunity for Map Collections and Map Users?” Kalf described this work and presented theoretical examples for *FRBR*’s use with maps, with this caveat: “One thing is clear: a serious theoretical approach must be developed prior to beginning any cataloging to organize information and build a coherent *FRBR* tree.”<sup>64</sup> Kalf noted that maps have characteristics—such as scale—that could be assigned either to the expression or manifestation level; she also enumerated other issues with map cataloging and *FRBR* and presented possible solutions.

“Linking Print and Electronic Books: One Approach” provided an example of how a library might use *FRBR* concepts to serve users more effectively.<sup>65</sup> Having purchased table-of-contents information to enrich records for many print books, the library wished to alert users to the availability of electronic manifestations for the same titles, though the records for the electronics versions lacked the TOC information. This was accomplished by using a nonstandard, local system-specific field to link the parallel records for the two manifestations; users who retrieved one record were provided a link to the record for the other format. Authors Simpson, Lundgren, and Barr “encouraged others to explore creative solutions that will overcome the absence of data . . . that might have facilitated navigation among different versions, but which were not added to records for cost reasons in the past.”<sup>66</sup>

## Cataloging Varied Resources

Cataloging literature reflected the range of media, physical formats, and content carriers of current library resources. While Yee’s *Moving Image Cataloging: How to Create and How to Use A Moving Image Catalog* covered several sets

of pertinent data content standards, it focused most on the principles underlying effective systems for organizing information: “Instead of telling you how to do it right, this textbook will try to teach you how to think about it right.”<sup>67</sup>

We recommend this bracing text to students, catalogers struggling with new media forms, practitioners with training responsibilities, and all those with an interest in continuing the role of human intervention in information organization. An operational contribution in this area of cataloging was Ho’s report on one academic library’s work to enhance bibliographic records for video recordings, for the benefit of OCLC and library users.<sup>68</sup> Ho’s article may, we hope, stimulate other libraries to engage in similar projects in the spirit of the LCWG’s recommendations.

A welcome update to the 2002 version, *Guide to Cataloging DVD and Blu-ray Discs Using AACR2r and MARC 21*, issued by Online Audiovisual Catalogers (OLAC), addressed the treatment of Blu-ray discs and other recent questions in DVD cataloging.<sup>69</sup> Each section, written by members of the OLAC DVD Cataloging Guide Update Task Force, provided guidance on using AACR2 and MARC 21 for DVDs of all kinds; included exemplars of packaging, labels, and on-screen credits; and showed how the information in those areas would be expressed in a MARC record. OLAC and the Music Library Association jointly produced the *Guide to Cataloging Playaway Devices Based on AACR2 Chapters 6 and 9* to aid catalogers struggling to describe the integration of content and playback device. It also included examples with photos of the device showing the locations of bibliographic information and the corresponding MARC records.<sup>70</sup>

*Descriptive Cataloging of Rare Materials (Books)*, the third version of a standard tool for treating rare printed monographs, added a chapter on objectives and principles and useful new appendixes on variations requiring the creation of new bibliographic records and collection- and core-level records, among other topics.<sup>71</sup> In a change from earlier editions, the LC played a supporting role in the development of the publication while primary responsibility was undertaken by a committee of the Association of College and Research Libraries’ Rare Books and Manuscripts Section. In “The Best of Both Worlds: Using CCO for Object Cataloging in Libraries and Special Collections,” O’Keefe encouraged the use of the data content standard Cataloging Cultural Objects (CCO) where appropriate, identified data elements unique in either AACR2 or CCO, and provided advice about mapping between the two schemes.<sup>72</sup>

Allgood’s ambitious “Serials and Multiple Versions, or The Inexorable Trend toward Work-Level Displays” examined the problems that multiple manifestations of serials continued to present to catalog users, and it analyzed current and possible solutions.<sup>73</sup> Among the most promising solutions were catalog code revision and the application of

FRBR principles to catalog displays. Allgood’s conclusion was applicable beyond the arcane territory of continuing resources cataloging:

It is time for librarians to determine if solutions to issues like the MulVer problem are complex because they have to be, or complex because librarians perpetuate practices that make them complex. . . . Within today’s world of proliferating information carriers, providing consistent access to the content users seek is inherently complex, but to users it must appear simple. The job of today’s librarians is to apply complex solutions to attain apparent simplicity—call it the Zen of librarianship.<sup>74</sup>

Latest versus successive entry cataloging of serial publications persisted as a topic of debate. In “Excessive Successive: Time for a Radical Change,” Baia called for revising cataloging codes to allow latest entry cataloging.<sup>75</sup> While acknowledging that introducing the concept of major and minor changes reduced the difficulties caused by successive entry practice, Baia noted that online journals—particularly those with all issues available in one place without regard to earlier titles under which issues may originally have been published—were not well served by successive entry cataloging; she proffered latest entry as a better fit with current technology. Greci, on the other hand, considered arguments against successive entry but concluded that current technology could enable us “to bring together holdings from an entire run of serial records into one user-friendly catalog display.” She advocated focusing efforts “on building and implementing online catalogs that make use of this potential.”<sup>76</sup> In “Latest Entry Legacies: Confessions of a Guerrilla Cataloger,” Randall recounted his experiences in a library that continued to use latest entry cataloging through the 1990s.<sup>77</sup> While Randall believed the practice was beneficial for library users, the library ultimately discontinued latest entry treatment when it became “a hindrance because of its variance from international standards” and caused problems in synchronizing records with other libraries (including those within the same university), MARC tagging, and use of standard identifiers.<sup>78</sup> Randall supported the development of sophisticated catalogs able to utilize linking data in the records: “If there is any solution to the problem of displaying the history of a serial in the catalog, I believe that it is in using encoded links between records to create a virtual record on the fly in response to a user’s search in the catalog.”<sup>79</sup> In “Confessions of a Correspondent from the Choice-of-Entry War,” Shroyer presented results of informal opinion surveys of catalogers and public service librarians about latest or successive entry preferences.<sup>80</sup> Both groups expressed a need and desire for catalogs to represent the history of a serial in a unified, composite display,

with public service librarians finding “the general state of access information for serials unsatisfactory, if not abysmal, for users.”<sup>81</sup> “The manner in which many of the public services librarian respondents to my survey characterized the current organization and presentation of serials information in our catalogs should shake up catalogers and rule-making bodies,” Shroyer concluded, and promoted changes in displays, linking practice, and serials cataloging rules.<sup>82</sup>

In “Catalog/Cataloging Changes and Web 2.0 Functionality,” Kemp provided a useful summary of developments in cataloging standards, theory, and catalog display, with an eye to their effect on serials practice, and she explored the ways “changes in the catalog will be influenced by a new conception of the way users interact with the web, dubbed Web 2.0.”<sup>83</sup> Using five tasks, from finding an unknown journal article on a topic to finding a complete journal run, Kemp predicted how these changes would affect serials search and display in the online public access catalog (OPAC). “Now that we know what is possible, our challenge is to determine how libraries can partner with vendors, publishers, and users to create a maximally effective catalog for finding serials and any other information resource in the library’s collection,” she concluded.<sup>84</sup>

Collins’ “Orphans Adopted Eighty-Nine Years After Conception” described the cataloging issues presented by updating loose-leaf publications—issues that tend to arise from their poor fit into the serials/monographs dichotomy—and reviewed the history of their neglect in American cataloging codes, with an emphasis on recent developments.<sup>85</sup> Collins concluded that with the 2002 revision of AACR2 chapter 12 and the corresponding changes to MARC 21, particularly the creation of the one-byte code “i” for integrating resources, cataloging practice at long last appropriately accommodated loose-leaves.

Fee’s “Do You Have Any Ditko?: Comic Books, MARC, FRBR and Findability” used Tarango’s “FRBR for Serials: Rounding the Square to Fit the Peg” and CONSER’s “An Approach to Serials with FRBR in Mind” as starting points for applying the *FRBR* model to the treatment of comic books as serials, with sample records to illustrate the results.<sup>86</sup> As Fee discussed, comic books present cataloging challenges because of their complex relationships between titles, questions of primary intellectual responsibility, and source-of-title issues, among other questions. Classification and subject treatment, including the use of genre terms, have often been locally defined, and the use of contents fields for analytic description also has presented problems. Fee proposed that despite the difficulties, better comics cataloging might promote format recognition as well as higher use.

Whether to allocate scarce resources to provide catalog access to free digital materials remained a vexing question for many libraries. To address the issue at their institution, Brown and Meagher tracked three years’ use of carefully

selected free digitized resources.<sup>87</sup> They found that use statistics provided support for the value of adding URLs to catalog records. Harcourt, Wacker, and Wolley described Columbia University Libraries’ approach to providing catalog records for free Internet resources, which involved selector input data, the use of an automated CGI program in Practical Extraction and Reporting Language (PERL) to harvest metadata, and access-level records.<sup>88</sup>

## Metadata and Cataloging in the Web World

As libraries continued to move into a more varied information environment, catalogers continued moving into less familiar territory as well. Developments in metadata standards and practice, interoperability between schema, and applications to library service were among the topics discussed in the literature. Writing about the Semantic Web seemed to fall into two categories: how to harness Semantic Web applications to library metadata and processes, and how to integrate library metadata into the data harvesting and processing activities in the interoperable environment of the Semantic Web—both directions, of course, indicating that the world of bibliographic control had grown bigger than library cataloging.

*Metadata*, by Zeng and Qin, presented an overview of metadata standards and encoding schema, with background on general principles, records structure, quality measurement, and interoperability.<sup>89</sup> Intended both as a textbook for students and a resource for practitioners, this work provided guidance in making informed decisions about metadata and digitizing projects.

Journal articles also provided introductory material. For example, Harpring presented a pithy introduction to CCO in “CCO Overview and Description.”<sup>90</sup> Whittaker described the salient points of Describing Archives: A Content Standard (DACS) and compared DACS with the developing RDA.<sup>91</sup> Alexander’s “Core Cataloging and Metadata Standards and Best Practices,” a very useful and concise overview of metadata schema for science and technology materials, described several schema, including their applications and full documentation, as well as encoding schemes, interoperability, crosswalks, and opportunities for continuing education in this area.<sup>92</sup> “Metadata for All: Descriptive Standards and Metadata Sharing Across Cultural Heritage Communities,” by Elings and Waibel, explained key concepts for understanding metadata standards; briefly discussed the primary standards used by library, archive, museum, and visual resources communities; and proposed reconceptualizing standards as “material specific, not limited to one particular community.”<sup>93</sup> “What emerges,” they concluded hopefully about the current metadata environment, “is not a picture of visual resources, libraries, archives

and museums promulgating different standards to describe the same materials, but a rich toolset of descriptive practices that are uniquely adapted to the particular material type they have been originally designed to characterize.<sup>94</sup>

*Using the Open Archives Initiative Protocol for Metadata Harvesting*, by Cole and Foulonneau, described the process of conceiving, implementing, and maintaining an Open Archives Initiative (OAI)-compliant digital repository.<sup>95</sup> This work defined the scope and purpose of OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) and provided context for its use and development in the scholarly communications process. In addition to detailing this protocol's technical aspects, the work examined organizational issues in implementing digital library services with aggregated metadata. The authors noted that "facilitating the transport of metadata is not the same as facilitating the generation and use of metadata end to end in a digital library context. Metadata creation and use remain in large part a human-mediated process, and human factors affect the success of the process."<sup>96</sup> Not unlike the LCWG, they reminded us that "users are looking for resources, not metadata."<sup>97</sup>

In "Knitting the Semantic Web," a thematic double issue of *Cataloging & Classification Quarterly* edited by Greenberg and Méndez, contributions were predominately (but not only) from information managers and information engineers.<sup>98</sup> McCathieNeville and Méndez summarized the value of using Resource Description Framework (RDF) in library metadata to support "global and interoperable Web information processing."<sup>99</sup> Harper and Tillet reviewed potential and theoretical uses of Semantic Web applications for controlled vocabulary, relator terms, authority files, and authority control in general, and suggested that applying these concepts to Semantic Web technology would provide better results for both human and machine.<sup>100</sup> Greenberg also examined ways in which traditional library functions could be applied to the Semantic Web and drew parallels between circulation and digital resource use, collection development and document selection, cataloging standards and metadata standards, schema, and ontologies.<sup>101</sup> Miles and Pérez-Agüera proposed that Simple Knowledge Organization System (SKOS)—a Semantic Web language currently in development for representing controlled, structured vocabularies—could be applied to resource collections that are part of the Semantic Web.<sup>102</sup>

"Framework for a Bibliographic Future," by Coyle and colleagues, proposed a metadata system with four layered components.<sup>103</sup> These components consisted of an "abstract metadata model" as the foundational level, providing basic structures and relationships; a "domain model," defining structures and relationships in the domain or application the metadata addresses; guidance for creating or assigning values; and encoding, with the assumption that metadata

might be encoded in different machine-readable formats while still being exchangeable. The draft document presented possibilities for a bibliographic description set based on Dublin Core Abstract Model, including some *FRBR* entities, and provided an example of a display.

Journal articles summarized successful metadata projects that other libraries could use as models. In "Integrating a Digital Library and a Traditional Library," Benedetti, Cody, and Hanerfeld described a project in which individual metadata records for digital objects created by faculty were harvested, crosswalked into MARC and Qualified Dublin Core, and migrated into the library's catalog and OCLC WorldCat.<sup>104</sup> "An Operational Model for Library Metadata Maintenance," by LeBlanc and Kurth, presented a model for metadata maintenance loosely based on J. A. Zachman's descriptive framework for information systems architecture and drawing on the assumption that libraries can apply catalog maintenance skills to a broader information environment.<sup>105</sup> Using their model, libraries can examine and create interdepartmental and interinstitutional metadata maintenance workflows beyond their MARC-based catalogs. "Moving Beyond MARC: Initiating and Embracing Change in a Traditional Technical Services Department," by Feltner-Reichert and Veve, described the process one department undertook to integrate metadata production resulting from a digitization project into its workflows, and the training, team-building, and workflow changes required to succeed.<sup>106</sup>

Brown and Harvey's "Adding Archival Finding Aids to the Library Catalogue: Simple Crosswalk or Data Traffic Jam?" reported on a project to convert Encoded Archival Description (EAD) finding aids for archival collections to MARC 21 for loading into the local catalog.<sup>107</sup> While archivists generally reject MARC 21 as too limited for archival purposes, integrating finding aids into the catalog may increase their exposure; linking to the full finding aid mitigates limitations of MARC 21. Brown and Harvey described the crosswalk between EAD and MARC 21 (a relatively simple process with software like MARCEdit) and compared an EAD record with the postconversion MARC record. While no data was yet available to show whether the project actually increased discovery of archival collections, the authors believed their "experience suggests that the possibilities of joint library and archives cataloguing project are viable and reap benefits for both parties."<sup>108</sup>

## Classification and Subject Access

Thirty new editions of LCC schedules were issued during 2007–8, and several included significant changes and new features.<sup>109</sup> For example, a number of law schedules included changes and simplifications made to form division

tables; Arabic script and Chinese characters, in addition to equivalent Latin script, appeared in pertinent areas of several schedules. The 2007 edition of *M, Music* had a number of new features, from updated captions, more explanatory notes, and many classification numbers not previously published, to a greatly revised index with many more terms from LCSH. Interestingly, *E–F, History of the Americas* included a reinstatement: Its preface noted that the “lengthy histories that had originally appeared under the name of each U.S. state and Canadian province in the 1913 edition but were removed from subsequent editions, have been restored.”<sup>110</sup>

New print editions of DDC were not issued during the biennium. However, the electronic versions of DDC—WebDewey and Abridged WebDewey—were regularly updated to incorporate new numbers and index entries. In addition, the Dewey website ([www.oclc.org/dewey/default.htm](http://www.oclc.org/dewey/default.htm)) provided mappings between new entries in LCSH and Dewey numbers. A new textbook in the use of DDC, Satija’s *The Theory and Practice of the Dewey Decimal Classification System*, published in England and written by a library and information science (LIS) professor at Guru Nanak Dev University in India, was itself an indication of the international scope of DDC use.<sup>111</sup> Satija emphasized the history, development, governance, and philosophy of DDC and perhaps included fewer exercises than might be found in a North American textbook on the subject; a reviewer in an Australian library journal praised the work’s clarity and emphasis on concepts and made the suggestion of using it in conjunction with “a practical course on number synthesis and number building.”<sup>112</sup>

A revised edition of *The Universal Decimal Classification: A Guide to Its Use*, by McIlwaine, was published by the UDC Consortium in 2007.<sup>113</sup> This edition did not represent a complete rewriting of the 2000 edition, but it did incorporate changes made through 2006 and presented updated examples. While the UDC is not used widely in North America, it continues to be popular internationally, and interest in its potential in automated applications—for example, as a mapping mechanism across domains or languages, or as a source of structured vocabulary—has grown in recent decades. In “Use of the Universal Decimal Classification: A World-Wide Survey,” Slavic presented results of her study using e-mail interviews with LIS professionals in 208 countries, and a literature review.<sup>114</sup> While the survey did not provide information about the number of institutions using UDC, results did show use in 124 countries (60 percent), with UDC the main classification system in 34 countries (28 percent) in Europe, Africa, and Asia. Slavic believed this was evidence that UDC remained “an international de facto standard in indexing.”<sup>115</sup>

In an article in *Knowledge Organization*, Green identified a crucial paradox of classification systems: “Relationships

are at the very heart of knowledge organization,” but knowledge organization schemes do not necessarily express relationships well.<sup>116</sup> After identifying the types of entity classes and relationships important in knowledge organization, Green discussed these classes and relationships in *FRBR*, bibliographic catalogs, DDC, and subject thesauri. She concluded, “Despite the centrality of relationships, their expression in knowledge organization schemes seldom rises to full and systematic expression.”<sup>117</sup>

While North American librarianship has focused on classification as a device for shelf organization, classification also can be used for identification and retrieval from catalogs and other databases. Bland and Stoffan described the development of a “classified browse” feature in the online public catalog of Western North Carolina Library Network (WNCLN).<sup>118</sup> Extracting LC class numbers and associated subject headings, the system built a hierarchical classification display with descriptive captions to offer searchers an additional way of identifying pertinent resources.

Following the practices of bookstores, some libraries have abandoned the use of traditional library classification for shelf arrangement in favor of shelving by genre, general subject categories, and so on. “Finding What You’re Looking For: A Reader-Centred Approach to the Classification of Adult Fiction in Public Libraries,” by Maker, proposed a variation of genre arrangement based on the target reader market (e.g., literary fiction versus popular fiction).<sup>119</sup> On the other hand, Brett’s “Classification Practice in Law Libraries: A Brief Survey” found an increasing use of subject classification schemes in law libraries in the United Kingdom and Ireland.<sup>120</sup> While Moys was the most frequently used scheme, many law libraries used classification schemes of their own devising. “The Making of a Classification Scheme for Libraries of Judaica” outlined the history, development, structure, and applications since the 1950s of *A Classification System for Libraries of Judaica*, from the point of view of one of its originators, David H. Elazar (brother of Daniel Elazar, the system’s other author).<sup>121</sup> The system was organized “according to Jewish concepts and based on Jewish thought and terminology,” which enhanced its suitability for browsing special collections of Judaica.<sup>122</sup> These articles demonstrated the wide divergence of classification practice that persists internationally and among different types of libraries, and even among libraries of the same type.

In “Thesauri and Facets and Tags, Oh My! A Look at Three Decades in Subject Analysis,” Schwartz reviewed developments in this area from the 1970s to the present, observing that facet analysis tended to lead to “the systematic discovery and assembly of the syndetic and semantic structure—the relationships intended to lead indexers and users around the vocabulary and promote match between query description and item description.”<sup>123</sup> Unfortunately, the relationship structure has not been as well implemented in the

online information setting as it has been in printed indexes or other tools. For example, in online catalogs, “relationships are typically displayed during search, but are neither clearly nor helpfully presented in most systems.”<sup>124</sup> Schwartz considered thesauri, folksonomy, and guided navigation as areas in which researchers in subject analysis were responding to the new information environment. Folksonomy (or social, collaborative, or open tagging) provides the user some personal organization of known items; though specific and personal, folksonomy has a future as an augment to controlled vocabulary: “They are not mutually exclusive.”<sup>125</sup> Facets in guided navigation, already implemented in Web-based commerce (a development that brought the work of Ranganathan to a wider audience, Schwartz noted), was beginning to show up in bibliographic catalogs.

“The Structure and Form of Folksonomy Tags: The Road to the Public Library Catalog,” by Spiteri, examined a sampling of user-created tags from Delicious, Furl, and Technorati to determine how they matched with the National Information Standards Organization (NISO) standards for controlled vocabulary.<sup>126</sup> Given the constraints of tagging conventions at the three sites (such as Delicious’s requirement that tags not contain spaces), the sample indicated that in terms of the types of concepts, the use of nouns, the use of alphabetic characters and correct spelling, and the use of a single word to represent a single concept, the tags closely followed the NISO standards. In the use of homophones and other ambiguous terms, nouns in the singular, and abbreviations, however, the sample tags did not conform to the NISO standard. Spiteri concluded with the suggestion that libraries that implement folksonomy creation should provide guidance to users in tag creation (particularly in the construction of multiterm tags), a link to reference works that would allow users to disambiguate homographs, instructions about using singular and plural forms of nouns, and an acceptable use policy. Peterson’s “Parallel Systems: The Coexistence of Subject Cataloging and Folksonomy” briefly described catalogs that employed both controlled subject vocabulary and user-supplied subject tags—including Amazon, the University of Pennsylvania’s PennTags, and the Montana State University electronic theses and dissertations database (ETD)—as well as projects that encouraged patrons to take data or records and reuse them in individual databases, and projects that encouraged users to contribute links and images to supplement the database.<sup>127</sup> While social tagging may have potential for popularity, use statistics for PennTags and ETD were not yet high. Peterson concluded that “the question of whether library users look for the social interaction features of Web 2.0 in the databases where they conduct library research remains.”<sup>128</sup>

Tag clouds, used by tagging services and more formally known as weighted lists, have generated interest as

retrieval mechanisms. “The Folksonomy Tag Cloud: When Is It Useful?” reported the results of an experiment in which researchers took a discrete amount of information, had study participants tag it, then set the participants specific information-seeking tasks using either a tag cloud or a search interface.<sup>129</sup> Open-ended and nonspecific searches were successful using the tag cloud; the tag cloud was useful for browsing or as a starting point for searches because the visual summary could be used to familiarize a searcher with the information domain. On the other hand, specific information was retrieved more easily with the search interface. Authors Sinclair and Cardew-Hall made this interesting observation about user-created metadata: “Individuals ostensibly create tags to serve their own needs, and in doing so, a consensus vocabulary emerges.”<sup>130</sup> Should this consensus vocabulary become a controlled vocabulary? Noruzi’s “Folksonomies: Why Do We Need Controlled Vocabulary?” an editorial in the online journal *Webology*, proposed that a folksonomy-based system *should* use a thesaurus.<sup>131</sup> Reasons included the resolution of the singular-plural bugaboo, correction of typographical errors, use of preferred terms in the case of synonyms, and the consistency of depth or specificity of tagging, among other concepts familiar to library catalogers. Noruzi concluded:

Folksonomy-based systems can employ optional *authority control* of subject keywords, place, personal or corporate names and resource titles by connecting the system to established *authority control* files or controlled vocabularies using new techniques. A folksonomy-based system needs a controlled vocabulary and a suggestion-based system. . . . In the future, it should be possible for search engine designers to design folksonomy-based engines with controlled vocabularies in different fields to improve web information retrieval.<sup>132</sup>

“Measuring the Extent of the Synonym Problem in Full-Text Searching” reported on a study in which a sample of ninety single-word synonym pairs was searched for, singly and jointly, in the Yahoo! database.<sup>133</sup> Authors Beall and Kafadar began their study using Google, but found they could not predict or explain significant discrepancies in the numbers of websites retrieved. Findings showed that the extent of the synonym problem depended on whether one searched the more common of the synonyms. As many as 30 percent of sites might be missed in the great majority of common word pairs; full retrieval required searching on synonyms. The authors believed “the data demonstrate the value of vocabulary control and cross references in providing more precise search results,” and suggested their method could be used to establish a benchmark data set regarding specific search algorithms’ abilities to minimize

the synonym problem.<sup>134</sup> As shown by this article, the possible applicability of approaches associated with traditional library cataloging—such as a system of cross references, whether seen or unseen by the searcher—to the broader universe of information control is clearly an area for further investigation and development.

### Questions of Diversity and Diverse Perspectives

Many libraries in North America are evaluating and modifying services and collections to include and better reflect diverse ethnicities, peoples, and perspectives. The literature of cataloging and classification in 2007–8 contributed to these initiatives. While the contributions discussed in this section might easily be included under different categories, we have collected them here to draw attention to this developing area of the literature.

Not all the authors would agree with this separation, we suspect. As Christensen noted, a “minoritizing view” in bibliographic access tools draws attention to difference, while a “universalizing view . . . calls for unmarked representation, terminology and hierarchical structure that don’t call attention to differences, emphasizing instead the unified whole.”<sup>135</sup> Using Greenblatt’s 1990 study on gay- and lesbian-related terms in LCSH as a starting point, Christensen’s “Minoritization vs. Universalization: Lesbianism and Male Homosexuality in LCSH and LCC” examined gay- and lesbian-related terminology in LCC, as well as changes in LCSH during the last two decades. He concluded, “Keeping up with current connotations and usages of various terms won’t ensure we please everyone, . . . but it will allow us to classify and describe concepts in a way consistent with current usage and with as much respect as possible to the various people these terms describe.”<sup>136</sup>

In “Dewey Deracialized: A Critical Race-Theoretic Perspective,” Furner introduced critical race theory as a framework for evaluating library classification schemes. Applying the theory to DDC and particularly to table 5 of DDC 21 and 22, Furner discussed both the historical significance and the results of changes in recent editions related to racial categories: With the changes, it became more difficult to find materials on topics related to racial categories, especially for racially mixed people. He observed, “We might consider that any decision taken to prevent classifiers and searchers from the use of racial categories is to ignore an everyday reality in which those categories are invoked not only in the distribution of social and political power, but also in individuals’ self-identification.”<sup>137</sup>

“Subject Headings for Aboriginals: The Power of Naming” examined the weaknesses of LCC and LCSH for classifying and describing materials from Native American

cultures in the United States and First Nations in Canada, and it described proposed solutions by Library and Archives Canada (LAC).<sup>138</sup> A consultative process between the Working Group on Collection Policies at the National Library of Canada (now LAC) and aboriginal groups, continuing with an advisory group called Committee on Aboriginal Resources and Services, resulted both in the identification of problem areas and concerns and in specific recommendations. Author Kam briefly identified related initiatives in the United States and in New Zealand and Australia; discussed the use of more culturally appropriate classification systems, such as Brian Deer Classification (developed by one of the earliest First Nations librarians in Canada); and described other efforts to develop accurate representations of First Nations and Aboriginals in catalogs. She observed that

progress is being made towards more culturally sensitive language for subject headings to describe First Nations materials. This progress is largely the result of a consultative process which, although lengthy, will most likely be a critical element in its success. . . . In essence, the existence of these revised terms will push the language boundaries of subject headings to accommodate different perspectives and worldviews leading to a richer and more dynamic reflection of societies and cultures.<sup>139</sup>

In “North American Indian Personal Names in National Bibliographies,” an essay in the volume *Radical Cataloging*, Exner (Little Bear) identified the characteristics of North American Indian name forms and naming traditions throughout life, showed how these name forms need special attention in terms of cataloging and authority control practice developed for European names, and analyzed the treatment of a test set of 185 North American Indian names in the online catalogs of ten national libraries.<sup>140</sup> His study revealed a lack of consistency that could be problematic for the international authority control movement. “Don’t Class Me in Antiquities! Giving Voice to Native American Materials,” also in this collection, presented a discussion between Webster and Doyle of issues related to cataloging and classifying these resources.<sup>141</sup> Included was a list of examples of alternative thesauri and classification schemes from North America, Australia, and New Zealand, developed to help provide accurate and culturally appropriate cataloging for indigenous works.

Chapman addressed resource discovery in the catalog from the perspective of visually impaired individuals.<sup>142</sup> Interestingly, in light of Bowman’s history of annotations in cataloging (see “Cataloging History” below), Chapman advocated the inclusion of summaries or abstracts in catalog records for the visually impaired, since this population is

likely to rely on the catalog entry rather than the dust jacket in making selections.

In “Subject Access for Readers’ Advisory Services: Their Impact on Contemporary Spanish Fiction in Selected Public Library Collections,” Hall-Ellis considered questions related specifically to readers’ advisory, but also more generally addressed how enhancing the catalog with Spanish subject headings could provide better access to the collections for Spanish-speaking patrons.<sup>143</sup> Unfortunately, the study revealed that even for the five public library systems with significant Spanish-speaking populations under study, there was no consistent addition of subject headings in Spanish. “Therefore, delivering readers’ advisory services to patrons who seek contemporary Spanish language adult fiction titles requires a reliance on independent knowledge on the part of the librarian or reader,” while a catalog with Spanish-language headings might enable independent discovery by those without this level of knowledge.<sup>144</sup>

Strottman analyzed problems in the subject treatment of cultures and history in a specific region of the United States in “Some of Our Fifty Are Missing: Library of Congress Subject Headings for Southwestern Cultures and History.”<sup>145</sup> To fill gaps in coverage, Strottman recommended the submission of new headings to the Subject Authority Cooperative Program (SACO) of the Program for Cooperative Cataloging (PCC); to correct misrepresentations and biased and inaccurate headings more systematically, she proposed establishing a SACO Southwest Funnel Group.

Jiang’s “Lost in Translation: The Treatment of Chinese Classics in the Library of Congress Classification” showed that the group of ancient works known as the Chinese Classics (called *jing* in Chinese) has been misinterpreted and therefore misplaced in the LCC; Jiang recommended moving them from “Literature” to “Philosophy” for a more accurate treatment.<sup>146</sup> Park used Korean examples to demonstrate issues in cross-language and cross-cultural access to names and subjects. Park observed that “natural language is not just mere arrangements of words, but the mirror of culture” and that the development of cross-lingual subject access schemes was “hindered by the lack of common conceptual mapping criteria that are interoperable across languages and culture.”<sup>147</sup> “The Colonial Bias: Library Classification in Aotearoa New Zealand,” by Bednarek, described the difficulties Western cultural assumptions create for Maori patrons trying to locate library materials or even comprehend the role of the library as an institution.<sup>148</sup> While most university libraries in Aotearoa New Zealand use LCC to classify their collections, Maori epistemology does not map well to the Western European epistemology that informs LCC.

“A Drum Speaks: A Partnership to Create a Digital Archive Based on Traditional Ojibwe Systems of Knowledge,”

by Powell, reflected on a project to catalog, preserve, and digitize objects created by the Ojibwe people of northern Minnesota.<sup>149</sup> This article asked a number of complex and difficult questions for librarians and others involved in the digitizing project, among them the following:

- Does digital media inadvertently encode western epistemologies into the programming, design, and interface of Web-based learning environments?
- What would a digital archive . . . look like if designed in close cooperation with respected members of Ojibwe communities?
- Is digital technology . . . better able to represent and integrate traditional “texts”—such as oral histories, beadwork, pictographs etched on birch bark, dance, drumming, and songs—than its predecessor, print culture?<sup>150</sup>

Powell reported that the project attempted to utilize “Ojibwe language, stories, and knowledge to shape the tags, codes, and metadata that constitutes the digital architecture of the site,” and that those involved in the project were

still struggling together to determine how this living system of knowledge can be translated into digital codes. We have begun to understand that stories should be more important than categories . . . and we are all beginning to realize that the Ojibwe are the most qualified to create this new folksonomy, once the digital architecture has been put into place.<sup>151</sup>

Ultimately, the article raised questions about the usefulness or appropriateness of applying U.S. librarianship’s codified structures to the intellectual work of native cultures and of attempts at culturally neutral and value-neutral descriptions and knowledge organization.

Providing further food for thought for subject catalogers was Olson’s “How We Construct Subjects: A Feminist Analysis.”<sup>152</sup> Olson explored possibilities for an alternative model of information organization that would emphasize “connected knowing” rather than hierarchical classification structures. She identified “four traces of connectedness—associative relationships, facets, FRBR, and [collaborative] tagging” already present in existing systems, and she proposed that FRBR, as an entity-relationship model, held further potential in this context, along with other possible approaches.<sup>153</sup> Using Olson’s suggestions to address the weaknesses in the expression of relationships in knowledge organization schemes, as observed by Green, and combining these insights with the insights gained in initiatives to include and reflect other worldviews, could result in new departures for knowledge organization and representation.<sup>154</sup>

### Other Reports of Practice and Research

Acquiring new materials that have already been cataloged and processed is an appealing strategy for libraries facing staff reductions. "Shelf-Ready Books Using PromptCat and YBP: Issues to Consider," by Walker and Kulczak, analyzed results of their library's shelf-ready project and presented information useful to other libraries contemplating their own outsourcing contracts.<sup>155</sup> The authors found that an individual review of each piece remained necessary to meet the library's quality standards (such as the presence of subject headings and the verification of URLs in 856 fields) or to conform to local library practice (such as locations based on size or classification practice for juvenile materials). While confirming that acquiring shelf-ready materials was a time-saver, the study also revealed that "accepting vendor-supplied records into our catalog without review would be to invite an unacceptable number of access errors."<sup>156</sup>

The use of MARC record services to provide records for individual electronic serials in dynamic databases is another strategy for maintaining complete catalog coverage with limited staff. "MARC Record Services: A Comparative Study of Library Practices and Perceptions," by Kemp, presented results of a survey of libraries using these services.<sup>157</sup> While the majority of responding libraries used the separate-records approach, a significant number used some form of single-record treatment for titles in multiple formats and databases. Many libraries modified the records once they were loaded, and some found that brief records caused problems because of a lack of links to related titles. While respondents expressed general satisfaction with a service that provided access that might otherwise not be available, many respondents also indicated that record accuracy could be improved.

"A Survey of Local Library Cataloging Tool and Resource Utilization," by Miksa, reported on the use of specific cataloging and classification tools by staff in 103 public libraries in north Texas.<sup>158</sup> Most of the libraries used an automated library system and performed their own cataloging (which mostly fell into the category of copy cataloging), and their overall use of standard cataloging documentation was low. Miksa also found "a disturbing lack of participation in the area of professional communication and the exchange of information" (e.g., subscriptions to electronic discussion lists or pertinent journals).<sup>159</sup> Miksa used the results as a springboard to pose a number of questions about professional preparation for the future and "how well cataloging educators have prepared students to be catalogers."<sup>160</sup>

Starting in spring 2005, the University of Michigan Library performed a comprehensive review of selection, acquisition, and cataloging workflow, prompted in part by migration to a new integrated library system and also by the library's partnership with Google "to digitize the entire University Library collection."<sup>161</sup> "Catalog Information and

User Expectations in an Amazoogole World: Too Much? Too Little?" by Knott and colleagues described how one of the library's working groups gathered information about user needs and behavior as part of this review: a brief online survey intended to reveal users' impressions of their own catalog search behavior, and data from actual search logs. Results were combined to provide support for proposals to enhance records, preferably by automated means; to link print and electronic versions of resources; and to continue current levels of subject analysis, shelflisting for collocation purposes, and authority control, among other recommendations. Ultimately, the library made changes that read like a prequel to the LCWG recommendations: more staff dedicated to digital resources, more cataloging staff time devoted to work with unique print resources, more cataloging and processing outsourcing for widely held materials, and more enhancements to bibliographic records through automated techniques.

"VIAF (Virtual International Authority File): Linking the Deutsche Nationalbibliothek and Library of Congress Name Authority Files" described a project to create a virtual international authority file for personal names from the name authority files of two national libraries to demonstrate both the feasibility and benefits of linking authority records from different national files.<sup>162</sup> Although some difficulties were encountered with the data, automated name matching algorithms were used to link 70 percent of names in common between the two files; linked bibliographic records were also used to enhance authority records to perform further matching and decrease the number of false matches. Authors Bennett and colleagues identified minor changes to authority records that could help with matching and increase the success of such a project in the future: "The long-term goal of the VIAF project is to combine the authoritative names from many national libraries and other significant sources into a shared global authority service."<sup>163</sup>

State and regional library journals may be an underappreciated source of brief articles of value to practitioners on aspects of cataloging operations. For example, "Showtime! Cataloging and Providing Access to Streaming Video Records in the Online Catalog," described issues encountered in cataloging these resources in a consortial environment.<sup>164</sup> Authors McDonald and Johnston gave specifics and examples about the use of MARC fields, reaching consortial cataloging decisions, and authenticating users. "Online Cataloging Tools Versus Print Cataloging Tools," by Jin and Branton, presented the results of a survey of catalogers participating in the electronic discussion lists of four southeast U.S. associations.<sup>165</sup> The authors concluded that online and print versions each had advantages for respondents; they suggested that designers of online tools, with their searching advantages, incorporate some features of the print tools. An unexpected survey result was the low number of recent MLS graduates who received training in the use

of online tools, which may be an echo of Miksa's findings.<sup>166</sup> "The Impact of the Library of Congress' Series Decision on LSU Libraries' Catalog: Minor Damage Now; Long-term Prospects Less Bright," by Nicholson, was one the few studies presenting data on the effects of the LC's 2006 decision to stop creating series authority records.<sup>167</sup> While the study uncovered limited impact thus far, Nicholson anticipated that the long-term effects on authority control and retrieval might well be larger. He concluded "that the best response to the Library of Congress's decision is neither angst nor indifference, but rather a forward-looking resolve to assume greater responsibility for authority control in our libraries and for our cataloging in general."<sup>168</sup>

### Catalogers: Education and Career

Education to produce future catalogers who are prepared for the immediate conditions of employment while ready to lead the specialty into new environments continued to concern an aging profession, and was the topic of many journal articles. Citing posts on AUTOCAT as background, Elrod's opinion piece, "The Case for Cataloguing Education," lamented a perceived decline in the quality of cataloging and linked the decline to the lack of cataloging instruction in library and information science (LIS) education.<sup>169</sup> Using the CONSER standard record (which may include corporate bodies and variant titles as access points without transcription or explanation in the record) and some RDA proposals as evidence, Elrod observed that "some library administrators and catalogers engaged in rule revision and standard setting seem to lack a basis in the principles of cataloguing which should have been a part of their professional education."<sup>170</sup>

Did the perception of a declining emphasis on education in cataloging match conditions "on the ground" in LIS education? In the second installment of his longitudinal study of cataloging education, Joudrey examined courses in organization of information (OI) in ALA-accredited graduate LIS schools in the United States and Canada.<sup>171</sup> He found that since 1998, there had been a 21 percent increase in the number of OI courses offered, though he reported that more than half of the courses included in the study fell "outside of what can be considered library cataloging courses," with metadata and organization courses likely to continue to grow in numbers.<sup>172</sup> Joudrey posited that the biggest threat to the future of cataloging might be the lack of qualified cataloging instructors, and urged, "Catalogers, it is time to get that PhD."<sup>173</sup>

In 2007, the Committee on Education, Training, and Recruitment for Cataloging of the Association for Library Collections and Technical Services (ALCTS) sponsored an ALA preconference intended to address the disconnect between library employers and educators concerning the preparation of cataloging professionals. Titled "What They

Don't Teach in Library School: Competencies, Education and Employer Expectations for a Career in Cataloging," the event inspired the publication of a theme issue of the *Journal of Education for Library and Information Science*: "Making the Connection: Focusing on the Disconnect between LIS Education and Employer Expectations." The issue's lead article, "Educating Cataloging Professionals in a Changing Information Environment," identified core competencies for cataloging professionals, consistent with the goal of producing LIS graduates "with a solid background in information organization and technology and the flexibility and creativity employers expected."<sup>174</sup> Author Hsieh-Yee proposed a concerted effort by educators and practitioners to include the following goals: raise awareness and encourage appreciation of information organization in general and cataloging in particular, prepare students to be both catalogers and metadata specialists, and prepare new leaders in the cataloging area. While this article focused on cataloging education, it was of interest to all catalogers for its acknowledgement of conditions demonstrated or alluded to by many contributions to the literature of the biennium:

What has become apparent is that cataloging as currently practiced in most libraries and other information settings is not a cost-effective solution for managing digital resources. Nevertheless, the principles of cataloging and many concepts related to cataloging, such as authority control and controlled vocabulary, are extremely valuable for bringing order to the ever-expanding information universe.<sup>175</sup>

Practica present a means of balancing education in theoretical and applied aspects of bibliographic control. Damasco and McGurr surveyed entry-level catalogers, now working at member institutions of the Association of Research Libraries, who had participated in a practicum during their library school experience.<sup>176</sup> They found that most of this group of catalogers agreed that a practicum should be a required part of the LIS curriculum because it augmented their classroom studies and provided a valuable means of learning the realities of work in a cataloging department.

"Employer Demands for Cataloger and Cataloger-Like Librarians and Implications for LIS" presented results of an exploratory study on employment in bibliographic control.<sup>177</sup> To determine employer expectations for computer knowledge and skills, traditional cataloging knowledge, prior library experience, and other qualifications, author Lussky analyzed seventy-six ads for cataloging-oriented positions appearing in an online LIS joblist. Lussky found that both traditional cataloging knowledge and knowledge of newer technology standards were required by the majority of ads; the most frequently desired (as opposed to required)

qualifications were familiarity with new technology tools, metadata schemes beyond MARC, and so on. Only 20 percent of jobs required a small amount of library experience. These findings should encourage LIS programs to incorporate practicals into degree programs and to maintain curricula in traditional as well as emerging bibliographic control skills and standards. Zhu analyzed job advertisements for heads of cataloging in academic libraries.<sup>178</sup> This study found—along with continuing core requirements and expectations concerning cataloging—some changes in job titles, responsibilities, and reporting lines, as well as growing expectations for knowledge of non-MARC metadata and digital resources: “The buzz word in the emerging titles was ‘metadata,’” she observed.<sup>179</sup>

Chapman analyzed job descriptions of metadata librarians in seven research libraries incorporating these positions into traditional technical services departments.<sup>180</sup> He identified their roles as collaboration, research, education, and development. He also discussed the benefits of assigning metadata responsibility to a professional librarian, noting that these roles “are descriptive of the responsibilities of professional, twenty-first-century librarians” generally.<sup>181</sup> In “Being a Librarian: Metadata and Metadata Specialists in the Twenty-First Century,” Calhoun considered the changing roles of librarians—especially catalogers and metadata specialists—and the implications for metadata specialists of such trends as technology-driven research and teaching environments, disintermediation in information-seeking behavior, and the emergence of a global infosphere.<sup>182</sup> Calhoun provided examples of effective knowledge management and twenty-first-century librarianship in a large research institution.

### Keeping Current: Columns and Blogs

With many libraries experiencing hiring freezes, layoffs, and furloughs, practitioners were stretched to cover essential services—and many had little time and energy available to review research literature. In this context, thoughtful summaries presented in regular columns or series by outstanding contributors to the field provided valuable continuing education. Blogs by respected practitioners and researchers filled similar roles without the editorial oversight or intervention that can add value to the resource but with great immediacy.

The periodical *Technicalities* described itself as an “information forum for the technical services professional,” and during 2007–8 it provided a significant service to practitioners through its well-selected array of columnists. Among the engaging discussions of issues in bibliographic control were McElfresh’s “When a Journal Isn’t a Journal: Patrons, Catalogs, and Monographic Series,” an entry in her column “Serially Speaking.”<sup>183</sup> McElfresh recounted

real-life difficulties her library’s patrons experienced in accessing individual titles in these series and described a way that improved search properties of new online catalogs might be used to provide better access to these resources. Intner’s “RDA: Progress or Problem?” an entry in her column “Dollars and Sense,” clarified concerns about the developing successor to AACR2.<sup>184</sup> In “Education for Librarianship in the Mid-20th Century, Part Two: Cataloguing, Classification, and Circulation,” an article in her column “Interfaces,” Weihs encouraged readers “to judge whether the fundamentals of librarianship have changed in the past fifty or sixty years and, if so, how great or how little the changes have been.”<sup>185</sup>

Donlan’s “Nexus: Where Reference and Technical Services Meet,” a regular column in *The Reference Librarian*, provided useful approaches for communicating with colleagues in public services. For example, “An Unfortunate Event for Series, or, LC Outsourced You Back” showed the possible impact of the LC’s controversial decision in answering reference queries.<sup>186</sup> Adamich’s regular contributions to *Knowledge Quest* covered topics related to cataloging in school libraries, from CE-MARC to FRBR. “RDA (Resource Description and Access): The New Way to Say ‘AACR2,’” briefly examined the development and characteristics of RDA and stressed the connection between RDA and FRBR.<sup>187</sup> It could serve as an introduction to RDA for staff in many types of libraries, not just those serving K–12 institutions. In his column in *The Unabashed Librarian*, Berman continued his efforts to update and improve LCSH practice. His “One Book, Many Missed Opportunities; or, Why Cataloging Matters (When It’s Done Right)” demonstrated how more useful and complete access might have been provided for Bliss Broyard’s *One Drop: My Father’s Hidden Life, a Story of Race and Family Secrets*.<sup>188</sup> Berman concluded with an “adage for critical catalogers: No matter who cataloged it first, try to make it better. (You usually can.)”<sup>189</sup>

Blogs have two major functions: delivering news and announcements, often with much cross-linking to other blogs, and providing an unfiltered podium—the Internet’s version of Hyde Park’s Speakers’ Corner. While many cataloging-related blogs of the biennium seemed to fall into the news-and-linking category and thus outside the scope of this essay, other blogs provided opinion pieces, professional musings, and personal essays. Dempsey’s Weblog on Libraries, Services and Networks (<http://orweblog.oclc.org>) covered emerging bibliographic control and metadata practices, library service to “network people,” and the application of Web 2.0 concepts to the library environment, as seen by a prominent member of the library profession, and was followed with interest—if not always agreement—by many catalogers. The blog 025.431: The Dewey Blog (<http://ddc.typepad.com/025431>) provided practical advice for the use of DDC, particularly for emerging topics and current events. More personally, Burke, in “The Grim Outlook for

Cataloging,” an August 16, 2008, posting on Brigid’s Blog, observed that “there is a lot of talk these days about the ‘future of cataloging’” and expressed the view that “the future of cataloging looks bleak and confusing. We have a new set of standards and tools being put in place that don’t seem very revolutionary, and yet the hype says they are.”<sup>190</sup>

## Cataloging History

Periods of intense change can prompt consideration of the past as well as the future, and the literature of this biennium included many treatments of historical topics, from cataloging as part of the history of information management, to development of current practices and standards, to memoir and biography. We began this review with literature speculating on cataloging’s future, and we end it with literature reflecting upon its past.

Aiming to “resist the tug of mystical techno-futurism and approach the story of the information age by looking squarely backward,” *Glut: Mastering Information through the Ages*, by information architect Wright, surveyed the entire history of human information management and technologies.<sup>191</sup> While libraries and their organization made many appearances in *Glut*, one chapter, “The Industrial Library,” about the development of library cataloging and classification, was particularly enlightening reading for catalogers, who do not often see Panizzi, Cutter, Dewey, Otlet, and Ranganathan presented within a much larger context. Indeed, *Glut* provided much surprising and useful background to ordinary day-laborers in information organization. And who knew that Cutter once wrote a futuristic essay about libraries in the late twentieth century in which there would be networked telegraphic access to library collections connected by a “fonographic foil”?<sup>192</sup>

“Boston Library Catalogues, 1850–1875: Female Labor and Technological Change,” by Mitchell, presented cataloging history in the context of two related themes: women’s employment history and the development of office organization and technology.<sup>193</sup> In the mid-nineteenth century, the Boston-Cambridge area had four large libraries with growing collections that needed cataloging. Starting in the 1850s, each of the four libraries began hiring women for this work. The women entering previously all-male library workplaces were mostly educated, described as assistants or clerks, and paid considerably less than their male counterparts. Most were unmarried, but not all (Sarah Appleton continued to work as a library assistant after marrying Charles A. Cutter, for example). Mitchell concluded that “libraries played an important role in the development of the female clerical workforce, and in Boston and Cambridge, it was seminal.”<sup>194</sup> Library cataloging, for good or ill, thus influenced the creation of the female office worker. Praising the card catalog as a “system of record management . . . so fundamental that

it can be considered a prototypical form of technology,” Mitchell proposed that

the rise of the card catalogue, and the concomitant entrance of female clerical workers into increasingly bureaucratized [sic] libraries, was a pivotal point not only in the history of libraries. The great library catalogues . . . were catalysts for an extraordinary moment of institutional growth and change.<sup>195</sup>

The evolution of cataloging rules and best practices has, historically, occurred in part within professional library organizations. *Commemorating the Past, Celebrating the Present, Creating the Future: Papers in Observance of the 50th Anniversary of the Association for Library Collections and Technical Services* presented essays documenting both the history of ALCTS (formerly Resources and Technical Services Division, RTSD), the U.S. cataloging community’s organizational “home” within the ALA, and some key developments in ALCTS specialties, including cataloging. Included were reminiscences of RTSD/ALCTS presidents, among them notable contributors to bibliographic control such as Ruth C. Carter, Robert Holley, Janet Swan Hill, Sheila S. Intner, and Michael Gorman.<sup>196</sup> In “The True History of AACR2, 1968–1988: A Personal Memoir, by One Who Was There,” Gorman, editor of AACR2, described the code’s development in the context of earlier codes and emerging international standards, and he recounted his own role in creating AACR2 right down to the music he listened to while writing the draft in longhand (“the Eagles, Linda Ronstadt, etc.”).<sup>197</sup> In addition to this personal view of recent descriptive cataloging history, Gorman proffered the opinion that AACR2, rather than RDA, was the true break with the past, and that, with modifications, AACR2 could still be used as the basis of international and national codes. “Others,” he noted, “think differently.”<sup>198</sup> In a related commemorative work, “Fifty Years of *Library Resources & Technical Services*,” Connell performed a content analysis of articles appearing in this journal from 1957 to 2006.<sup>199</sup> She found that the first thirty years of the journal’s contents concentrated strongly on cataloging topics, but the focus had more recently been shifting: “Cataloging and classification articles still make up the majority on [sic] the content, but the proportion is decreasing.”<sup>200</sup> In addition to national professional organizations, members of the cataloging community have historically been active in regional associations. An interesting contribution by Clemons and Goldberg, “Ohio Valley Group of Technical Services Librarians: A History,” documented the history of this group from its formation as the Ohio Valley Regional Group of Catalogers in 1924 to the present.<sup>201</sup> During the 1950s, the group’s scope expanded into other library technical services; during the 2000s, topics in bibliographic control expanded to include FRBR, XML,

and metadata schemes.

“Cartographic Materials: A Century of Cataloging at Library of Congress and Beyond,” by Mangan, traced the history of map cataloging at the LC during the twentieth century and the concurrent development of cataloging treatment of these resources.<sup>202</sup> Among the topics considered were recurring disputes about rules for the primary access point (e.g., by geographic area or by creator or issuing corporate body); the rejection by LC catalogers of the initial atlas schedule in LC class G and the later cooperative development and acceptance of an LC schedule for both maps and atlases; the development and implementation of MARC for maps; the ups and downs of cataloging rule development; and the effect of digital spatial data on cartographic materials cataloging. “The history of map cataloging history clearly illustrates the problems of having nonspecialists making decisions dealing with the organization, description, and classification of cartographic materials,” Mangan believed.<sup>203</sup> This review of the difficulties of a distinct resources community within library cataloging may be instructive as we attempt to establish collaborative relationships with other communities of practice.

H. E. Bliss (1870–1955) was the creator of the Bliss or Bibliographic Classification scheme and arguably “the first person to attempt a comprehensive and formal statement of the theory of classification.”<sup>204</sup> Broughton’s elegant “Henry Evelyn Bliss: The Other Immortal, or a Prophet Without Honour?” examined the broader context of Bliss’s philosophy of classification, the main features of his work, the originality of his contributions, and the extent to which many of his principles have been fully absorbed into modern classification and indexing theory—and, hence, are often not attributed to him. Of particular interest were discussions concerning a precursor to the concept of citation order in Bliss’s work, his placement of classification within a social context and the resonance of this concept with contemporary theories of social classification or folksonomy, and his contributions to the development of facet structure. Also intriguing were speculations about recent developments in search engines that suggest “the classificatory approach is now regarded as more valuable than it may have been in the early days of digital resource discovery.”<sup>205</sup> Broughton concluded that “Bliss stands shoulder to shoulder with Ranganathan in terms of intellect and influence.”<sup>206</sup>

“Women in Australian Librarianship: The Example of Jean Fleming Arnot,” by McLeod, analyzed the career of Arnot (1903–95), a prominent cataloger and an activist for equal pay for women.<sup>207</sup> The article focused on her working life from the perspective of her experience of “an intellectual and creative task—the art of cataloguing—at which Arnot excelled,” and posited “a link between her professional identity and research skills and her contribution to the women’s movement.”<sup>208</sup> This is a starkly different perspective from

the more usual view of women catalogers historically contributing to “an atmosphere of genteel propriety” and confined “to largely manual skills of copying and filing.”<sup>209</sup>

*Cataloger, Editor, and Scholar: Essays in Honor of Ruth C. Carter*, edited by Holley, honored Carter’s varied career with contributions in many areas of cataloging and classification, including a number of essays on historical topics. Of particular note was Bowman’s “Annotation: A Lost Art in Cataloging,” which described the rise and decline of annotations in cataloging and suggested that new forms of annotation, appropriate for the digital environment, might reemerge.<sup>210</sup> For cataloging historians of the future, the volume included several articles related to Carter’s own contributions, including an interview with Carter by Ewbank, a biographical essay by Henderson, and “*Cataloging & Classification Quarterly*, 1990–2006,” by Roe, Culberston, and Jizba, an analysis of the contents of that journal, an important producer of cataloging literature, during Carter’s tenure as editor.<sup>211</sup>

### Conclusion: Cataloging Literature 2009–10 and Beyond?

This review demonstrates, we believe, the vast extent of the literature of cataloging and classification during 2007–8, its international nature, diverse methods and approaches, varied authorship, expanding concerns, and, especially, its vitality. For example, the literature referring to the concepts of *FRBR* revealed great energy: eagerness to understand *FRBR*’s possibilities, expand *FRBR*’s scope, explain its value to others in the library and information communities, and develop varied applications that used *FRBR* concepts to create links and displays promoting more successful search and retrieval by users. Further opportunities for research include the development of additional strategies for identifying work sets and the comparison of observed search and selection by users in “*FRBR*-ized” catalogs versus traditional OPACs. As additional implementations of *FRBR* concepts develop, additional research questions will surface. For example, is the application of *FRBR* concepts more helpful to some categories of users or to users in some information environments, or when applied to some types of manifestations of works (e.g., editions of nineteenth-century novels) as opposed to other categories of manifestations (e.g., successive editions of works in which currency is critical)? *FRBR* concepts and their possible uses or limitations are likely to inspire the research and publication of cataloging literature in the current biennium, as they did in 2007–8.

Other areas of innovative practice, research, and development should offer abundant opportunities to contribute to the literature as well. One need only look at the flurry of

operational research and commentary resulting from the implementation of AACR2 to anticipate that testing and implementation (or not) of RDA will present rich fields for exploration. The globalization and internationalization of library information sources and applications will require attention to issues of transporting data across languages, character sets, and cultures. The development of metadata standards appropriate for specific cultures and peoples—by those peoples—and questions concerning the desirability or even possibility of creating a culture-neutral metadata standard also are important areas for examination. Projects and research that evaluate the application of some library cataloging approaches to nonlibrary and non-MARC environments could contribute to the larger universe of information access. While library uses of non-MARC metadata are proliferating, we believe the field needs to define underlying principles for optimizing the use of metadata originating in multiple sources and in multiple formats, thereby preventing the user from having lowest-common-denominator access. These topics are among those suggested by recent literature.

Several of the recommendations of the LCWG concerned the development of a culture of research and evaluation regarding issues in bibliographic control to create a stronger evidence base. We anticipate that cataloging and classification literature of the current biennium will be as extensive and vigorous as in 2007–8. With all the talk of the “end of cataloging,” is this burst of productivity the flowering before the tree dies?

The answer, we believe, will not be solely determined by catalogers and researchers in bibliographic control, but will also emerge from the developing information environment and from libraries’ decisions about their roles and how to fulfill them. It is clear that library cataloging will not be the only tree in the bibliographic forest. Nevertheless, considering the richness of much cataloging data, the adaptability of the profession in finding new ways to use this data, and the energy surrounding concepts such as *FRBR* and questions of diversity, we speculate that library cataloging and its literature will continue to develop and make distinctive contributions to the universe of information control.

### References and Notes

1. Why the approximate numbers? We attempted to remove any remaining duplicates and reviews, but this often proved difficult before we connected to the resources themselves. Downloaded citations from sources contained varying amounts of information and made the firm identification of content difficult.
2. Association of College and Research Libraries Rare Books and Manuscripts Section Bibliographic Standards Committee, (ACRL RBMS BSC) and Library of Congress Cataloging Policy and Support Office (LC CPSO), *Descriptive Cataloging of Rare Materials (Books)* (Washington, D.C.: Cataloging Distribution Service, Library of Congress, 2007).
3. *On the Record: Report of the Library of Congress Working Group on the Future of Bibliographic Control* (Washington, D.C.: Library of Congress, 2008).
4. Ibid.
5. Ibid., 4.
6. See, for example, Jeannette Ho, “Enhancing and Upgrading Records for Video Recordings in OCLC’s WorldCat Database: One Participant’s Experience at Texas A&M University,” *Cataloging & Classification Quarterly* 46, no. 3 (2008): 238–68.
7. *On the Record*, 34.
8. Janet Swan Hill, “Entering an Alternate Universe: Some Consequences of Implementing Recommendations of the Library of Congress Working Group on the Future of Bibliographic Control,” *Library Resources & Technical Services* 52, no. 4 (Oct. 2008): 218.
9. *On the Record*, 2.
10. Ibid., 29.
11. Diane I. Hillmann, “Comments on Report of the Library of Congress Working Group on the Future of Bibliographic Control,” [http://docs.google.com/View?docid=dn8z3gs\\_51dsqc77](http://docs.google.com/View?docid=dn8z3gs_51dsqc77) (accessed Dec. 12, 2007).
12. Elaine Sanchez, *Emerging Issues in Academic Library Cataloging & Technical Services* (New York: Primary Research Group, 2007): 132.
13. Library of Congress and Cataloging in Publication Division, *CIP Poised for Change: Survey Findings and Recommendations of the 2006 CIP Review Group* (Washington, D.C.: Library of Congress, Cataloging in Publication Division, 2007).
14. Ibid., 94; *On the Record*, 15.
15. David Bade, “Structures, Standards, and the People Who Make Them Meaningful,” in *Responsible Librarianship: Library Policies for Unreliable Systems* (Duluth, Minn.: Library Juice Press, 2007): 115–35.
16. Ibid., 127.
17. Martha M. Yee, “Will the Response of the Library Profession to the Internet be Self-Immolation?” *Unabashed Librarian* no. 144 (2007): 3.
18. Thomas Mann, ““On the Record’ But Off the Track: A Review of the Report of the Library of Congress Working Group on The Future of Bibliographic Control, With a Further Examination of Library of Congress Cataloging Tendencies,” [www.guild2910.org/WorkingGrpResponse2008.pdf](http://www.guild2910.org/WorkingGrpResponse2008.pdf) (accessed Aug. 15, 2008).
19. David Banush and Jim LeBlanc, “Utility, Library Priorities, and Cataloging Policies,” *Library Collections, Acquisitions & Technical Services* 31, no. 2 (2007): 99.
20. Deanna Marcum, “The Future of Cataloging: Address to the EBSCO Leadership Seminar, Boston, Massachusetts, January 16, 2005,” [www.loc.gov/library/reports/CatalogingSpeech.pdf](http://www.loc.gov/library/reports/CatalogingSpeech.pdf) (accessed Sept. 4, 2009), quoted in Banush and LeBlanc, “Utility,” 101.
21. Banush and LeBlanc, “Utility,” 101, 103.
22. Ibid., 107.

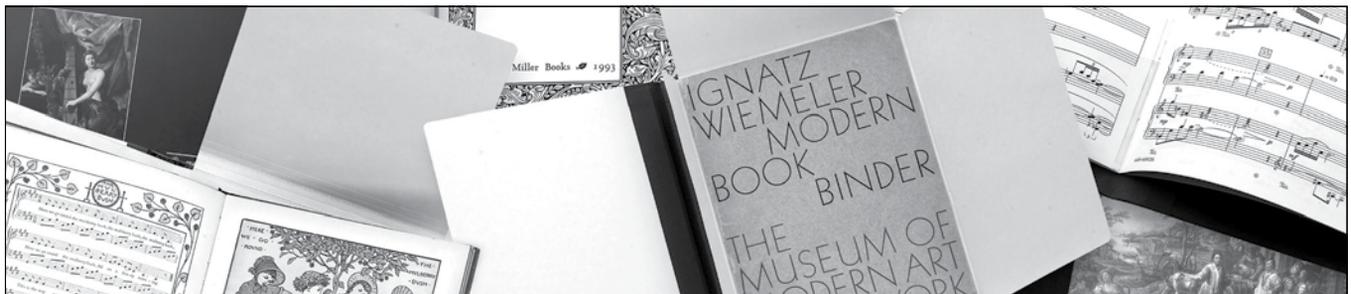
22. Jeffrey Garrett, "Subject Headings in Full-Text Environments: The ECCO Experiment," *College & Research Libraries* 68, no. 1 (Jan. 2007): 69-81. Garrett quotes from Deanna Marcum, "The Future of Cataloging," *Library Resources & Technical Services* 50, no. 1 (2006): 5.
23. Garrett, "Subject Headings," 75.
24. Sheila Bair and Sharon Carlson, "Where Keywords Fail: Using Metadata to Facilitate Digital Humanities Scholarship," *Journal of Library Metadata* 8, no. 3 (2008): 249-62.
25. *Ibid.*, 259-60.
26. Karen Markey, "The Online Library Catalog: Paradise Lost and Paradise Regained?" *D-Lib Magazine* 13, no. 1/2 (2007), [www.dlib.org/dlib/january07/markey/01markey.html](http://www.dlib.org/dlib/january07/markey/01markey.html) (accessed Apr. 25, 2008).
27. *Ibid.*
28. *Ibid.*
29. Tina Gross, "Who Moved My Pinakes?" in *Radical Cataloging: Essays from the Front*, ed. K. R. Roberto (Jefferson, N.C.: McFarland, 2008): 146.
30. Joint Steering Committee for Development of RDA, RDA: Resource Description and Access: Full Draft of RDA, [www.rda-jsc.org/rdafulldraft.html](http://www.rda-jsc.org/rdafulldraft.html) (accessed Aug. 13, 2009); *Anglo-American Cataloguing Rules*, 2nd ed., 1988 revision (AACR2R) (Chicago: ALA, 1988).
31. Jean Weihs and Lynne C. Howarth, "Designating Materials: From "Germane Terms" to Element Types," *Cataloging & Classification Quarterly* 45, no. 4 (2008): 3-24.
32. Laurence S. Creider, "Family Names and the Cataloger," *Library Resources & Technical Services* 51, no. 4 (Oct. 2007): 254-62.
33. Matthew Beacom, "Cataloging Cultural Objects (CCO), Resource Description and Access (RDA), and the Future of Metadata Content," *Visual Resources Association Bulletin* 34, no. 1 (Spring 2007): 81-85.
34. Karen Coyle and Diane I. Hillmann, "Resource Description and Access (RDA): Cataloging Rules for the 20th Century," *D-Lib Magazine* 13, no. 1/2 (2007), [www.dlib.org/dlib/january07/coyle/01coyle.html](http://www.dlib.org/dlib/january07/coyle/01coyle.html) (accessed Apr. 25, 2008).
35. *Ibid.*
37. Michael Gorman, "RDA: Imminent Debacle," *American Libraries* 38, no. 11 (Dec. 2007): 65.
38. *On the Record*, 29.
39. IFLA Study Group on the Functional Requirements for Bibliographic Records and International Federation of Library Associations and Institutions Section on Cataloguing Standing Committee, *Functional Requirements for Bibliographic Records: Final Report*, UBCIM Publications-New Series, vol. 19 (Munich: K.G. Saur, 1998), [www.ifla.org/files/cataloguing/frbr/frbr.pdf](http://www.ifla.org/files/cataloguing/frbr/frbr.pdf) (accessed Mar. 1, 2010).
40. International Conference on Cataloguing Principles (Paris: 1961). *Report*. (London: International Federation of Library Associations, 1963): 91-96.
41. Barbara B. Tillett, "Draft Statement of International Cataloguing Principles: Overview," in *IFLA Cataloguing Principles: Steps Towards an International Cataloguing Code, 4: Report from the 4th IFLA Meeting of Experts on an International Cataloguing Code, Seoul, Korea, 2006*, ed. Barbara B. Tillett, Jaesun Lee, and Ana Lupe Cristán (Munich: K.G. Saur, 2007): 291-308.
42. IFLA Cataloguing Section and IFLA Meetings of Experts on an International Cataloguing Code, Statement of International Cataloguing Principles, [www.ifla.org/en/publications/statement-of-international-cataloguing-principles](http://www.ifla.org/en/publications/statement-of-international-cataloguing-principles) (accessed Sept. 17, 2009).
43. International Federation of Library Associations and Institutions ISBD Review Group and International Federation of Library Associations and Institutions Cataloguing Section Standing Committee, *International Standard Bibliographic Description (ISBD)* (Munich: K.G. Saur, 2007).
44. Lois Mai Chan and Theodora Hodges, *Cataloging and Classification: An Introduction* (Lanham, Md.: Scarecrow, 2007).
45. Sheila S. Intner and Jean Riddle Weihs, *Standard Cataloging for School and Public Libraries*, 4th ed. (Westport, Conn: Libraries Unlimited, 2007).
46. G. G. Chowdhury and Sudatta Chowdhury, *Organizing Information: From the Shelf to the Web* (London: Facet, 2007).
47. *Functional Requirements for Bibliographic Records Final Report*.
48. Robert L. Maxwell, *FRBR: A Guide for the Perplexed* (Chicago: ALA, 2008).
49. Pat Riva, "Introducing the 'Functional Requirements for Bibliographic Records' and Related IFLA Developments," *Bulletin of the American Society for Information Science and Technology* 33, no. 6 (Aug./Sept. 2007): 7-11.
50. Timothy J. Dickey, "FRBRization of a Library Catalog: Better Collocation of Records, Leading to Enhanced Search, Retrieval, and Display," *Information Technology & Libraries* 27, no. 1 (Mar. 2008): 23-32.
51. Arlene G. Taylor, ed., *Understanding FRBR: What It Is and How It Will Affect Our Retrieval Tools* (Westport, Conn: Libraries Unlimited, 2007).
52. Steven C. Shadle, "FRBR and Serials: One Serialist's Analysis," in Taylor, *Understanding FRBR*, 153-74; Sherry Vellucci, "FRBR and Music," in *Ibid.*, 131-51.
53. Murtha Baca and Sherman Clarke, "FRBR and Works of Art, Architecture, and Material Culture," in Taylor, *Understanding FRBR*, 103.
54. Richard Smiraglia, "Bibliographic Families and Superworks," in Taylor, *Understanding FRBR*, 73-86.
55. Glenn E. Patton and IFLA Working Group on Functional Requirements and Numbering of Authority Records, *Functional Requirements for Authority Data: A Conceptual Model* (Munich: K.G. Saur, 2009).
56. IFLA Working Group on Functional Requirements for Subject Authority Records, *Functional Requirements for Subject Authority Data (FRSAD): Conceptual Model, 2nd Draft* (June 10, 2009), [nkos.slis.kent.edu/FRSAR/report090623.pdf](http://nkos.slis.kent.edu/FRSAR/report090623.pdf) (accessed Sept. 19, 2009): 7.
57. Jan Pisanski and Maja Žumer, "Functional Requirements for Bibliographic Records: An Investigation of Two Prototypes," *Program: Electronic Library & Information Systems* 41, no. 4 (2007): 400-417.
58. *Ibid.*, 416.
59. *Ibid.*
60. Allyson Carlyle, Sara Ranger, and Joel Summerlin, "Making the Pieces Fit: *Little Women*, Works, and the Pursuit of

- Quality," *Cataloging & Classification Quarterly* 46, no. 1 (2008): 35–63.
61. *Ibid.*, 61.
  62. Marija Petek, "Bibliographic Families and Relationships among Family Members in COBIB," *OCLC Systems & Services* 24, no. 2 (2008): 91–100.
  63. *Ibid.*, 98.
  64. Ruth Kalf, "FRBR: An Opportunity for Map Collections and Map Users?" *Liber Quarterly: The Journal of European Research Libraries* 18, no. 2 (Sept. 2008): 280.
  65. Betsy Simpson, Jimmie Lundgren, and Tatiana Barr, "Linking Print and Electronic Books: One Approach," *Library Resources & Technical Services* 51, no. 2 (Apr. 2007): 146–52.
  66. *Ibid.*, 150.
  67. Martha M. Yee, *Moving Image Cataloging: How to Create and How to Use a Moving Image Catalog* (Westport, Conn: Libraries Unlimited, 2007): xiii.
  68. Ho, "Enhancing and Upgrading Records."
  69. DVD Cataloging Guide Update Task Force, Cataloging Policy Committee, Online Audiovisual Catalogers, *Guide to Cataloging DVD and Blu-ray Discs Using AACR2r and MARC 21, 2008 Update*, www.olacinc.org/drupal/capc\_files/DVD\_guide\_final.pdf (accessed Dec. 12, 2009).
  70. Playaway Cataloging Joint Task Force: Online Audiovisual Catalogers Cataloging Policy Committee, Music Library Association Subcommittee on Descriptive Cataloging, *Guide to Cataloging Playaway Devices Based on AACR2 Chapters 6 and 9*, (last updated Aug. 21, 2008), www.olacinc.org/drupal/capc\_files/playawaysPDF.pdf (accessed May 4, 2009).
  71. ACRL RBMS BSC and PL CPSO, *Descriptive Cataloging of Rare Materials (Books)*.
  72. Elizabeth O'Keefe, "The Best of Both Worlds: Using CCO for Object Cataloging in Libraries and Special Collections," *Visual Resources Association Bulletin* 34, no. 1 (Spring 2007): 86–96.
  73. Julian Everett Allgood, "Serials and Multiple Versions, or the Inexorable Trend toward Work-Level Displays," *Library Resources & Technical Services* 51, no. 3 (July 2007): 160–78.
  74. *Ibid.*, 175.
  75. Wendy Baia, "Excessive Successive: Time for a Radical Change," *The Serials Librarian* 53, no. 1 (2007): 57–80.
  76. Mary Greci, "How Successive Entry Cataloging Stacks Up in Today's World," *The Serials Librarian* 53, no. 1/2 (2007): 97.
  77. Kevin M. Randall, "Latest Entry Legacies: Confessions of a Guerrilla Cataloger," *The Serials Librarian* 53, no. 1/2 (2007): 81–91.
  78. *Ibid.*, 86.
  79. *Ibid.*, 90.
  80. Andrew D. Shroyer, "Confessions of a Correspondent from the Choice-of-Entry War: Review of Responses to a Set of Informal Opinion Surveys," *The Serials Librarian* 53, no. 1 (2007): 99–123.
  81. *Ibid.*, 116.
  82. *Ibid.*, 118.
  83. Rebecca Kemp, "Catalog/Cataloging Changes and Web 2.0 Functionality: New Directions for Serials," *The Serials Librarian* 53, no. 4 (2008): 99.
  84. *Ibid.*, 109.
  85. Carol Morgan Collins, "Orphans Adopted Eighty-Nine Years After Conception, Or, Cataloging Updating Loose-Leaf Publications," *The Serials Librarian* 55, no. 4 (2008): 598–614.
  86. William T. Fee, "Do You Have Any Ditko?: Comic Books, MARC, FRBR and Findability," *Serials Review* 34, no. 3 (2008): 175–89; Adolfo Tarango, "FRBR for Serials: Rounding the Square to Fit the Peg," presented at the Operations Meeting, Apr. 24, 2008, www.loc.gov/acq/conser/FRBR-for-serials.pdf (accessed June 10, 2009); CONSER Task Force on Universal Holdings, "An Approach to Serials with FRBR in Mind," draft, rev. Diane Hillmann, Jan. 24, 2004, www.lib.unc.edu/cat/mfh/serials\_approach\_frbr.pdf (accessed June 10, 2009).
  87. Christopher C. Brown and Elizabeth S. Meagher, "Cataloging Free E-Resources: Is It Worth the Investment?" *Interlending & Document Supply* 36, no. 3 (2008): 135–41.
  88. Kate Harcourt, Melanie Wacker, and Iris Wolley, "Automated Access Level Cataloging for Internet Resources at Columbia University Libraries," *Library Resources & Technical Services* 51, no. 3 (July 2007): 212–25.
  89. Marcia Lei Zeng and Jian Qin, *Metadata* (New York: Neal-Schuman, 2008).
  90. Patricia Harpring, "CCO Overview and Description," *Visual Resources Association Bulletin* 34, no. 1 (Spring 2007): 34–44.
  91. Beth M. Whittaker, "DACS and RDA: Insights and Questions from the New Archival Descriptive Standard," *Library Resources & Technical Services* 51, no. 2 (Apr. 2007): 98–105.
  92. Mary S. Alexander, "Core Cataloging and Metadata Standards and Best Practices," *Science & Technology Libraries* 28, no. 1/2 (2008): 63–85.
  93. Mary W. Elings and Günter Waibel, "Metadata for All: Descriptive Standards and Metadata Sharing Across Cultural Heritage Communities," *Visual Resources Association Bulletin* 34, no. 1 (Spring 2007): 12.
  94. *Ibid.*, 13.
  95. Timothy W. Cole and Muriel Foulonneau, *Using the Open Archives Initiative Protocol for Metadata Harvesting* (Westport, Conn.: Libraries Unlimited, 2007).
  96. *Ibid.*, 185–86.
  97. *Ibid.*, 189.
  98. Jane Greenberg and Eva Méndez, eds., "Knitting the Semantic Web," special issue, *Cataloging & Classification Quarterly* 43, no. 3/4 (2007).
  99. Charles McCathieNevile and Eva Méndez, "Library Cards for the 21st Century," *Cataloging & Classification Quarterly* 43, no. 3/4 (2007): 30.
  100. Corey A. Harper and Barbara B. Tillett, "Library of Congress Controlled Vocabularies and Their Application to the Semantic Web," *Cataloging & Classification Quarterly* 43, no. 3/4 (2007): 47–68.
  101. Jane Greenberg, "Advancing the Semantic Web Via Library Functions," *Cataloging & Classification Quarterly* 43, no. 3/4 (2007): 203–25.
  102. Alistair Miles and Jose R. Pérez-Agüera, "SKOS: Simple Knowledge Organisation for the Web," *Cataloging & Classification Quarterly* 43, no. 3/4 (2007): 69–83.

103. Karen Coyle et al., "Framework for a Bibliographic Future," May 2007 rev., <http://futurelib.pbworks.com/Framework> (accessed Sept. 23, 2009).
104. Susannah Benedetti, Sue Ann Cody, and Arlene Hanerfeld, "Integrating a Digital Library and a Traditional Library: Librarians and Scientists Collaborating for Sustainability," *Technical Services Quarterly* 24, no. 3 (2007): 15–27.
105. Jim LeBlanc and Martin Kurth, "An Operational Model for Library Metadata Maintenance," *Library Resources & Technical Services* 52, no. 1 (Jan. 2008): 54–59.
106. Melanie Feltner-Reichert and Marielle Veve, "Moving Beyond MARC: Initiating and Embracing Change in a Traditional Technical Services Department," *Against the Grain* 19, no. 5 (Nov. 2007): 28, 30, 32, 34.
107. Geoff Brown and Kathryn Harvey, "Adding Archival Finding Aids to the Library Catalogue: Simple Crosswalk Or Data Traffic Jam?" *Partnership: The Canadian Journal of Library & Information Practice & Research* 2, no. 2 (2007): 1–18.
108. *Ibid.*, 13.
109. A full list of the most recent editions of classification schedules is available at [www.loc.gov/cds/classif.html](http://www.loc.gov/cds/classif.html).
110. Library of Congress Cataloging Policy and Support Office, *Library of Congress Classification. E–F. History of the Americas* (Washington, D.C.: Library of Congress, Cataloging Distribution Service, 2007): iii.
111. Mohinder Partap Satija, *The Theory and Practice of the Dewey Decimal Classification System* (Oxford, England: Chandos, 2007).
112. Doreen Sullivan, "DDC Gem," *Australian Library Journal* 57, no. 4 (Nov. 2008): 484.
113. Ia McIlwaine, *The Universal Decimal Classification: A Guide to Its Use* (The Hague: UDC Consortium, 2007).
114. Aida Slavic, "Use of the Universal Decimal Classification: A World-Wide Survey," *Journal of Documentation* 64, no. 2 (2008): 211–28.
115. *Ibid.*, 224.
116. Rebecca Green, "Relationships in Knowledge Organization," *Knowledge Organization* 35, no. 2/3 (2008): 150.
117. *Ibid.*, 158.
118. Robert N. Bland and Mark A. Stoffan, "Returning Classification to the Catalog," *Information Technology & Libraries* 27, no. 3 (Sept. 2008): 55–60.
119. Richard Maker, "Finding what You're Looking for: A Reader-Centred Approach to the Classification of Adult Fiction in Public Libraries," *Australian Library Journal* 57, no. 2 (May 2008): 169–77.
120. Rachel Brett, "Classification Practice in Law Libraries: A Brief Survey," *Legal Information Management* 8, no. 1 (Spring 2008): 61–63.
121. David H. Elazar, "The Making of a Classification Scheme for Libraries of Judaica," *Judaica Librarianship* 14 (2008): 15–25. See also David H. Elazar and Daniel Elazar, *A Classification System for Libraries of Judaica*, 3rd. rev. ed. (Northvale, N.J.: Jason Aronson, 1997).
122. Elazar, "The Making of a Classification Scheme for Libraries of Judaica," 16.
123. Candy Schwartz, "Thesauri and Facets and Tags, Oh My! A Look at Three Decades in Subject Analysis," *Library Trends* 56, no. 4 (Spring 2008): 833.
124. *Ibid.*, 834.
125. *Ibid.*, 837.
126. Louise F. Spiteri, "The Structure and Form of Folksonomy Tags: The Road to the Public Library Catalog," *Information Technology & Libraries* 26, no. 3 (Sept. 2007): 13–25.
127. Elaine Peterson, "Parallel Systems: The Coexistence of Subject Cataloging and Folksonomy," *Library Philosophy & Practice* (2008): 1–6, <http://unllib.unl.edu/LPP/e-peterson3.pdf> (accessed June 5, 2008).
128. *Ibid.*, 5.
129. James Sinclair and Michael Cardew-Hall, "The Folksonomy Tag Cloud: When is It Useful?" *Journal of Information Science* 34, no. 1 (2008): 15–29.
130. *Ibid.*, 16.
131. Alireza Noruzi, "Folksonomies: Why Do We Need Controlled Vocabulary?" *Webology* 4, no. 2 (2007), [www.webology.ir/2007/v4n2/editorial12.html](http://www.webology.ir/2007/v4n2/editorial12.html) (accessed June 5, 2008).
132. *Ibid.*
133. Jeffrey Beall and Karen Kafadar, "Measuring the Extent of the Synonym Problem in Full-Text Searching," *Evidence Based Library & Information Practice* 3, no. 4 (2008): 18–33, <http://ejournals.library.ualberta.ca/index.php/EBLIP/article/view/4081/3697> (accessed Mar. 16, 2009).
134. *Ibid.*, 29.
135. Ben Christensen, "Minoritization Vs. Universalization: Lesbianism and Male Homosexuality in LCSH and LCC," *Knowledge Organization* 35, no. 4 (2008): 236.
136. Ben Christensen, "Minoritization Vs. Universalization: Lesbianism and Male Homosexuality in LCSH and LCC," *Knowledge Organization* 35, no. 4 (2008): 236, 238. See also Ellen Greenblatt, "Homosexuality: The Evolution of a Concept in the Library of Congress Subject Headings," in *Gay and Lesbian Library Service*, ed. Cal Gough and Ellen Greenblatt (Jefferson, N.C.: McFarland, 1990): 75–101.
137. Jonathan Fumer, "Dewey Deracialized: A Critical Race-Theoretic Perspective," *Knowledge Organization* 34, no. 3 (2007): 165.
138. D. Vanessa Kam, "Subject Headings for Aboriginals: The Power of Naming," *Art Documentation: Bulletin of the Art Libraries Society of North America* 26, no. 2 (Fall 2007): 18–22.
139. *Ibid.*, 21.
140. Frank Exner (Little Bear), "North American Indian Personal Names in National Bibliographies," in *Radical Cataloging: Essays at the Front*, ed. K. R. Roberto (Jefferson, N.C.: McFarland, 2008): 150–64.
141. Kelly Webster and Ann Doyle, "Don't Class Me in Antiquities! Giving Voice to Native American Materials," in *Ibid.*, 189–95.
142. Ann Chapman, "Resource Discovery: Catalogs, Cataloging, and the User," *Library Trends* 55, no. 4 (Spring 2007): 917–31.
143. Sylvia D. Hall-Ellis, "Subject Access for Readers' Advisory Services: Their Impact on Contemporary Spanish Fiction in Selected Public Library Collections," *Public Library Quarterly* 27, no. 1 (2008): 1–18.
144. *Ibid.*, 14.
145. Theresa A. Strotzman, "Some of our Fifty are Missing: Library of Congress Subject Headings for Southwestern Cultures

- and History," *Cataloging & Classification Quarterly* 45, no. 2 (2007): 41.
146. ShuYong Jiang, "Lost in Translation: The Treatment of Chinese Classics in the Library of Congress Classification," *Cataloging & Classification Quarterly* 45, no. 1 (2007): 3–14.
  147. Jung-Ran Park, "Cross-Lingual Name and Subject Access: Mechanisms and Challenges," *Library Resources & Technical Services* 51, no. 3 (July 2007): 186.
  148. Antje Bednarek, "The Colonial Bias: Library Classification in Aotearoa New Zealand," *New Zealand Library & Information Management Journal—Nga Purongo* 50, no. 3 (Oct. 2007): 233–45. Note: The author's name is misspelled as Bednareks at the beginning of the published article and appears in this incorrect form in some indexes.
  149. Timothy B. Powell, "A Drum Speaks: A Partnership to Create a Digital Archive Based on Traditional Ojibwe Systems of Knowledge," *RBM: A Journal of Rare Books, Manuscripts, & Cultural Heritage* 8, no. 2 (Fall 2007): 167–79.
  150. *Ibid.*, 169.
  151. *Ibid.*, 176, 179.
  152. Hope A. Olson, "How We Construct Subjects: A Feminist Analysis," *Library Trends* 56, no. 2 (Fall 2007): 509–41.
  153. *Ibid.*, 531.
  154. Green, "Relationships in Knowledge Organization."
  155. Mary Walker and Deb Kulczak, "Shelf-Ready Books using PromptCat and YBP: Issues to Consider (An Analysis of Errors at the University of Arkansas)," *Library Collections, Acquisitions, & Technical Services* 31, no. 2 (2007): 61–84.
  156. *Ibid.*, 82.
  157. Rebecca Kemp, "MARC Record Services: A Comparative Study of Library Practices and Perceptions," *The Serials Librarian* 55, no. 3 (2008): 379–410.
  158. Shawne D. Miksa, "A Survey of Local Library Cataloging Tool and Resource Utilization," *Journal of Education for Library & Information Science* 49, no. 2 (Spring 2008): 128–46.
  159. *Ibid.*, 145.
  160. *Ibid.*, 144.
  161. Martin L. Knott et al., "Catalog Information and User Expectations in an Amazoogole World: Too Much? Too Little?" *Against the Grain* 19, no. 5 (Nov. 2007): 40.
  162. Rick Bennett et al., "VIAF (Virtual International Authority File): Linking the Deutsche Nationalbibliothek and Library of Congress Name Authority Files," *International Cataloguing & Bibliographic Control* 36, no. 1 (Jan./Mar. 2007): 12–18.
  163. *Ibid.*, 12.
  164. Elizabeth McDonald and Joyce Johnston, "Showtime! Cataloging and Providing Access to Streaming Video Records in the Online Catalog," *Tennessee Libraries* 58, no. 2 (2008): 1–5.
  165. Miao Jin and Ann Branton, "Online Cataloging Tools Versus Print Cataloging Tools," *Southeastern Librarian* 55, no. 3 (Fall 2007): 18–22.
  166. Miksa, "A Survey of Local Library Cataloging Tool and Resource Utilization."
  167. Joseph R. Nicholson, "The Impact of the Library of Congress' Series Decision on LSU Libraries Catalog: Minor Damage Now; Long-Term Prospects Less Bright," *Louisiana Libraries* 71, no. 1 (2008): 3–10.
  168. *Ibid.*, 9.
  169. J. McRee Elrod, "The Case for Cataloguing Education," *The Serials Librarian* 55, no. 1/2 (2008): 1–10.
  170. *Ibid.*, 5.
  171. Daniel N. Joudrey, "Another Look at Graduate Education for Cataloging and the Organization of Information," *Cataloging & Classification Quarterly* 46, no. 2 (2008): 137–81.
  172. *Ibid.*, 175.
  173. *Ibid.*, 178.
  174. Ingrid Hsieh-Yee, "Educating Cataloging Professionals in a Changing Information Environment," *Journal of Education for Library and Information Science* 49, no. 2 (Spring 2008): 101.
  175. *Ibid.*, 93.
  176. Ione T. Damasco and Melanie J. McGurr, "A Survey of Cataloger Perspectives on Practicum Experiences," *Cataloging & Classification Quarterly* 45, no. 4 (2008): 43–64.
  177. Joan P. Lussky, "Employer Demands for Cataloger and Cataloger-Like Librarians and Implications for LIS," *Journal of Education for Library & Information Science* 49, no. 2 (2008): 116–27.
  178. Lihong Zhu, "Head of Cataloging Positions in Academic Libraries: An Analysis of Job Advertisements," *Technical Services Quarterly* 25, no. 4 (2008): 49–70.
  179. *Ibid.*, 54.
  180. John W. Chapman, "The Roles of the Metadata Librarian in a Research Library," *Library Resources & Technical Services* 51, no. 4 (Oct. 2007): 279–85.
  181. *Ibid.*, 283.
  182. Karen Calhoun, "Being a Librarian: Metadata and Metadata Specialists in the Twenty-First Century," *Library Hi Tech* 25, no. 2 (2007): 174–87.
  183. Laura Kane McElfresh, "Serially Speaking: When a Journal Isn't a Journal: Patrons, Catalogs, and Monographic Series," *Technicalities* 27, no. 1 (Jan./Feb. 2007): 1, 11–13.
  184. Sheila S. Intner, "Dollars and Sense: RDA: Progress or Problem?" *Technicalities* 28, no. 4 (2008): 1–15.
  185. Jean Weihs, "Interfaces: Education for Librarianship in the Mid-20th Century, Part Two: Cataloguing, Classification, and Circulation," *Technicalities* 28, no. 5 (2008): 9.
  186. Rebecca Donlan, "An Unfortunate Event for Series, Or, LC Outsourced You Back," *Reference Librarian* 47, no. 98 (2007): 95–99.
  187. Tom Adamich, "RDA (Resource Description and Access): The New Way to Say, 'AACR2,'" *Knowledge Quest* 36, no. 4 (Mar./Apr. 2008): 64–69.
  188. Sanford Berman, "One Book, Many Missed Opportunities: Or, Why Cataloging Matters (When It's Done Right)," *The Unabashed Librarian*, no. 149 (2008): 20–23.
  189. *Ibid.*, 23.
  190. Brigid Burke, "The Grim Outlook for Cataloging," online posting, Aug. 16, 2008, Brigid's Blog, <http://brigidburke.blogspot.com/2008/08/grim-outlook-for-cataloging.html> (accessed Dec. 18, 2009).
  191. Alex Wright, *Glut: Mastering Information through the Ages* (Washington, D.C.: Joseph Henry Press, 2007): 3.
  192. *Ibid.*, 182.
  193. Barbara A. Mitchell, "Boston Library Catalogues, 1850–1875: Female Labor and Technological Change," in *Institutions of Reading: The Social Life of Libraries in the United States*, ed.

- Thomas Augst and Kenneth Carpenter (Amherst: Univ. of Massachusetts Pr., 2007): 119–47.
194. *Ibid.*, 145.
195. *Ibid.*, 147.
196. Pamela Bluh, ed., *Commemorating the Past, Celebrating the Present, Creating the Future: Papers in Observance of the 50th Anniversary of the Association for Library Collections & Technical Services* (Chicago: ALA, 2007).
197. Michael Gorman, "The True History of AACR2, 1968–1988: A Personal Memoir, by One Who Was There," in Bluh, ed., *Commemorating the Past*, 69.
198. *Ibid.*, 72.
199. Tschera Harkness Connell, "A Look at Fifty Years of Library Resources & Technical Services," *Library Resources & Technical Services* 51, no. 4 (Oct. 2007): 237–53.
200. *Ibid.*, 248.
201. Angel Clemons and Tyler Goldberg, "Ohio Valley Group of Technical Services Librarians: A History," *Serials Review* 33, no. 2 (2007): 103–13.
202. Elizabeth Mangan, "Cartographic Materials: A Century of Cataloging at Library of Congress and Beyond," *Journal of Map & Geography Libraries* 3, no. 2 (2007): 23–44.
203. *Ibid.*, 24.
204. Vanda Broughton, "Henry Evelyn Bliss: The Other Immortal, Or a Prophet without Honour?" *Journal of Librarianship & Information Science* 40, no. 1 (Mar. 2008): 46.
205. *Ibid.*, 55.
206. *Ibid.*
207. Louise McLeod, "Women in Australian Librarianship: The Example of Jean Fleming Arnot," *Australian Library Journal* 56, no. 3/4 (November, 2007): 322–34.
208. *Ibid.*, 322, 328.
209. Thomas Augst, introduction to *Institutions of Reading: The Social Life of Libraries in the United States*, ed. Thomas Augst and Kenneth Carpenter (Amherst: Univ. of Massachusetts Pr., 2007): 18.
210. J. H. Bowman, "Annotation: A Lost Art in Cataloging," in *Cataloger, Editor, and Scholar: Essays in Honor of Ruth C. Carter*, ed. Robert P. Holley (Binghamton, N.Y.: Haworth, 2007): 95–111.
211. Linda C. Ewbank and Ruth C. Carter, "An Interview with Ruth C. Carter," in *Cataloger, Editor, and Scholar: Essays in Honor of Ruth C. Carter*, ed. Robert P. Holley (Binghamton, N.Y.: Haworth, 2007): 19–38; Kathryn Luther Henderson, "The Story of Ruth: The Life and Contributions of Ruth C. Carter," in *Cataloger, Editor, and Scholar*, 7–17; Sandra K. Roe, Rebecca Culbertson, and Laurel Jizba, "Cataloging & Classification Quarterly, 1990–2006," in *Cataloger, Editor, and Scholar*, 39–52.



# ARCHIVAL.COM

## INNOVATIVE SOLUTIONS FOR PRESERVATION

### Call for a complete catalog

<i>Pamphlet Binders</i>	<i>Polypropylene Sheet</i>
<i>Music Binders</i>	<i>&amp; Photo Protectors</i>
<i>Archival Folders</i>	<i>Archival Boards</i>
<i>Manuscript Folders</i>	<i>Adhesives</i>
<i>Hinge Board Covers</i>	<i>Bookkeeper</i>
<i>Academy Folders</i>	<i>Century Boxes</i>
<i>Newspaper/Map Folders</i>	<i>Conservation Cloths</i>
<i>Bound Four Flap</i>	<i>Non-Glare Polypropylene</i>
<i>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](mailto:custserv@archival.com)

Web: [archival.com](http://archival.com)

# Rethinking Research Library Collections

## A Policy Framework for Straitened Times, and Beyond

By Dan Hazen

*Academic and research libraries today confront daunting financial pressures. Their faltering budgets also compound an intensifying existential crisis resulting from profound shifts in information, scholarship, technology, and academic organizations. The purposes of collections are particularly uncertain in this radically fluid context. Analyzing the most salient elements in today's collections landscape can help to frame the guiding principles that will inform adaptive new approaches to collections and content.*

Research libraries today contend with shrinking budgets that compound a long-standing structural mismatch between available resources and community expectations. The broader landscapes of information, scholarship, technology, and academic organizations also are in flux. The community's collections strategies must therefore adapt to a radically fluid context that is brimming with both opportunities and demands. This essay describes some key elements in today's collections landscape and also offers a simple model for information types and their uses. The framework in turn suggests a set of principles to inform a redirected strategy for collections and content.

### Universities, Information, and Library Collections: An Environmental Scan

#### The Information Landscape: Continuity and Change

##### Information

The supply of information resources has mushroomed across all formats. Emerging countries and also traditional publishing centers are producing more than ever before. Recession and deflation may mitigate these trends, and some categories of publications—print newspapers are a likely example—may decline or even disappear. Nonetheless, predictions that hard copy publications will soon be overwhelmed by an avalanche of electronic resources are emphatically premature, particularly in the developing world and other areas of emergent modernity. Analog materials remain both prevalent and indispensable as the digital explosion continues apace.

Despite the persistence of print, large-scale digitization is transforming the library world. The scale of the resources available via Internet search engines

**Dan Hazen** (dchazen@fas.harvard.edu) is Associate Librarian for Collection Development, Harvard University, Cambridge, Massachusetts.

Submitted November 9, 2009; accepted November 19, 2009, pending modest revision. Revised and submitted December 4, 2009, and accepted for publication.

This essay is based on a longer discussion paper the author prepared for the Harvard College Library in the spring of 2009.

substantially exceeds that of any single research library, and also of the research library community as a whole. The immense range of virtual materials now at hand has in turn undermined the quantitative measures by which we have traditionally judged our holdings. Libraries and librarians, along with other agencies devoted to our intellectual and cultural heritage, are experiencing a dual crisis of purpose and identity. Most students and scholars perceive less cause for concern.

The day's expanding array of information resources, in all formats, is complemented by intense price pressures that consistently outpace inflation. Publisher conglomerates, which already wield oligopolistic control over the scientific, technical, and medical information universe, are now expanding into other market segments. Greater outputs of increasingly expensive published materials characterize developing regions throughout the world. The weakened dollar, which lost about one-third of its value relative to both the Euro and the Pound between 2000 and 2008, compounds the challenge for libraries with heavily international collections. Information may want to be free, but it also is a commodity—and scholarly resources are in thrall to the marketplace.

#### Scholarship

Research and teaching continue to evolve. Until recently, pedagogical models and research strategies privileged the “core” or “canonical” writers and sources. In most fields, scholarship was considered an orderly and necessarily cumulative enterprise in which new inquiries both relied and built upon the earlier studies that composed the scholarly record. The bulk of research and learning was confined within rigid disciplinary boundaries, with each field claiming its own foundational literature and a unique suite of closely aligned methodologies.

Today's expectations are profoundly different. Cross-disciplinary inquiry, participatory learning, an obsession with primary resources and original documentation in all formats, and hybrid methodologies are increasingly the norm. The record of scholarship, while still important, has in many fields become less central. Multimedia research products, scholarship that relies upon massive and remotely hosted data sets, and team-based inquiry are other features of this emerging panorama. The appropriate locus of support for these resources is not yet clear.

#### Technology

In the past, libraries tended to acquire and warehouse hard-copy materials as passive objects for students and scholars to ferret out and then interpret on their own. Digital resources, by contrast, are energized from the start. The catchphrase

“If it isn't on Google, it doesn't exist” only begins to capture this dynamism. For a simple example, keyword searches across JSTOR can lead researchers to sources that would have remained invisible in a context limited by the traditional apparatus of field-specific bibliographies, indexes, and abstracts. On a more mechanical but likewise transformative level, linked footnotes allow the seamless pursuit of citation threads that would be unrealistic to track through physically dispersed books and journals. Mash-ups and other digitally recombinant possibilities encourage projects that transcend an exclusively textual framework. Libraries are now called to help users by contextualizing all of these energized resources within a broadly activated system of information, tools, and expert staff. Integrating our deep stores of analog holdings into this high-energy electronic network remains a central challenge.

Digital information predominates within some fields and is increasingly prevalent across the board. Digital technologies also are affecting scholarly inquiry and output as well as teaching and learning. Large data sets—numerical data, text corpora, image banks, etc.—invite structured inquiries across masses of information on a scale that can easily exceed the capabilities of any single institution. Tools for analysis, manipulation, and visualization may best be developed as community efforts. The “cloud” is becoming the locus for more and more data and applications in contrast to past models, which had readily identified developers and sites. “Community engagement” is likewise the byword for social networking initiatives, whose shorthand is the panoply of Web 2.0 products and services. Many new approaches to teaching and learning similarly rely upon open-source collaborative and participatory tools.

We do not yet understand the scholarly significance of large swaths of the digital universe. Blogs are often compared to diaries; e-mails are likened to letters and memos. The analogies are not only imperfect, but they may also complicate our decisions about what we need to capture and preserve. More familiar products like learning objects and computer software can be difficult to assess. Websites are typically dynamic and multilayered, requiring thoughtful protocols to determine what to retain. Social networking spaces are again unfamiliar. Instant Messages and cell phone videos pose challenges of their own. Scholars, users, creators, technologists, librarians, and digital objects themselves all have roles to play in clarifying our possibilities and needs.

One critical, perturbing, and unresolved element within the electronic universe concerns the requisites for preservation, which are most effectively addressed at the moment of digital conception. Today's technologies for access control—digital rights management, streaming systems, legal and contractual limitations, and so on—often work at cross purposes to permanence. Consensus-based regimes to ensure digital persistence are far from certain.

### Organizations and Institutions

The academic world is moving beyond structures defined primarily by discipline. Newly minted centers, institutes, programs, and initiatives today provide homes for interdisciplinary scholarship even as traditional departments remain strong.

The scholarly community also is affected by other kinds of structures and constraints. Intellectual property regimes channel access to and uses of many information resources. Google is an archetype for the commercial players that now occupy an expanding and disquieting space within the realms of information and academia. Traditional higher education is itself struggling against intense financial pressures, with for-profit institutions promoting an essentially distinct vocational model.

Cooperative arrangements and consortia are further reshaping the institutional environment. Economies of scale, aggregated expertise, new synergies and unexpected opportunities, and strengthened political coalitions and operational capacities are among the potential benefits. Local autonomy is less possible or desirable than ever—even as institutional competition remains a hallmark of American higher education.

### Modeling Information, Collections, and Content

Academic institutions create and also consume information. Libraries play a critical role within this ecology as they ensure the community's continuing access to the information resources that sustain research and learning. Conceptual frameworks, as well as practical tools, enable libraries to understand and then manage the torrents of information that now overflow the landscape. The following heuristic model has been created to help clarify our options. This model asserts that information resources in all forms and formats, whether viewed individually or in broader groupings, can be clumped into four ideal categories that reflect their academic uses as well as their origins: core resources and curricular support, the record of scholarship, primary resources, and data. Libraries, along with scholarly disciplines, departments and programs, and individual students and scholars, play critical roles in enacting this classification.

#### Core Resources and Curricular Support

All academic libraries provide the basic bibliographies and reference works, reading list materials, foundational literatures, and other core sources that are required for teaching and learning. Curricular support is a fundamental activity for every college and university library. Local definitions of each field's core resources also tend to carry across from one institution to the next.

### The Record of Scholarship

Academic libraries in institutions that support original research and advanced study further aspire to capture some or all of the record of scholarship. New studies in many fields are framed within a broader context of ongoing inquiry, as manifest in the scholarly record. Holdings that recapitulate this record thus remain critical in sustaining the cumulative process of creating new knowledge. This category includes the published outputs of colleges and universities, think tanks and scholarly societies, commercial laboratories and trade organizations, academies and associations, specialized agencies, and ad hoc research groups. A particular library's collecting appetite may vary within this large realm—perhaps only American university press publications; a multinational, multilingual sampler; or (at least in theory) exhaustive coverage. Levels of coverage can also vary among fields. Electronic publishing and new access technologies may mitigate the need for comprehensive local collections of the scholarly record by providing alternate ways to locate and use these resources.

The record of scholarship manifests itself above all in books and journals. Scholarly journals, which are important in all fields, make up the primary vehicle for validating new findings in science and technology. Market dynamics are pushing serials toward digital formats, through which they can also be disseminated at multiple levels of aggregation (bundled journal packages, individual serial titles, and specific articles)—always within a context of escalating costs. Scholarly monographs are then particularly central to the humanities. Despite experiments in electronic publishing and forecasts of ubiquitous print on demand, these materials are currently at risk.

#### Primary Resources

This immense third category comprises all organized human expression, or the full range of primary sources. These raw materials for scholarly work have become ever more eclectic. Many libraries have always pursued a broad range of noncanonical creative writing—novels, drama, poetry, and so on. Local and international newspapers, as well as government documents, are enduring mainstays as well. The scholarly record and synthetic works themselves serve as primary sources for researchers studying intellectual history and broader shifts in ways of thought. Rare book holdings and many special collections fall within this category as well.

Other primary sources have only more recently gained a place within the library (and scholarly) pantheon. Ephemera and grey literature, pamphlets, popular magazines, comic books, visual imagery, films and video, manuscript and archival collections, and sound recordings are all by now

accepted as legitimate collections categories. Websites, blogs, and other digital outlets, sometimes created as social endeavors and sometimes to represent a single perspective, are more recent additions. On a global scale, the gradually diminishing digital divide still affects information production and collections strategies across areas that differ in terms of affluence or openness.

### Data

Unorganized or minimally structured raw data represent a category of information that we are only beginning to understand. Scholars' unprocessed laboratory notes and research transcripts—unruly file cabinets, boxes of scribbles and scrawls—provide a quaintly venerable and readily managed example. The realm of raw data has assumed greater importance as the research tools associated with “big science” drive more and more scholarship. Digital satellite imagery, DNA and genome sequences, remote sensing data, raw survey responses, meteorological measurements, and text and image corpora are among the data sets and data streams that now pose daunting challenges of capture, interpretation, and curation.

Each field's scholarship and teaching draw upon different blends of information from these four categories. Research in medieval studies, for example, relies upon an array of original sources and texts that is by now pretty much fixed, at least when compared to the endless tidal wave of new materials that inform scholarship in fields like film studies, chemistry, and political science. For medievalists, exhaustive access to contemporary scholarship is therefore essential. Even this ground, of course, is not entirely solid—for instance, the field's research has broadened beyond a fairly confined textual canon to include the evidence of archeology and material culture. The high energy physics community, by contrast, relies heavily on the almost comprehensive availability of research findings in the arXiv (the archive for electronic preprints of scientific papers hosted by Cornell University) server. Peer-reviewed journals then invest specific reports with validation and prestige. Vast streams of raw data, for instance those generated by CERN's Large Hadron Collider, are crucial as well. Benchmark monographs serve to recapitulate the field's state of the art at particular points in time.

The academy's lore depicts the library as the humanist's laboratory, implying that historical materials and primary sources are less central in other scholarly realms. We need a more nuanced understanding. For example, historic field surveys are indispensable for botanical and zoological research. Star maps and celestial observations from both past and present are essential for astronomers. Scholars' uses of noncurrent literature in disciplines like chemistry or physics may follow in the path of medical researchers, who

have fruitfully engaged in text mining across large sets of historic data and reports.

The academic uses of information resources are shifting, sometimes in unexpected ways, across all four categories. While discipline-specific research paradigms often remain important, more agile models for scholarship and inquiry also suggest more fluid approaches. Today's pedagogical models routinely require students to grapple with primary sources and special collections, as well as secondary works and synthetic texts. Just as individual resources have become energized in the current environment, so has the entire structure of information. Our traditional collecting expectations, which were far more static and staid, no longer serve us well.

### The Changing Contexts and Expanding Scale of Collection Development

Models for information resources provide one potentially useful window into library collections and collecting. Another perspective focuses on the changing context within which our collections are now being built.

#### Collections of Record, Collections for Use

With a few exceptions, such as consciously duplicated core materials, reserve readings, and high-use recreational works, research libraries have sought to build collections that will persist through time. Carefully selected individual items, in their aggregate, make up definitive representations of the associated topics and fields. Libraries then care for these assemblages so that they will be permanently available. Creating and stewarding this patrimony constitutes a vocation of broad cultural consequence.

Looking to the future, research libraries will in some areas continue to build enduring collections of record. In others, they will settle for use-driven holdings while seeking neither comprehensive coverage nor long-term retention. The availability of digital surrogates or of remotely maintained archival copies may also affect local choices. Ideally, libraries will seek to ensure that some institution is providing ongoing preservation and care for everything they hold—but there may be instances in which current-use materials are acquired and discarded regardless of provisions for persistence. The continuum of curation will become more diverse.

#### From Collections to Collections and Content

Most academic libraries will continue to acquire both analog and digital materials for their on-site collections. However, their focus will expand ever more emphatically beyond acquisitions as they also provide access to intellectual content that

is leased rather than acquired, or to which they only point. Some libraries will likewise continue to create new, primarily digital, resources on their own. The increasing ubiquity and utility of highly diverse digital resources will require adjustments in all library operations. “Content”—a category that encompasses everything to which a library enjoys ready physical or digital access regardless of ownership status—is central to all that we do.

The diffuse knowledge that is embedded within and suffused throughout every university is a form of local content that most institutions have barely begun to tap. Energizing and leveraging this largely latent capacity is critical to the academy’s future. The process will most fruitfully engage faculty, staff members, and also the students whose research pilgrimages—mental and physical—foreshadow tomorrow’s scholarly agendas. Knowledge management will be a necessary element in our emerging content strategy.

#### Enlarging the Field: Partners and Players

All academic libraries are under intense financial pressure. The possibilities and the shared challenges associated with digital resources, the scale of today’s information needs, and examples of consortial achievements together make cooperation more appealing than ever before. The production of information resources, as well as conjoined consumption and processing, can become shared functions within a virtual environment.

Collective action may allow libraries to more fully shape both the landscape and the marketplace for electronic resources. Collections cooperation has traditionally emphasized the obscure, low-demand, sometimes expensive resources that can be shared between partners with minimal inconvenience to occasional local users. The compelling argument holds that shared physical resources made available through interlibrary loan can effectively reduce the need for redundant acquisitions at many different sites. Collecting scale, geographic and programmatic proximities, and resonances with other cultural institutions further shape the potential results. Structured acquisitions programs and streamlined processes for resource sharing have allowed limited progress in some relatively specific collections niches.

This model might now also be turned on its head as members of consortia together identify and arrange for digital access to core materials. Particularly in the electronic age, cooperative activities can cut across all four categories of collection resources. For large data sets, collaboration will be essential in building both infrastructure and tools because of the sheer scale of the task.

Collaborative action might encompass other dimensions as well. Libraries, archives, and museums are often co-located. They also share similar aspirations and missions. New opportunities for service and deeper complementarity

may be at hand. Research library cooperation has been most successful in focused efforts between groups of limited size, for example, intensive partnerships between two or three peer institutions, and relatively compact consortia such as the Committee on Institutional Cooperation or the California Digital Library. Cooperative initiatives that achieve enduring operational success seem to be bound by intractable limitations of organizational structure and scale, even in today’s technological age.

#### Libraries as Storehouses, Libraries as Tool Sheds

The mass of information resources now available on the Web, many of them free, is fundamentally changing the library community’s thinking about collections. High-quality and openly accessible scholarly resources—digitized maps and medieval manuscripts, books and journals, images from archives and art museums, music scores and sound recordings, and so on—can be found in staggering profusion without even considering the medium’s less scholarly emanations. Links to freely available digital content, metasearch capabilities that cut across products and platforms, and local aggregations of electronic resources, will all play a growing role in libraries’ collections and content strategies. This in turn will also reduce the physicality of library holdings and alter the functionalities of their spaces. But we need to go further.

Three aspects of Web-based content require close attention. First, the search engines that today allow users to find materials on the Web are neither transparent nor fully revealing of useful content in predictable ways. Google Scholar, for example, relies upon opaque search algorithms and relevance rankings that appear not to fully exploit the wealth of standards-based metadata that libraries routinely provide. But most libraries do little better, investing their cataloged resources with robust metadata that our discovery tools rarely handle well. Second, sources on the Web—whether websites themselves or the data, images, objects, and documents embedded within them—are notoriously unstable. Content is added, changed, and removed; links shift around and disappear. Scholarship relies on enduring access to constant content, a goal that remains elusive in the digital domain. Capture, curation, and digital preservation are all implicated in this conundrum. Third, dispersed and disparate Web content requires tools that can work across amalgamated sets of sources in predictable and repeatable ways. Some of the uses are well understood while others reflect a new realm of inquiry that includes text mining, pattern recognition, visualization, and simulation. The needs are perhaps most pressing around massive accumulations of raw data.

Libraries, working together and also with academics and information technologists, have an evolving role in creating

and supporting the tools that will enable students and scholars to take full advantage of the digital world. It is not yet clear whether lead roles can or should be preordained; arrangements that embody flexibility and contingency seem most likely to succeed.

#### Scarcity: Measure of Prestige or Consequence of Manipulation

Research libraries have traditionally built their reputations on the basis of their collection size and also the depth and breadth of their rare book holdings and their special collections. Scarce or unique artifacts, as well as uniquely comprehensive collections, remain primary measures of quality. Prestige based on both size and scarcity may diminish as large-scale digitization weakens the once obvious benefits of local ownership. The structural scarcity associated with rare artifacts is less compelling in a rich digital environment.

Paradoxically, our most coveted resources now include those digital materials whose uses are limited by contractual restrictions. Electronic gatekeepers can create scarcity (and also compromise long-term persistence) by manipulating license agreements and relying upon restrictive delivery technologies, even as the underlying resources could in theory be available without limit. “Scarcity,” in a traditional sense, reflects materials that are physically rare or unique. Today’s environment adds in the artificial scarcity created through restrictive manipulations of the digital marketplace.

#### Authorship and Authority

Academic libraries have historically served as custodians for carefully selected, authoritative information. Library holdings were then taken to embody the highest standards of analytical and methodological rigor. The weighty bound tomes associated with research libraries and traditional scholarship carried their own aura of permanence and security. Norms for careful reading and measured scholarly discourse further suggested prudence, stability, confidence, and authority. Deeply embedded synergies between artifact and text played an essential role in research and teaching.

Our excursions into the exuberantly expressive realm of primary resources have effectively destroyed these presumptions. All manner of deliberately ephemeral products circulate at high velocity, undermining anyone’s attempts to delimit agency, define a “canon,” or codify quality. Ours is instead a prolific universe of spontaneous, unmediated, non-validated information. Web 2.0 both reflects and engenders “Authority 2.0” as users, singly or in cohorts, participate in an electronic free-for-all. Platforms and formats are likewise provisional. Experimental and ephemeral expressions may evolve into dominant manifestations and forms, though extinction (think WordStar in the pedestrian realm of word

processing programs) is a real possibility as well. Libraries are on uncertain ground as they engage with this fractious, seductive, alien, and essential universe.

### Guiding Principles for Collections and Content

The trends here described suggest several general principles to guide academic libraries as they move toward the future:

1. Most information—core materials, the record of scholarship, trade publications, an increasing proportion of recorded human expression, and data—is becoming available in digital formats. The emergent electronic realm will, in time, relegate new analog materials to a diminishing subset of primary sources. Digital resources will increasingly define both the information and the scholarly landscapes. *Our future is digital: libraries must prepare for and promote this shift.*
2. Digital resources are produced, become available, and then behave differently than hardcopy objects. Among many other features, few of them can be owned in the same way as books or journals. Libraries must therefore frame their information goals in terms of providing access to content that they do not possess, as well as on-site holdings. *Libraries must broaden their focus to encompass both collections and an evolving range of content, whether owned or not.*
3. As budgets decline and priorities shift, many academic libraries will steer their acquisitions toward the basic texts and sources required for curricular support. These holdings will be heavily redundant across different institutions. Conversely, more and more noncore materials may be entirely missed. Cooperative efforts—international, national, regional, and local—can at once increase efficiencies around everyone’s need for duplicative materials and also maximize the collections reach of those libraries that are capable of pursuing scarce or unique resources. *Cooperative activities will become increasingly central to library programs and strategies.*
4. The commercialization of scholarly information, on top of long-standing trends toward monetization and privatization in the realms of mass expression and entertainment, threaten the free flow of information that the academy requires. Prohibitive costs and artificial scarcity are among the consequences. Many experiments and initiatives, with those broadly clustered under the open-access

rubric among the most promising, are now in play. *Academic libraries must actively engage in reformulating information flows and scholarly communications to protect future research and learning.*

5. Libraries have always sought to make information both accessible and usable. Catalog records link users with the sources relevant to their interests; reference (or “research and learning”) services then help those users extract the fullest possible benefit from what they have found. Digital resources—particularly large-scale, cloud-based data—require new, standards-based tools and services for description, access, use and manipulation, and preservation. *Libraries, acting independently and through external partnerships, must participate in developing all of these tools and services.*
6. Academic libraries must be aligned with and accountable to their parent institutions. Yet information is becoming more diffuse and library activities, across the board, are ever more cooperative in nature and expansive in scope. Closely consultative

processes within each campus will remain essential, but may no longer be sufficient. *Universities and libraries must devise models for governance that both ensure local accountability and encourage cooperative activities.*

The world of library collections is one in which once solid certainties no longer obtain. The range of relevant materials has shifted and grown, though the relative centrality of tangible resources under the library’s direct control is in decline. Libraries will increasingly work to identify and describe information that they will never own, and to provide the tools that enable their students and scholars to discover and use these resources effectively. The sources themselves will take on new dimensions whose continued usability will demand different kinds of support. More and more, libraries will have to engage in partnerships and collaborative efforts to achieve their goals. While the mandate to ensure ready access to a comprehensive array of information resources will remain, the “what” and “how” will seem quite different.

---

### Index to Advertisers

ALCTS.....	124
Archival Products.....	114
Library of Congress.....	cover 2
Library Technologies.....	cover 3

# Book Reviews

Edward Swanson

*Usage Statistics of E-Serials.* Ed. David C. Fowler. Binghamton, N.Y.: Haworth, 2007. 297p. \$40.00 soft cover (ISBN 978-0-7890-2988-1), \$70.00 hard cover (ISBN 978-0-7890-2987-4). Published simultaneously as *The Serials Librarian* 53, suppl. no. 9, 2007.

The drawback of many of Haworth's monographs that are copublished as journal issues is that while they are a collection of decent journal articles on a certain topic, they lack the cohesive sense that is characteristic of a monograph. *Usage Statistics of E-Serials* is a collection of "articles" authored by more than twenty-five people on a wide range of practical and theoretical topics related to usage statistics; read as a whole, the chapters are highly redundant in certain aspects, making reading the book in its entirety a monotonous endeavor. Almost every one of the seventeen chapters touches on the importance of usage statistics, how statistics can be used in decision making, how they help us understand users, the inconsistency of usage statistics, and definitions of COUNTER (Counting Online Usage of Networked Electronic Resources) and ICOLC (International Coalition of Library Consortia). Understandably, all these topics are relevant to a discussion of usage statistics, but the book would be a much more useful monograph if it began with a general introductory chapter outlining these basics followed by chapters delving into the viewpoints and experiences of each of the chapter authors.

Drawbacks aside, the book has many chapters that will be of interest to librarians (or students studying library and information science) who are new to the benefits and processes

(and problems) of collecting and analyzing the usage data that come from vendors and publishers. Eight of the chapters attend to the basics of how to do this. Chapter 1 addresses "practical considerations" related to the "processing, standardizing, and dissemination" of statistics (6). The chapter titled "Application of Electronic Serial Usage Statistics in a National Laboratory" nicely addresses how to evaluate a resource by analyzing usage statistics, and how to put that resource in context. Several other chapters concentrate on the methods of calculating and analyzing cost-per-use (or cost-per-download) and issues related to using the information in decision-making. Norm Medeiros, in "Uses of Necessity or Uses of Convenience? What Usage Statistics Reveal and Conceal About Electronic Serials," discusses the history of usage statistics and why we track use, but he also points out the pitfalls of "error-laden data" pertaining to cost-per-use information (237). Several authors stress the importance of using usage data to assist in formulating informed decisions on collections but caution against putting too much emphasis on usage data when making decisions to cancel resources. Rickey Best, author of the chapter "Lies, Damn Lies, and Usage Statistics: What's a Librarian to Do?" concludes that "understanding the limitations of usage data allows librarians the opportunity to integrate the data with collections policies, and to focus upon the needs of the user community" (212).

The remaining nine chapters delve into a wide variety of topics. Two of them address specifics related to products available to assist with collecting and managing usage statistics

(chapters 7 and 17); however, this book was published more than two years ago. Many more products are now available to libraries to assist them in managing electronic resources and usage. Of particular interest is the chapter "Shared Purchase-Shared Responsibility: A Stewardship Tool for Consistent E-Usage Evaluation," in which Susanne Clement shares her experience developing a model for managing centrally funded databases at the University of Kansas. The model involves assigning each centrally funded database a "steward" who annually assesses the scope, coverage, and usage statistics of the resource and investigates competing products. Chapters such as Heather Morrison's on the impact of usage statistics as an economic factor in scholarly communication, Eleonara Dubicki's on how statistics can help drive marketing resources to patrons, Susan Kendall and Celia Burke's on gathering usage statistics for government resources, and Carol Tenopir's on the MaxData project, which "seeks to determine how to engineer a system that adequately measures the effectiveness" of resources, demonstrate the broad range of issues related to usage statistics included in this book (73).

Elise Anderson's chapter, "The Next Steps in Developing Usage Statistics for E-Serials," is one of the best in the book. It touches on the basics of usage data (such as accessing, retrieving, and storing data) while looking to the future to assess ways to "develop and improve the utility of usage statistics" (245). Anderson offers ideas to solve a situation that is "ripe for improvement" (249) and discusses ways that bibliographers, collection managers, and library instructors

could collaborate to increase the usage of existing resources.

Without a doubt, *Usage Statistics of E-Serials* imparts the idea that collecting, assessing, and storing usage data involves time, personnel, and money. The book also stresses that usage statistics can inform much more than just collection management decisions. Used creatively, statistics can help librarians increase users' awareness of their existing resources, learn more about user needs and activities, and justify their budgets. Although this book was copublished as a *Serials Librarian* supplement, the chapters are not indexed anywhere that I could find, nor are they available electronically. What a shame. Most readers will find a portion of the book useful to them, but not the book as a whole.—*Karen Fischer, (karen-fischer@uiowa.edu), University of Iowa, Iowa City.*

**Newspapers Collection Management: Printed and Digital Challenges. *La gestion de colecciones de periodicos: desafios en impresos y digitales. Proceedings of the International Newspaper Conference, Santiago de Chile, April 3–5, 2007.*** Ed. Hartmut Walravens. Munich: K. G. Saur, 2008. 396p. \$152.00 (IFLA members \$114.00) (ISBN 978-3-598-22039-5). IFLA Publications, 133.

*Newspapers Collection Management: Printed and Digital Challenges* continues the tradition set by Hartmut Walravens, chair of the International Federation of Library Associations and Institutions (IFLA) Newspaper Section, of publishing the proceedings of the various IFLA international newspaper conferences and section meetings. The focus of this book is on the meetings in Santiago, Chile, in May 2007 and at the section meetings at the IFLA World Library and Information Congress (WLIC) in Durban, South Africa, in August 2007.

The first part of the book centers

on conference papers presented in Santiago, Chile, which had a special focus on Latin American newspaper collections and looked at activities, needs, and various issues with which those libraries and countries have had to deal with as they take on more digitalization projects. The majority of the articles are presented in both English and Spanish, but a few articles are only in Spanish. These Spanish-only articles come from Mexico and Peru, and include a slideshow presentation from Thomson Gale. One article solely in English discusses the preservation and digitalization of Latin American newspapers in the United States. Other articles in Spanish and English come from authors in Chile, Costa Rica, Cuba, Colombia, Guatemala, as well as from Finland, the United Kingdom, Canada, the Netherlands, the United States, France, Germany, and Spain. Topics cover a wide range of digitalization issues, but a large percentage of the articles focus on the history of printing in a particular country, the history and evolution of various digitalization projects, and projects that are currently underway. Chapters covering unique topics include “The Evolution of the Aboriginal Presses in Canada” by Sandra Burrows, “Colombian Newspapers of the 19th Century: Treasures and Memory” by Sandra M. Angulo Mendez, and “Cuban Newspapers in the XVIII and XIX Centuries: Conservation of Unique Existing Issues” by Hilda Perez Sousa. What these and other articles clearly illustrate is the passion, collaboration, time, and effort that librarians and libraries are putting forth in their efforts to save the past and make it easily available for their users.

The second section of the book includes the four papers that were given at WLIC in Durban, South Africa. Three articles cover newspaper collections at the National Library of South Africa, the Library of Congress Office in Nairobi, Kenya, and the Makerere

University Library in Uganda; the fourth looks at the future of African newspaper collections in American educational institution (which sadly shows that newspapers from a few countries from Sub-Saharan Africa are not being collected). These four articles also are presented in English and Spanish. Newspaper collections in Africa also deal with many of the same issues, such as microfilming, cost, new technology, indexing, and access, found in South America.

The book does a great job presenting new and interesting research, but readers might be disappointed with the lack of citations and documentation in many of the articles. Also, because these are conference papers, grammar, spelling, and style formats are not uniform.

This book is a fascinating account of what libraries around the world are doing with their newspaper collections and the many issues with which they deal in managing their digital, print, and microfilm collections. Ximena Cruzat Amunategui, director of the National Library of Chile, writes in her article that “digital libraries share many things with the ‘real’ (traditional, physical) libraries that we know and love: 1. It is focused on readers and patrons. Without readers, there can be no library. 2. It is more than a mere repository. It is a space for encounters and knowledge creation. The road is full of new challenges and discoveries, but the road itself is older than Alexandria” (11). While reading *Newspapers Collection Management: Printed and Digital Challenges*, one can easily see the new challenges and discoveries that await those on the digital road. This book is highly recommended for those libraries with newspaper collections, as well as those universities with archival, library science, and journalism programs.—*Melissa Aho, (ahox0017@umn.edu), University of Minnesota, Minneapolis.*



pass it on...

# Preservation Week

May 9-15, 2010

With spring cleaning underway, Americans across the country will wonder how to preserve new-found treasures.

From May 9 – 15 communities will turn to their local library to celebrate the first-ever Preservation Week @ your library, a time when libraries across the country will provide information and expertise on how to preserve collectables, photos, family records and other valuable materials.

During Preservation Week @ your library, themed “Pass It On: Saving Heritage and Memories,” take this opportunity to offer special programs and services to connect your users with preservation tools for their own collections; promote the importance of preservation; and enhance everyone’s knowledge of preservation issues.

Library professionals can find more information on Preservation Week @ your library, including programming ideas and tools, at [www.ala.org/preservationweek](http://www.ala.org/preservationweek).

Direct patrons and the general public to the @ your library web site, [www.atyourlibrary.org](http://www.atyourlibrary.org), for ideas, tips, and stories about saving their treasures.

# *Authority Express*

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

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

## *Authority Express*

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

*"Authority Control for the 21st Century"*



## **LIBRARY TECHNOLOGIES, INC.**

2300 Computer Avenue, Suite D-19 Willow Grove, PA 19090

(215) 830-9320 Fax: (215) 830-9422

(800) 795-9504 email: [LTI@LibraryTech.Com](mailto:LTI@LibraryTech.Com)

Visit our website at:  
[www.LibraryTech.Com](http://www.LibraryTech.Com)



# ALCTS

Association for Library Collections & Technical Services

a division of the American Library Association

50 East Huron Street, Chicago, IL 60611 • 312-280-5038

Fax: 312-280-5033 • Toll Free: 1-800-545-2433 • [www.ala.org/alcts](http://www.ala.org/alcts)