

## EDITORIAL BOARD

Editor and Chair

Peggy Johnson

*University of Minnesota*

### Members

Kristin Antelman, *North Carolina State University*

Elise Calvi, *Indiana Historical Society*

Allyson Carlyle, *University of Washington*

Elisa Coghlan, *University of Washington*

Leslie Czechowski, *University of Pittsburgh Health Sciences Library*

Lewis Brian Day, *Harvard University*

Dawn Hale, *Johns Hopkins University*

October Ivins, *Ivins eContent Solutions*

Edgar Jones, *National University*

Birdie MacLennan, *University of Vermont Libraries*

Randy Roeder, *University of Iowa*

Carlen Ruschoff, *University of Maryland*

Sarah Simpson, *Tulsa City-County Library System*

Lori Terrill, *University of Wyoming Libraries*

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

Lynn N. Wiley, *University of Illinois at Urbana-Champaign*

### Ex-Officio Members

Charles Wilt, Executive Director, ALCTS

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

Norm Medeiros, *Haverford College*  
Book Review Editor, *LRTS*

# Editorial

Peggy Johnson



I want to draw your attention to an important paper, “Assessing the Cost and Value of Bibliographic Control,” by Erin Stalberg and Christopher Cronin, in this issue. *Library Resources and Technical Services (LRTS)* does not normally publish the work of Association for Library Collections and Technical Services (ALCTS) committees, interest group, and task forces. As a peer-reviewed journal, *LRTS* publishes papers that take a critical approach to the questions and challenges facing librarians and libraries and publishes both research papers and thoughtful explorations of operational issues that have value and implications for other libraries. “Assessing the Cost and Value of Bibliographic Control,” while incorporating a great deal of the final report of the Task Force on Cost/Value Assessment of Bibliographic Control (appointed by the ALCTS Heads of Technical Services in Large Research Libraries Interest Group), is much more than a task force report or a summary of the task force’s work. It met *LRTS*’ rigorous requirements and was vetted through our double blind review process. “Assessing the Cost and Value of Bibliographic Control” is a thoughtful analysis of a complex topic and should be required reading for all who care about the future of bibliographic control and technical services.

Stalberg and Cronin’s paper takes a critical approach to the questions and challenges facing technical services librarians today. How do we measure costs, benefits, and value of bibliographic control? How can we document the value of investment in bibliographic control, which is a central function of technical services operations? How do we communicate this value to library administrators? The authors place these questions in a historical context through their literature review, noting that answering the key questions about cost/benefit have been hampered by the absence of definitions of value as well as methods for assessing both cost and value.

The Task Force on Cost/Value Assessment of Bibliographic Control initially set out to develop and articulate metrics for evaluating the cost and value of cataloging activities, but it realized that a common vocabulary for what constitutes value and an understanding of how value is attained are needed first. At the heart of Stalberg and Cronin’s paper are seven operational definitions of value:

- Discovery success
- Use
- Display understanding
- Ability of library bibliographic data to operate on the open web and inter-operate with vendors and suppliers in the bibliographic supply chain
- Ability to support the Functional Requirements of Bibliographic Records (FRBR) user tasks
- Throughput and timeliness
- Ability to support the library’s administrative and management goals

What strikes me in re-reading these definitions is how useful they are in discussing with administrators what makes cataloging both effective and successful—and how they describe what is too often lacking in many of the vendor-supplied records for e-books.

Stahlberg and Cronin, echoing the task force, stress the need for additional research into which bibliographic elements have the greatest value to, and impact on, users. The authors propose numerous areas for research that will increase understanding of the costs and values associated with bibliographic control. Some of these are feasible as individual research projects, some will require institutional

resources and support, and some are possible only with grant-funded projects. I hope you will read this paper with care and heed the call to advance understanding of the value that bibliographic control brings to the library and its mission to support discovery, access, and, ultimately, research and scholarship.

This issue contains book reviews commissioned by Norm Medeiros, whom I am delighted to welcome back to the *LRTS* editorial board in his new role. If you are interested in writing book reviews for *LRTS*, please contact him ([lrtsbookreview@ala.org](mailto:lrtsbookreview@ala.org)).

# Assessing the Cost and Value of Bibliographic Control

Erin Stalberg and Christopher Cronin

*In June 2009, the Association for Library Collections and Technical Services Heads of Technical Services in Large Research Libraries Interest Group established the Task Force on Cost/Value Assessment of Bibliographic Control to address recommendation 5.1.1.1 of On the Record: Report of the Library of Congress Working Group on the Future of Bibliographic Control, which focused on developing measures for costs, benefits, and value of bibliographic control. This paper outlines results of that task force's efforts to develop and articulate metrics for evaluating the cost and value of cataloging activities specifically, and offers some next steps that the community could take to further the profession's collective understanding of the costs and values associated with bibliographic control.*

The technical services community has long struggled with making sound, evidence-based decisions about bibliographic control. This has been demonstrated recently by controversy over the 2006 Library of Congress (LC) decision to change its practices for series authority control, concern over the impending implementation of Resource Description and Access (RDA), the increasing need to better integrate library bibliographic data with nonlibrary web data, and requests from library administrators to document the value of investment in cataloging operations. The ability to make evidence-based decisions has been hindered by a lack of both operational definitions of value and methods for assessing cost and value within larger institutional constructs. To date, libraries have not developed robust cost/benefit metrics, and those for bibliographic control are even further lacking. The development of cost/benefit analyses for libraries may be difficult, but faced with limited resources and an array of directions in which to move forward, libraries find that articulating the varied cost/value propositions in measured and concrete ways is increasingly necessary.

In June 2009, the Heads of Technical Services in Large Research Libraries Interest Group of the Association of Library Collections and Technical Services (ALCTS) sponsored the Task Force on Cost/Value Assessment of Bibliographic Control (hereafter referred to "the Task Force") to begin to identify measures of cost, benefit, and value of bibliographic control. This paper offers a literature review, outlines the work of that Task Force, explores operational definitions of value associated with bibliographic control, suggests research areas that will further the profession's understanding of the value of cataloging activities, discusses possible cost measures, and considers interdependencies between creators and consumers of bibliographic data.

**Erin Stalberg** (serin\_stalberg@ncsu.edu) is Head, Metadata and Cataloging, North Carolina State University Libraries, Raleigh, and **Christopher Cronin** (croninc@uchicago.edu) is Director, Metadata and Cataloging Services, University of Chicago Library.

Submitted September 8, 2010; tentatively accepted, pending revision November 15, 2010; revised and resubmitted December 13, 2010; accepted December 24, 2011.

The authors wish to thank Ann-Marie Breaux, John Chapman, Karen Coyle, Myung-Ja Han, Jennifer O'Brien Roper, Steven Shadle, and Roberta Winjum, members of the Task Force on Cost/Value Assessment of Bibliographic Control, who along with the authors, participated in the work delivered to the Heads of Technical Services in Large Research Libraries Interest Group at the American Library Association Annual Conference, Washington, D.C., June 25, 2010, on which this paper is based.

## Literature Review

The literature gives evidence of a lengthy dialogue about the cost and value of cataloging, often tied to a discussion about the impact of advancing technology. Of interest is how similar that dialogue has been over time. In an address to the New York State Library School in 1915 titled *Cataloging as an Asset*, Bishop asked his audience “of what value is a knowledge of cataloging?”<sup>1</sup> Only fifteen years after the LC had begun distributing cards, Bishop—who at the time was the superintendent of the LC’s reading room of the Library of Congress—remarked,

Seventy-five per cent of the cards needed in the various libraries of the country are being supplied by the Library of Congress. It is not unnatural, in fact it is almost inevitable, that there should have come a lessening of interest in cataloging work, and even a dearth of catalogers. . . . The successful adaptation of a manufactured product is seldom as interesting as the making itself. . . . Catalogs and catalogers are not in the forefront of library thought. In fact a certain impatience with them and their wares is to be detected in many quarters. Shallow folk are inclined to belittle the whole cataloging business. And there have not been wanting persons to sit in the seat of the scornful.<sup>2</sup>

The tension between the increasing availability of what we would now call “copy” and the resulting value of employing professional catalogers was clearly palpable almost a century ago. The crux of Bishop’s argument is that cataloging forms the core of the profession because the catalog itself is a valuable and essential instrument for the reader to do his or her work. Recognizing that libraries and their indexes, shelf lists, and public catalogs were growing exponentially, Bishop was concerned about the implications for cataloging, the catalog, and the values placed on them. He wrote,

We have continued to use an instrument whose value for small collections is well established, and we have built it up until it fairly threatens to break down of its own size and weight. . . . But we have not seemed to realize that all our skill and all our abilities are now needed to make our huge card catalogs workable. We shall need every bit of energy, vigor, and knowledge that we possess to adapt the card catalog to libraries of the future.<sup>3</sup>

Even in 1915, Bishop was keenly aware of the shared network that would eventually develop for library cataloging and record keeping, the impact that an investment in network-level operations could have on the profession,

and the value of the rules developed to create cataloging. In thinking of the task of keeping track of books scattered across branches of public libraries, for example, he noted,

What a complicated thing is a modern “union shelf list,” a “combined catalog!” And how near we are to the day of union catalogs or “repertories” designed to show the resources of cities, or regions, perhaps of the entire country! Can you imagine anyone unversed in practical cataloging undertaking to supervise such records? . . . The future is a day of co-operation, and co-operation in most cases on the common basis of one set of cataloging rules governing a supply of contributed entries. You will begin to see something of the value of those rules.<sup>4</sup>

Cost measures, especially for technical services, have formed part of the library literature since at least the latter part of the nineteenth century.<sup>5</sup> In his historical review of discussions of cataloging costs, Harris highlighted Congress’s own complaints that it cost 22 cents per book to catalog the library that Thomas Jefferson sold to them after the original Library of Congress was burned by the British in 1812.<sup>6</sup> Harris surmised that this incident “yielded one of the earliest figures on cataloging in this country and probably the first recorded protest over the high cost of cataloging.”<sup>7</sup> By 1941, Metcalf wrote that surveys of cataloging costs had become “vogue,” but that they “accomplished little, except to make us understand that costs were high and that there seemed nothing to be done about it.”<sup>8</sup>

By the 1950s, librarians sought to not just list the types of costs associated with cataloging and formulas to calculate them, but also to contribute to the interpretation of those costs. Swank wrote that calculating the per-unit cost of cataloging is not enough on its own, one also must understand the evolving context of those unit costs, which will not remain static over time.<sup>9</sup> For instance, unit costs increase “as ever sharper distinctions must be drawn among ever larger quantities of materials. The ‘no conflict’ principle is increasingly more difficult to apply, and the definition of the relationships among books becomes more and more subtle.”<sup>10</sup> Reading this in 2010 is eerily timely, considering the proliferation of access to information afforded by the Internet and efforts like the Google Books Project, as well as the amount of current discussion on the costs of implementing RDA, which focuses largely on describing relationships between resources.<sup>11</sup>

Swank also touched on the then-half-century mark of LC card distribution and the fact that many libraries had still not realized the economic benefits of centralized copy production; they were not accepting copy as is, but were reviewing and altering cards at length. While libraries at the time argued that they were placing a greater value on the institution’s specialized needs over universal or shared

economies of scale (“Uniformity at the local level may seem to be more important than conformity at the national level”), Swank observed that entrenchment may be at the root of this value compromise.<sup>12</sup> He recognized the hidden costs of intangibles—such as morale, organizational culture, and the provision of adequate training—and encouraged moving beyond computing unit-costs alone to also analyzing and measuring the work itself. Swank was dismayed that, at the time, no studies had related costs to values or results in a way that could be used by other libraries; he was even more dismayed that no studies had yet evaluated the product of cataloging efforts against the needs of the readers.

One notable, early attempt to apply cost/benefit analysis across an entire library system was performed by MIT Libraries in 1969 and reported by Raffel and Shishko.<sup>13</sup> However, despite acknowledging that cataloging used 21.2 percent of MIT Library’s general and research collection budget at the time—3.4 percent higher than the purchasing budget itself—the authors dedicate only four pages of the book to analysis of cataloging operations. Reviews of the cost/benefit analysis methods applied to MIT were not positive across the profession, largely because the analysis was performed by an economist and a political scientist, not a librarian. McAnally observed that “the authors suffer from two severe handicaps—relative ignorance about the details of libraries, learning, and research, and also the absence of clear objectives and good measures of success or effectiveness in the university library world.”<sup>14</sup> Indeed, Raffel and Shishko themselves had recognized the limitations of their method, noting that “much of the analysis that has been presented so far has relied primarily upon impressionistic judgments of the benefits associated with a given system and of comparisons among systems serving different objectives.”<sup>15</sup> The study did not have broad impact either at MIT or elsewhere.

To Bishop, the rapidly increasing size of collections justified the investment in cataloging.<sup>16</sup> More than seventy years after his address to the New York State Library School in 1915, the importance of working at the network level was clearly understood on the philosophical level, but not-yet fully realized operationally. In the late 1980s, Lahiri wrote that the “proliferation of information has surfaced as a persistent prospect and problem to society. This proliferation further complicates the ways and means of bibliographic control and, more crucially, the justification of cost and benefit of providing access/exposure to the users.”<sup>17</sup> Lahiri noted that automation in the 1960s attempted but failed to solve problems associated with cataloging the growing corpus of information because those efforts were largely concentrated in research libraries

with little conceptual basis which could be used for a broader national context. Instead they empha-

sized that the unique or special aspects of cataloging must meet the needs of their own institutions first. Even the MARC format, despite its far-reaching impact in cataloging automation at the national and international levels, was not free from such proclivity.<sup>18</sup>

Earlier in 1981, Koel had expressed similar concern about the concentration of cataloging costs at the institutional level and promoted forming a centralized federal Agency of Bibliographic Control that would not only steward master records, but also perform cost/benefit analyses of existing practices that would ensure improved efficiencies of scale.<sup>19</sup> Lahiri explained that even though criticisms of the “slow and costly” nature of cataloging have not declined over time, there was still

not much or any calculation, clarification, explanation, and justification for the value of an authoritative catalog. An absence of the discussion on the benefits of bibliographic control is equally conspicuous. Some believe that in order to increase widespread demand for information, bibliographic systems will have to provide their worth in dollars. To estimate that value we should use economic criteria.<sup>20</sup>

Unfortunately, even with this explicit recognition of the lack of applicable measures, few are offered.

In advocating for the idea of a national database, Lahiri was acutely aware of the same challenges we recognize even today, stating that cost/benefit analysis of cataloging is complicated by the fact that users who access information do so outside of a typical marketplace economy.<sup>21</sup> Access, he wrote, is largely provided in a noneconomic environment, making difficult the justification of building new technologies and systems to support bibliographic data, the value of which is elusive. Lahiri’s own assessment of the library literature of the time was that the concepts of “benefit” and “effectiveness” were used in interchangeable and confusing ways, and that most experts at the time felt that library services as a whole could not be measured in monetary terms. In relation to the idea of building a national integrated bibliographic database, he asserted that while devising “ways for the measurement of the value of national bibliographic control activity” is critical, most existing studies are ambiguous, ill-defined, misleading, unreliable, and “clouded with contradictions.”<sup>22</sup> He stated that the extraneous benefits of improvements to “research, education, and social well-being” are so intangible in character “as not to be susceptible to appraisal in monetary terms.”<sup>23</sup>

Tangible measures for cost, value, and benefits of cataloging are rare in library literature. Lahiri posited that the effects of centralizing a national database could be measured

by the speed at which records could be reliably used by others, that benefits could be measured in terms of increases to scholarly output, but that ultimately, quantification would be nearly impossible. Even if the benefits could be quantified,

they cannot be valued by any market criteria and are generally termed as intangibles. . . . Although some quantitative assessment of benefits is possible, the multiplicity of benefits and their diffusion among different aspects of life will normally be such that their precise quantification is difficult to trace . . . [and would be] an attempt to measure the unmeasurable.<sup>24</sup>

In 1997, the Council on Library Resources commissioned a detailed study on the value of library and information services.<sup>25</sup> The study had two primary objectives: to analyze issues related to the value of library and information services in order to develop a conceptual structure that could serve as a theory of “use-oriented value of information and information services,” and to apply the theoretical framework to propose methods for similar studies of other information services generally.<sup>26</sup> The report, issued in two parts, discussed the difficulty and complexity of defining value. Despite generating a strong taxonomy of values, the study did not apply those value structures directly to monetary or other economic measures, concluding, “While studies and determination of value are a difficult and involved proposition, they are only the first step in meeting a larger challenge. The challenge is to connect studies of value with some appropriate economic indicators.”<sup>27</sup>

Missingham, and Imholz and Arns have reviewed and summarized studies that attempted to quantify, in monetary terms, the benefits of library services generally. However, they did not quantify bibliographic control specifically, except to show calculations of cost savings derived from using copy cataloging records extracted from national bibliographic databases.<sup>28</sup> A recent attempt at cost/benefit analysis within large academic research libraries is the University of Illinois at Urbana-Champaign’s attempt to correlate library costs with direct monetary benefits in the form of grant funds.<sup>29</sup> Using the results of a study on return on investment at one institution, during only one year, has obvious and acknowledged limitations in terms of wider applicability, but it does serve as an example of one approach that could be adjusted for other environments or institution-types. Again, the context of the Illinois study is the organization as a whole and the model put forth does not specifically address bibliographic control. Cornell Library’s more informal approach to demonstrating return on investment simply lists how the library is used and how it generates more value than the money expended to support its operations.<sup>30</sup> Again, however, they do not focus on bibliographic control and metadata provision.

Some studies have been careful to note that costs are not necessarily comparable across institutions when those institutions have variations in such factors as number and levels of staff, types of resources, levels of cataloging, and the number of records processed as a result. McCain and Shorten analyzed the results of a survey of academic libraries, which focused on staffing levels, the number of items processed, the presence and size of a backlog, the automation system in use, and perceptions of efficiency.<sup>31</sup> They presented measures of efficiency and effectiveness for cataloging departments on the basis of those factors. Morris and colleagues described a longitudinal study by Iowa State University measuring cataloging time and costs, as well as the tasks that staff performed.<sup>32</sup> Their article categorized the tasks (e.g., copy cataloging, original cataloging, authority control, recataloging, and monographic and serials cataloging) in detail, and analyzed the productivity of each task with staff. A subsequent study analyzed tasks librarywide.<sup>33</sup>

Measuring the relationships between copy cataloging, original cataloging, and partial-original cataloging, and levels of staffing has been a frequent focus of studies, perhaps because they are considered more tangible or well-understood categories of bibliographic control. Miksa recently reflected on her 2005–6 study of technical services operations in rural, urban, and suburban libraries in North Texas, which asked respondents to give the average number of hours per week dedicated to original or partial-original cataloging activities.<sup>34</sup> Of the 103 respondents, 8 libraries reported 0 hours, 59 reported fewer than 10 hours, and only 5 reported 31 to 40 hours. Miksa also described her own anecdotal experiences with her cataloging students and library staff across Texas, many of whom have expressed concerns about the diminishing value their organizations place on cataloging. She offered authority control as an example of a cost/value compromise made by many libraries in her survey, with only 12.5 percent of respondents reporting weekly or monthly maintenance of authorities databases. While the survey does not elicit reasons behind the time spent on maintaining authority databases, Miksa reported her own impression that perhaps there is a lack of understanding of the purpose and value of authority control or a belief that outsourced records are good enough or “it may simply be rooted in the more realistic lack of funding.”<sup>35</sup>

Miksa further posited that investment in cataloging is an investment in adding value to resource description and access. She stated that poor or “dead-end” metadata is a reflection of the lack of value placed on quality cataloging, as evidenced by the decreased emphasis on cataloging in graduate library programs. She wrote, “I strongly suspect that we are seeing in our catalogs the result of the disturbing lack of knowledge of many cataloging librarians and library administrators that resulted from relegating traditional courses to the back burner over the past decade or so.”<sup>36</sup>

Hider's application of the contingent valuation method to estimate the monetary value added to a collection by the technical services operations of an Australian city public library demonstrates a recent (Hider claims the first) attempt to place a dollar figure on the value of technical services.<sup>37</sup> Contingent valuation employs survey methods to establish value for resources and services that are non-market (i.e., not sold). For the study, Hider presented three scenarios to gauge the relationship between cost and the respondent's willingness to pay: a referendum was held to ask citizens to pay a monthly levy to maintain library services at present levels or the library would close the library was converted to "self-service," wherein the library would maintain the catalog and the collection as it exists today; and the self-service library consisted solely of the collection, with no catalog. The benefit/cost ratio for the first scenario was 1.33:1; the ratio for the second scenario was 1.8:1; the ratio for the third scenario—that is, for technical services specifically—was 2.4:1, demonstrating an especially good value provided by technical services.<sup>38</sup> While Hider's study is of a small city public library and does not focus on bibliographic control in an academic library environment, he articulated interesting methodological issues that could translate across any library size or type.

Gorman related the value of cataloging to the value of the resources cataloged. He posited that the two main problems with bibliographic control of electronic resources are that the majority are themselves "of no value, little value, very localized value, or temporary value," and they are "inherently unstable and shape-shifting."<sup>39</sup> He stated that cataloging resources that are valueless, of limited value, or that could be changed or altered in the future is neither rational nor efficient. Instead, the value of cataloging will only be realized if resources that are actively assessed to have value and permanence are cataloged.

RDA, the proposed successor to the Anglo-American cataloging rules, includes cost efficiency as one of its objectives: "The data should meet functional requirements for the support of user tasks in a cost-efficient manner."<sup>40</sup> Institutions participating in the U.S. national RDA test will, as a part of that process, contribute surveys for every bibliographic record created during the test, including details on how long it took to catalog a resource. The results of the test, and what might be learned from that process about the costs of implementing RDA or the values it will help realize, remain unknown.

## Background

The objective of the Task Force was not to develop a complete model of costs and value for bibliographic data, but to begin to identify sound measures that can inform decisions

by those engaged in the creation, exchange, and use of bibliographic data. The establishment of the task force was one response to the 2008 *On the Record: Report of the Library of Congress Working Group on the Future of Bibliographic Control*, in which the working group noted that the community has inadequate measures for moving forward on sound decision-making.<sup>41</sup> The Heads of Technical Services in Large Research Libraries Interest Group therefore charged the Task Force with identifying measures of the cost, benefit, and value of bibliographic control for key stakeholder communities, taking into account interdependencies between creators and consumers of bibliographic data, and developing a plan for implementing these measures.<sup>42</sup> Measures of cost and value, the charge read, could be granular and relative; for example, it could address the cost/value of controlled and uncontrolled name headings in different contexts or compare the differences between descriptive practices and standards used by libraries with those commonly used by the publishing or book trade industry. The charge also stated that stakeholders should include not only the end-users of library materials, but the parties and processes involved in the management of information resources and data, such as book vendors, system vendors, and software applications. Cost and value would be considered in relation to all sizes and types of libraries (public, academic, special, school, etc.). Interdependencies between creators and consumers of bibliographic data would be identified, since the benefits of bibliographic control may be separated from the current cost centers by multiple business processes, or may be cumulative over time.

The Task Force interpreted its charge broadly, encouraged to do so by the Working Group on the Future of the Bibliographic Control:

The phrase "bibliographic control" is often interpreted to have the same meaning as the word "cataloging." The library catalog, however, is just one access route to materials that a library manages for its users. The benefits of bibliographic control can be expanded to a wide range of information resources both through cooperation and through design. The Working Group urges adoption of a broad definition of bibliographic control that embraces all library materials, a diverse community of users, and a multiplicity of venues where information is sought.<sup>43</sup>

The Task Force therefore challenged itself to consider the value of bibliographic data in a variety of contexts and from a variety of perspectives. In doing so, the Task Force sought out a useful vocabulary for discussing value in relation to bibliographic control, but ultimately found none.

The Task Force also addressed vocabulary around

which to discuss cost. While one can outline elements contributing to the cost of cataloging (and work has been done in this area), evaluating those metadata costs and determining whether those costs are currently too high, without first having a clear understanding of their value, is difficult. When the LC changed its treatment of authorizing series headings in 2006, it opted for a cost-lowering technique without community metrics for assessing the value impact of that decision.<sup>44</sup> In the course of its work, the Task Force attempted to consider cost and value separately. Separate analyses can be pursued simultaneously to a point, but one cannot simply lower costs (unless one can figure out how to achieve exactly the same outcome for less cost) without discussing what would be lost in value. Before useful and specific measures could be written, therefore, the Task Force needed to reframe its work to propose community definitions for value. The Task Force, in its final report, suggested a research agenda for the community and recommended that the Heads of Technical Services in Large Research Libraries Interest Group identify institutions within this group or solicit partners from the community who are willing to contribute to an evolving effort of applying metrics to assess cost and value of bibliographic control.<sup>45</sup>

The context for the Task Force's charge was provided by section 5.1 of *On the Record: Report of the Library of Congress Working Group on the Future of Bibliographic Control*:

Bibliographic control occurs in a complex system of participants (contributors and users), information resources products and services, and technological capabilities. There are increasing numbers of participants, information formats and media, and information technologies. Contributors of bibliographic data and services may have different and sometimes conflicting agendas. Multiple user communities may have changing and expanding needs and expectations. In this increasingly complex environment, the actions taken by key players can have downstream impacts on others. *Unfortunately, there are still inadequate measures of the costs, benefits, and value of bibliographic information and almost no information on the interdependencies within the broader bibliographic control environment, including the impact of internationalization.*

Although the use of cost-benefit analysis for service organizations such as libraries is problematic, *all organizations must achieve goals and provide value.* Bibliographic control may be considered by many to be a public good, but it has real costs attached to it, just as, presumably, it has real value.<sup>46</sup> [emphasis added]

With the publication of *On the Record* in January 2008, the ALCTS board established the Task Group on the LC Working Group Report to analyze the recommendations put forward in the report and to identify those recommendations that ALCTS is well suited to address. In April 2008, the Task Group released of ten recommendations for the ALCTS community.<sup>47</sup> The ALCTS board then formed the Implementation Task Group, charged to identify ALCTS committees and others outside of ALCTS to take responsibility for moving one or more of the ten recommendations forward.

ALCTS's eighth recommendation was to bring together key participants to agree to implement a set of measures of costs, benefits, and value of bibliographic control for each group of participants and to identify interdependencies between participants.<sup>48</sup> At the 2009 American Library Association (ALA) Midwinter Meeting, the chair of the ALCTS Implementation Task Force proposed that the Heads of Technical Services in Large Research Libraries sponsor a task force to look at those measures of cost, value, and benefit for bibliographic control. The Heads of Technical Services in Large Research Libraries have the authority and leadership to bring key players together from their libraries and others to forge agreement on costs and benefits of bibliographic control—an effort that would serve not only research libraries, but potentially be of interest to libraries of all sizes and constituencies. The Task Force on Cost/Value Assessment of Bibliographic Control began work following the 2009 ALA Annual Conference.

The Task Force's report outlines its discussions of four fundamental questions necessary to defining metrics for value: (1) Can value be measured in ways that are non-numeric? (2) Is discussing relative value over intrinsic value helpful? (3) Does *value* equal *use*? and (4) Is it possible to define a list of bibliographic elements that are "high-value" and others that are "low-value"?<sup>49</sup> Given the difficulty in answering these questions, the lack of research into the area of value for bibliographic control, and the Task Force's desire to advance discussions about quantifying the value of bibliographic control in an environment where the vocabulary for doing so does not yet exist, the Task Force proposed seven operational definitions of value and offered suggestions for research in these areas. While the charge was to develop measures for value, the Task Force determined that doing so would not be helpful until the community has a common vocabulary for what constitutes value and an understanding of how value is attained, and until more user research into which bibliographic elements result in true research impact is conducted. The Task Force chose to scope the problem in a way to encourage discussion about value from various perspectives and provide next steps for institutions interested in taking on these crucial questions.

## Operational Definitions of Value and a Research Agenda

At the core of the Task Force's report are seven operational definitions of value with recommendations for a research agenda and strategies for advancing that research.

### Discovery Success

The Task Force identified discovery success as a key element of value and proposed research into which bibliographic elements produce useful retrieval results. While research exists into which elements are used in bibliographic data (largely MARC records), this research generally speaks to inputs—what catalogers are entering, based on what the rules proscribe—but does not directly or in measurable ways speak to which elements are of value to users. The MARC Content Designation Utilization Project ([www.mcdu.unt.edu](http://www.mcdu.unt.edu)) provides a wealth of statistical data on MARC tag use. Publications from that project also address the correlation between MARC tag use and cooperative cataloging guidelines and instructions. Smith-Yoshimura and colleagues have explored the implications of MARC tag use on library metadata practices.<sup>50</sup>

The Task Force suggested research in the following areas:

- Recognizing the inadequacy of log data currently generated by MARC-based systems, use search terms from user logs to evaluate which bibliographic elements match those search terms.<sup>51</sup> Non-MARC bibliographic systems might exist in which this data can be more easily and accurately captured.
- In addition to log analysis, directly watch user behavior to determine which records users clicked through to and why.
- Test discovery success in two systems when indexing two versions of the same record with and without certain metadata fields available. How does the presence or absence of elements affect users' ability to retrieve?
- In projects where brief records are being upgraded, capture the initial record set pre-upgrade and compare with discovery success post-upgrade.
- Identify delivery systems where one system indexes table of contents data and the other does not; research impacts on discovery from user log data.

Research into these areas presents challenges. Data across institutions would vary because of indexing and system design issues (such as last in, first out sorting decisions or relevancy). Assessing such data across institutions would cause

the community to ask questions about whether such differences are based on indexing decisions, display decisions, the nature of the collections, and other variables. While proving correlations between trends in findings and any particular factor would be difficult if the institutions comparing results ran tests under different conditions, by using statistical techniques such as meta-analysis, this approach could identify useful value similarities and would have the advantage of enabling analyses of the value of bibliographic data in the information ecosystem that includes systems design.

### Use

Use, represented largely as circulation (which may include in-house circulation for those libraries that capture these data), is a helpful measure of value. Use, quantified by circulation counts, has been examined for collection development and maintenance purposes, but not to assess the impact of library resources on a user's research. Hit counts on metadata records in a digital library environment are problematic because they are not always considered reliable measures of the user experience. Bollen, Van de Sompel, and Rodriguez wrote that usage data has great potential for analyzing scholars' use of resources.<sup>52</sup> Perneger argued that hit counts are not reliable measures of actual resource use because the number only reflects the visits to the website.<sup>53</sup> Miller wrote that hit counts are considered ambiguous because they include "all of the complex elements that are loaded separately to comprise that page as well as the Web crawlers."<sup>54</sup> Without standards to record and exchange the data, understanding the exact meaning of use data is difficult. Because they are numeric and quantifiable, use statistics may be a tempting but ultimately inadequate measure for articulating value. They are only one piece of a complex puzzle.

What does use mean for non- or low-circulating materials in libraries that have strong commitments to preserve the cultural record, including rarely requested primary source materials? And is value not derived from bibliographic control when a user decides from the metadata record that a particular item is *not* useful? Although use is clearly only part of the value equation, two questions are of critical interest. Do items with "better" records circulate more frequently or are electronic resources with "better" records more highly used? Is fuller bibliographic information valuable enough to be worth the cost?

The Task Force proposed a method for addressing these questions: where collections were shelved in open, browseable stacks before cataloging, compare circulation statistics of the same items before and after full cataloging. Criteria for choosing institutions would necessarily include running an integrated library system (ILS) that logs the catalog record's date of completion and that contains sufficient

historical circulation data. Alternatively, this could be a longitudinal study going forward.

### Display Understanding

Several research questions address this operational value. How much of the data that catalogers create do users understand? How frequently does a user go from a brief display to a full display? When a user does go to the full display for more information, what information is he or she seeking? When users request items from storage or through interlibrary loan (ILL), what is missing in the bibliographic display that would help them assess the usability of that item before requesting it? Assuming that some percentage of users of ILL or stored items request them to evaluate their usefulness for their research, how might the bibliographic record help improve this evaluation step?

Various research projects using user studies, including focus groups and other behavioral research, could address these questions.

- Ask what in a particular display is not understood, and what in the display helps the user decide this item is what he or she is seeking. Test the metadata with users from multiple approaches (i.e., the presence or absence of certain metadata, the displays of certain data elements for ease of use, and the rate of use and perceived usefulness of specific metadata elements). Particular attention should be paid to the elements that are beyond basic description, such as subject access, uniform titles, and classification. Another set of questions could involve user-assigned data—what would a user add if he or she could add something to a record to help the next person encountering it to determine whether the resource would be useful.
- Conduct testing of two iterations of the same interface (A/B testing), displaying different metadata elements.
- Survey users at the point of return of storage and ILL items.

The possible research projects are not without problems. Assessing the value of metadata separately from the quality of any particular discovery interface would be difficult. Data across institutions would vary because of system design issues. Assessing such data across institutions would cause the community to ask questions about whether such differences are based on indexing decisions, display decisions, the nature of the collections, and other reasons. Proving correlations between trends in findings and any particular factor would be difficult if the institutions comparing results ran tests under different conditions. However, using statistical techniques such as meta-analysis, this approach could identify useful value similarities and would have the advantage of enabling

analyses of the value of bibliographic data in the information ecosystem that includes systems design.

### Ability of Library Bibliographic Data to Operate on the Open Web and Interoperate with Vendors and Suppliers in the Bibliographic Supply Chain

The question here is where would libraries derive value if library bibliographic data were more integrated with web services (separate from or in addition to making library data more valuable to nonlibrary entities)? Certainly, the extent to which data employs a syntax that is machine processable contributes to the value of library data. Significant work has been undertaken in this area in preparation for RDA.<sup>55</sup> The community also needs further study on how much nonlibrary entities know about and understand library data and how the use of ONIX data is affected the library supply chain.

Suggested areas for research are:

- Research ONIX uptake throughout the bibliographic community. With several concrete ONIX-MARC projects underway, analysis can now be done to determine the extent to which ONIX data are valuable for cataloging workflows.<sup>56</sup>
- Select a set of ONIX records from a known publisher and track over time how that metadata are used throughout the supply chain to vendors of bibliographic data, OCLC, libraries (i.e., Program for Cooperative Cataloging upgrade) and out to the open web (Amazon, Google, LibraryThing, etc.) as a gauge of value and a measure of success in sharing data beyond library borders.
- Determine how much library data is currently being used outside the library ecosystem. While the potential here lies in the RDA Vocabularies as linked data, doing research on this now would give the community a baseline for comparing the extent of usability of library data in nonlibrary contexts now with what is hoped will happen when library data become more truly accessible on the open web.
- Analyze the extent to which library bibliographic data are successfully interacting with other programs in the user's bibliographic toolset (EndNote, Zotero, etc.).

### Ability to Support the Functional Requirements of Bibliographic Records (FRBR) User Tasks

RDA does not explicitly address which RDA elements support which FRBR user tasks. While RDA speaks directly to the bibliographic entities (work/expression/manifestation/item), the element lists do not speak directly to the facilitation of the user tasks (find/identify/select/obtain). RDA

includes discussion of the user tasks in the introductory matter for the relevant chapters, the chapters have been mapped to the FRBR user tasks, and a number of the chapter names reference a user task (e.g., “Identifying Manifestations and Items”).<sup>57</sup> Documentation around the development of the core element list explains that core elements were determined by assessing the value of those elements according to how they support the user tasks.<sup>58</sup> However, much of this work is buried in narrative, and a direct mapping of RDA elements to FRBR user tasks has not been issued. Surfacing these data more explicitly would be useful.

Suggestions for research into the ability of RDA to support the FRBR user tasks are:

- Undertake a mapping of the RDA elements to the FRBR user tasks.
- Undertake usability research to determine if, in fact, these elements do provide value towards facilitating the user tasks.

The Task Force sought to aggregate various datasets and documents to create a mapping of RDA bibliographic elements to FRBR user tasks and to illustrate a value ranking. The aggregated data are presented in an appendix that accompanied the Task Force’s final report.<sup>59</sup> The 2009 OCLC report *Online Catalogs: What Users and Librarians Want* calls particular attention to user desires for elements supporting delivery.<sup>60</sup> Users also requested discovery-related data, such as the ability to preview the book, cover art, summary and abstract data, and tables of contents data. While not all of these are covered by the RDA element set, summarization of the content, for example, was rated by IFLA as “low” for the identify task and “medium” for the select task. Work has been published by the MARC Content Designation Utilization Project ([www.mcd.u.unt.edu](http://www.mcd.u.unt.edu)) showing how catalogers code MARC tags in support of the FRBR user tasks, but it does not provide value research into the user perspective. The Task Force recommended that further work be done in this area to aim for a common understanding of stated value for individual bibliographic elements and to test the value of an element for a user task.

- In conjunction with other operational definitions of value above, determine which of these elements are commonly indexed, which are commonly displayed, which users pay attention to, which users understand, etc.
- Consider integrating an RDA-to-FRBR User Tasks mapping analysis into the RDA Toolkit ([www.rda-toolkit.org](http://www.rda-toolkit.org)). Such a resource could provide guidance to catalogers, particularly in light of the RDA Toolkit workflow functionality.

### Throughput and Timeliness

The extent to which data-creation processes facilitate timeliness in resource availability is a measure of value. Users cannot access materials that are sitting (literally or digitally) in uncataloged backlogs. Additionally, the value of editing existing records over cataloging materials completely lacking description is, of course, questionable. Research into this area ideally would demonstrate the effect on a community of not having new materials made quickly available.

The following areas are suggested as appropriate for research:

- Measure the uptake of the data created by catalogers. In cases where the resource itself is available to users both before and after release of metadata in the library’s discovery systems, compare resource use before full metadata has been loaded with use (in a defined timeline) after release of the metadata.<sup>61</sup>
- Identify older imprints newly added to WorldCat and then determine how quickly other institutions add their holdings once the record has been input. This metric would not demonstrate direct user impact, but it could show something about how quickly uptake of new cataloging occurs throughout the MARC bibliographic ecosystem. If OCLC does not retain long-term retrospective data on record edits in WorldCat, would performing a prospective rolling analysis on records newly added to the database be possible?
- For a set of materials, analyze publication dates against the dates when items were first acquired, first cataloged, and first circulated to identify trends in resource discovery and use. While other variables that affect discovery and use would be difficult to control, having an understanding of how quickly newly cataloged materials circulate could help determine appropriate throughput expectations.

### Ability to Support the Library’s Administrative and Management Goals

The question of which bibliographic elements provide value to the library for collection development, acquisitions, auditing, and inventory purposes beyond the value they provide for discovery or use by patrons needs to be addressed. One approach would be to survey the community to understand the value of the bibliographic data elements for librarians involved in managing collections.

### Value Multipliers

The Task Force discussed aspects of value that affect the operational definitions above:

- The extent to which bibliographic data are normalized
- The extent to which data support collocation and disambiguation in discovery
- The extent to which data have used controlled terms across format and subject domains
- The extent to which level of granularity matches what users expect
- The extent to which data enable a formal and functional expression of relationships (links between resources) to find “like” items
- The extent to which data are accurate
- The extent to which data enhancements are able to proliferate to all derivative records

All these items contribute to how valuable library data are in conjunction with the operational definitions proposed above. These are identified as “value multipliers” because they contribute to value, but the degree of contribution cannot be assessed until further research is done on the operational definitions outlined above and the community’s value goals become clearer.

### Measures of Cost

The Task Force also struggled with defining a vocabulary around which to discuss cost. While elements contributing to the cost of cataloging can be outlined (work has been done in this area), evaluating those metadata costs and determining whether those costs are currently too high is difficult without first having a clear understanding of value. Broadly, the following elements contribute to cost:

- Salary and benefits multiplied by the time for new record creation (for all bibliographic control activities, including searching for copy, original description, MARC encoding, classification, subject analysis, authority work, and local practices that vary from greater accepted practice)
- Cataloging tools (including Cataloger’s Desktop, Classification Web, OCLC, RDA Toolkit, WebDewey, etc.)
- Database maintenance (salary and benefits multiplied by the time on bibliographic and access (URL) corrections, vended authority control services, vended record upgrade notification services, activities such as “typo of the day,” etc.)
- Overhead (training, policy development, documentation, cooperative cataloging arrangements, the systems that they are built on, the practices that grow up around them, etc.)

While calculating cost for the creation of individual

elements or even areas of cataloging (such as authority control) by doing time studies is possible, doing so is most useful against a value question. The level of granularity needed to make the most meaningful analysis is not clear. The community also needs to be clear on the purpose. If the purpose is to bring down the costs generally, the method would be to calculate the costs listed above (in a way agreed on by the community) and work to develop systems or an infrastructure that would help lower those costs. If the purpose is to ask whether the tasks *are worth* the costs, then better research into the value questions above is needed first.

The Task Force discussed how the community might capture the costs of many individual bibliographic elements and, while even small costs add up over time given the way bibliographic description is done, imagining how one might calculate the cost of creating individual bibliographic elements is difficult. This direction also puts the emphasis on initial record creation and overlooks costs of maintaining the integrity of bibliographic databases over time. Much of the cost of bibliographic control is not in the original data creation but in metadata-maintenance activities that come later in the lifecycle.

Alternatively, the Task Force considered suggesting an extremely simple solution, such as the number of volumes cataloged divided by salaries, but this type of calculation will fail to illustrate true costs. *Festschrift* is an emblematic example. The act of coding the fixed field value is not where the cost lies, but that is the cost most easily captured. True costs (and true savings, if catalogers were to stop coding this value or many others) are in the overhead category of training and documentation, which are significantly harder to quantify.

Deviations from standard practices carry added cost and consume additional processing time, often in the work and always in the overhead (training, policy development, documentation, etc.). The Task Force challenged libraries to determine how much their local practices are costing and to undertake conversations about whether they are of appropriate value to their constituencies. All libraries must actively decide what they are willing to pay to support their priorities. Local practices are often brought into question by outside forces: consultants, new administration or staff, planning for vended cataloging services, and so on. These influences force articulation and assessment of local priorities. The Task Force encouraged cataloging departments to embrace a culture of continuous cost/value discussion and assessment and, when possible, to invite objective, external influences to the discussion that will elicit attentiveness to library priorities and their associated costs.

Finally, opportunity costs need to be quantified. Time spent on low-value activities (no matter which operational definition is used for “value”) is time *not* spent on high-value activities. Having materials sitting in technical

services waiting for copy to appear while libraries edit existing records inhibits discovery and use of collections. In its final report, the Task Force considered the research conducted by R2 Consulting, who completed a report to the LC in the midst of the Task Force discussions.<sup>62</sup> This 2009 report, *Study of the North American MARC Records Marketplace* states,

Our survey results also confirm our direct observation of many “aging” backlogs in libraries. Because of their own staffing constraints, or unwillingness to bear the cost of original record creation, many libraries simply wait for another library to catalog an item they have already received. On average those items are held for three to six months, with periodic searches of OCLC to determine whether another library has blinked. While this makes sense as a way of controlling costs, it does not provide optimal service for users.<sup>63</sup>

The extent to which data-creation processes facilitate timeliness in resource availability is a crucial component of value. Additionally, the failure to contribute meaningful edits to the national community causes the community at large to repeatedly pay editing costs. While creating a metric that would calculate that cost may be possible, it is undeniably larger than zero. Any number larger than zero is no longer sustainable.

### Interdependencies between Creators and Consumers of Bibliographic Data

The final element of the Task Force’s charge was to identify the interdependencies between creators and consumers of bibliographic data. The Task Force believed that this was well documented for the MARC record ecosystem by the *Study of the North American MARC Records Marketplace*. Appendix B of this report outlines the stakeholders in the MARC record ecosystem.<sup>64</sup> While the Task Force noted that the R2 work was scoped to MARC, the Task Force did not identify any missing components particular to the ARL community within that scope. R2’s focus on MARC records specifically and within the realm of LC production puts a different slant on cost and value than in the Task Force charge. The Task Force focused on the value of metadata to users, while R2 focused on the value of bibliographic data to libraries as organizations (as measured by cost reduction). The Task Force felt that, if significant metadata-production changes are made at the LC, these will affect methods the community must develop. The non-MARC bibliographic marketplace is significantly less defined, but the creators and consumers of bibliographic data in that ecosystem can

be placed in the R2 context as well. Libraries are creating original non-MARC data, vendors are creating (and selling) original non-MARC data, and aggregators (commercial and noncommercial) are creating non-MARC data through services such as OAI.

The Task Force therefore believed that appendix B of the R2 report captures the stakeholder relationships for the ARL community in the MARC ecosystem and may be extended to encompass non-MARC metadata-creation partners as well.

### Conclusion

In 1956, in an article titled “Cataloging Cost Factors,” Swank wrote,

For purposes of evaluation, studies of the use of the catalog would be helpful if related to costs. If we could know, for example, the utility of various added entries and could tell the difference in cost if they were or were not made, we might be able to pass reasonable judgment. But even studies of the use and cost of the catalog would leave much to the imagination, because they would still fail to inform us about the relations of the catalog to other kinds of bibliography. Even though it were demonstrated that a job needs to be done and could be done at reasonable cost in the catalog, there would still be the possibility that the same job might be done better at less cost in some other way. The most valuable single kind of study that could be made at this time, I believe, would be case studies of the experience of readers in using the entire range of a library’s bibliographical services studies that could then be related to analyses of the costs of the entire range of services. . . . The whole area is a great maze which will never be untangled until (a) adequate studies of readers’ needs have been made, (b) the most economical bibliographical [*sic*] means of satisfying those needs have been determined, and (c) the role of the catalog as one of those means has been established. This is a big order, perhaps an impossible one.<sup>65</sup>

Fifty-four years later, these unknowns persist. In the current economic climate, the community must strive to untangle the maze. The Task Force found its charge difficult, but believed that—with more research into value—developing measures of cost and value that communities could agree on is possible. The Task Force submitted its report with the hope to engage conversation about what constitutes value for bibliographic control. The Task Force

outlined seven operational definitions of value:

- Discovery success
- Use
- Display
- Ability of library bibliographic data to operate on the open web and to interoperate with vendors and suppliers
- Ability to support FRBR user tasks
- Throughput and timeliness
- Ability to support the library's administration and management goals

These were offered as a means to frame the problem and to encourage discussion into value from various perspectives, scope the value questions—which are overwhelmingly large at first pass—into segments that are more accessible to undertake, and provide next steps for institutions interested in engaging in these crucial questions.

As representatives of the large research community, the Task Force submitted its report to the Heads of Technical Services in Large Research Libraries Interest Group, but also felt that ownership of these questions is shared by entities of all sizes and types (including vendors of bibliographic data) within the library community. The Task Force hoped that the community would amass enough data to start analyzing the results. While some strategies would best be undertaken by a single, centralized entity, individual institutions could do much of this work on a smaller scale in line with their institutional missions and begin to pool that information in search of aggregate commonalities and differences. While the Task Force directs its recommendations in the final report to the Heads of Technical Services Interest Group, its desire is that any institution with an interest in addressing any of the recommendations therein should feel free and is encouraged to do so.

#### References and Notes

1. William Warner Bishop, *Cataloging As an Asset: An Address to the New York State Library School, May 15, 1915* (Baltimore: Waverly, 1916): 4.
2. *Ibid.*, 7.
3. *Ibid.*, 8.
4. *Ibid.*, 16–17.
5. Early attempts at discussing, applying, and debating cost measures include Charles A. Cutter, “Dr. Hagan’s Letter on Cataloging,” *Library Journal* 1 (1877): 219; James L. Whitney, “On the Costs of Cataloging,” *Library Journal* 10 (1885): 214–16; William Warner Bishop, “Some Considerations on the Cost of Cataloging,” *Library Journal* 30 (1905): 10–14; Aksel G. S. Josephson, “Committee on Cost and Method of Cataloging,” *Library Journal* 39 (1914): 598–99; Aksel G. S. Josephson, “The Cataloging Test: Results and Outlook,” *ALA Bulletin* 10 (1916): 242–44; “Plan for an Investigation into and Report on -the Cost of Cataloging,” *Bulletin of the American Library Association* 19 (1925): 278–86; Adah Patton, “The Cost of Cataloging,” *Library Journal* 51 (1926): 140–41; “The Cost of Cataloging,” *Library Journal* 32 (1927): 239; Fremont Rider, “Library Cost Accounting,” *Library Quarterly* 6, no. 4 (1936): 331–81; Robert A. Miller, “Cost Accounting for Libraries: Acquisition and Cataloging,” *Library Quarterly* 7, no. 4 (1937): 511–36; Jerrold Orne, “We Have Cut our Cataloging Costs!” *Library Journal* 73 (1948): 1475–87; Felix Reichmann, “Cost of Cataloging,” *Library Trends* 2 (1953): 290–317.
6. George Harris, “Historic Cataloging Costs, Issues, and Trends,” *Library Quarterly* 59, no. 1 (1989): 1–21.
7. *Ibid.*, 1.
8. Keyes D. Metcalf, “Attitude of the Library Administrator Toward Cataloging,” *ALA Bulletin* 35, no. 8 (Sept. 1941): 48.
9. R. C. Swank, “Cataloging Cost Factors,” *Library Quarterly* 26, no. 4 (1956): 303–17.
10. *Ibid.*, 307.
11. Joint Steering Committee for Development of RDA, *Resource Description and Access: RDA* (Chicago: ALA, 2010–).
12. Swank, “Cataloging Cost Factors,” 309.
13. Jeffrey A. Raffel and Robert Shishko, *Systematic Analysis of University Libraries: An Application of Cost–Benefit Analysis to the MIT Libraries* (Cambridge, Mass.: MIT Pr., 1969).
14. Arthur McAnally, review of *Systematic Analysis of University Libraries*, by Jeffrey A. Raffel and Robert Shishko, *Library Quarterly* 40, no. 3 (1970): 355.
15. Raffel and Shishko, *Systematic Analysis of University Libraries*, 46.
16. Bishop, *Cataloging As an Asset*.
17. Amar K. Lahiri, “Toward a Bibliographic Common Cause,” *Cataloging & Classification Quarterly* 8, no. 1 (1987): 66.
18. *Ibid.*
19. Ake I. Koel, “Bibliographic Control at the Crossroads: Do We Get Our Money’s Worth?” *Journal of Academic Librarianship* 7, no. 4 (1981): 220–22.
20. Lahiri, “Toward a Bibliographic Common Cause,” 70.
21. *Ibid.*
22. *Ibid.*, 71–72.
23. *Ibid.*, 73.
24. *Ibid.*, 74–76.
25. Tefko Saracevic and Paul B. Kantor, “Studying the Value of Library and Information Services. Part I. Establishing a Theoretical Framework,” *Journal of the American Society for Information Science* 48, no. 6 (1997): 527–42; Saracevic and Kantor, “Studying the Value of Library and Information Services. Part II. Methodology and Taxonomy,” *Journal of the American Society for Information Science* 48, no. 6 (1997): 543–63.
26. *Ibid.*, 527.
27. *Ibid.*, 562.
28. Roxanne Missingham, “Libraries and Economic Value: A Review of Recent Studies,” *Performance Measurement and Metrics* 6, no. 3 (2005): 142–58; Susan Imholz and Jennifer Weil Arns, *Worth Their Weight: An Assessment of the Evolving Field of Library Valuation* (New York: Americans for Libraries Council, 2007), www.bibliotheks

- portal.de/fileadmin/0themen/Management/dokumente/WorthTheirWeight.pdf (accessed Sept. 7, 2010).
29. Judy Luther, *University Investment in the Library: What's the Return? A Case Study at University of Illinois at Urbana-Champaign*, Library Connect White Paper no. 1 (San Diego: Elsevier, 2008), <http://libraryconnect.elsevier.com/whitepapers/lcwp0101.pdf> (accessed Sept. 1, 2010); Paula T. Kaufman, "The Library as Strategic Investment: Results of the Illinois Return on Investment Study," *Liber Quarterly* 18, no. 3/4 (2008): 424–36.
  30. Cornell University Library, *Cornell University Library Research & Assessment Unit: Making Data Make Sense, Library Value Calculations*, <http://research.library.cornell.edu/value> (accessed Sept. 1, 2010).
  31. Cheryl McCain and Jay Shorten, "Cataloging Efficiency and Effectiveness," *Library Resources & Technical Services* 46, no. 1 (2002): 23–31.
  32. Dilys E. Morris et al., "Cataloging Staff Costs Revisited," *Library Resources & Technical Services* 44, no. 2 (2000): 70–83.
  33. Dilys E. Morris et al., "Where Does the Time Go? The Staff Allocations Project," *Library Administration & Management* 20, no. 4 (2006): 177–91.
  34. Shawne D. Miksa, "You Need My Metadata: Demonstrating the Value of Library Cataloging," *Journal of Library Metadata* 8, no. 1 (2008): 23–36.
  35. *Ibid.*, 27.
  36. *Ibid.*, 34.
  37. Philip Hider, "How Much Are Technical Services Worth? Using the Contingent Valuation Method to Estimate the Added Value of Collection Management and Access," *Library Resources & Technical Services* 52, no. 4 (2008): 254–62.
  38. *Ibid.*, 258.
  39. Michael Gorman, *Our Enduring Values: Librarianship in the 21st Century* (Chicago: ALA, 2000): 114.
  40. Joint Steering Committee for Development of RDA, *Resource Description and Access: RDA*, Objective 0.4.2.2.
  41. Library of Congress Working Group on the Future of Bibliographic Control (LC Working Group), *On the Record: Report of the Library of Congress Working Group on the Future of Bibliographic Control*, Jan. 9, 2008, [www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf](http://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf) (accessed Aug. 31, 2010).
  42. The charge and membership of the group is included in the *Final Report of the Task Force on Cost/Value Assessment of Bibliographic Control*, submitted to the Heads of Technical Services in Large Research Libraries Interest Group, June 18, 2010, <http://connect.ala.org/node/106017> (accessed Sept. 1, 2010).
  43. LC Working Group, *On the Record*, 10.
  44. Library of Congress, "Series at the Library of Congress: June 1, 2006," [www.loc.gov/catdir/cpsos/series.html](http://www.loc.gov/catdir/cpsos/series.html) (accessed Nov. 14, 2010).
  45. *Final Report of the Task Force on Cost/Value Assessment for Bibliographic Control*.
  46. Library of Congress Working Group on the Future of Bibliographic Control, *On the Record*, 37–38.
  47. ALCTS Task Group on the LC Working Group Report, "On the Record: Report of the LC Working Group on the Future of Bibliographic Control: Ten Actions for ALCTS," [www.ala.org/ala/mgrps/divs/alcts/ianda/bibcontrol/lcwgtop10.cfm](http://www.ala.org/ala/mgrps/divs/alcts/ianda/bibcontrol/lcwgtop10.cfm) (accessed Aug. 31, 2010).
  48. *Ibid.*
  49. *Final Report of the Task Force on Cost/Value Assessment for Bibliographic Control*, 5–6.
  50. Karen Smith-Yoshimura et al., *Implications of MARC Tag Usage on Library Metadata Practices* (Dublin, Ohio: OCLC Research, 2010), [www.oclc.org/research/publications/library/2010/2010-06.pdf](http://www.oclc.org/research/publications/library/2010/2010-06.pdf) (accessed Aug. 31, 2010).
  51. *Ibid.*, 15–16
  52. Johan Bollen, Herbert Van de Sompel, and Marko A. Rodriguez, "Towards Usage-Based Impact Metrics: First Results from the MESUR Project," Apr. 23, 2008, <http://arxiv.org/abs/0804.3791> (accessed Sept. 2, 2010).
  53. Thomas V. Perneger, "Relation between Online 'Hit Counts' and Subsequent Citations: Prospective Study of Research Papers In the BMJ," *BMJ* 329, no. 7465 (Sept. 4, 2004): 546–47.
  54. Rush G. Miller, "Shaping Digital Library Content," *The Journal of Academic Librarianship* 28, no. 3 (2002): 100.
  55. Diane Hillmann et al., "RDA Vocabularies: Process, Outcome, Use," *D-Lib Magazine* 16, no. 1/2 (Jan./Feb. 2010), <http://dlib.org/dlib/january10/hillmann/01hillmann.html> (accessed May 15, 2011); Open Metadata Registry, "The RDA (Resource Description and Access) Vocabularies," <http://metadataregistry.org/rdabrowse.htm> (accessed Aug. 31, 2010).
  56. Projects underway include Next Generation Cataloging, [www.oclc.org/partnerships/material/nexgen/nextgencataloging.htm](http://www.oclc.org/partnerships/material/nexgen/nextgencataloging.htm) (accessed Sept. 7, 2010); OCLC Metadata Services for Publishers, <http://publishers.oclc.org/en/metadata/default.htm> (accessed Aug. 31, 2010); Library of Congress, ONIX Pilot, <http://cip.loc.gov/onixpro.html> (accessed Aug. 31, 2010); research being conducted by Myung-Ja Han at University of Illinois at Urbana-Champaign, and Carol Jean Godby, *Mapping ONIX to MARC* (Dublin, Ohio: OCLC, 2010), [www.oclc.org/research/publications/library/2010/2010-14.pdf](http://www.oclc.org/research/publications/library/2010/2010-14.pdf) (accessed Sept. 2, 2010).
  57. Tom Delsey, RDA Editor, "RDA, FRBR/FRAD, and Implementation Scenarios," memorandum to Joint Steering Committee for Development of RDA, Jan. 23, 2008, [www.rda-jsc.org/docs/5editor4.pdf](http://www.rda-jsc.org/docs/5editor4.pdf) (accessed Aug. 31, 2010).
  58. Deirdre Kiorgaard, "RDA Core Elements and FRBR User Tasks," memorandum to Joint Steering Committee for Development of RDA, Nov. 5, 2008, [www.rda-jsc.org/docs/5chair15.pdf](http://www.rda-jsc.org/docs/5chair15.pdf) (accessed Aug. 31, 2010).
  59. "RDA Value Matrix, Task Force on Cost/Value Assessment of Bibliographic Control," appendix to Final Report of the Task Force on Cost/Value Assessment for Bibliographic Control, June 18, 2010, <http://connect.ala.org/node/106017> (accessed Feb. 5, 2011).
  60. OCLC, *Online Catalogs: What Users and Librarians Want* (Dublin, Ohio: OCLC, 2009), [www.oclc.org/reports/onlinecatalogs/fullreport.pdf](http://www.oclc.org/reports/onlinecatalogs/fullreport.pdf) (accessed Feb. 5, 2011).
  61. Existing research demonstrates increased use of e-books when MARC records are integrated into the online public

access catalog, but it does not consider the importance of timeliness in record creation and integration. See, for example, Jacqueline Belanger, "Cataloguing E-Books in UK Higher Education Libraries: Report of a Survey," *Program: Electronic Library & Information Systems* 41, no. 3 (2007): 203–16; Dennis Dillon, "E-Books: The University of Texas Experience, Part 2," *Library Hi Tech* 19, no. 4 (2001): 350–62; Susan Gibbons, "NetLibrary eBook Usage at the University of Rochester Libraries, Version 2," Sept. 27, 2001,

- [www.lib.rochester.edu/main/studies/analysis.pdf](http://www.lib.rochester.edu/main/studies/analysis.pdf) (accessed Aug. 31, 2010).
62. R2 Consulting, *Study of the North American MARC Records Marketplace*, Oct. 2009, [www.loc.gov/bibliographic-future/news/MARC\\_Record\\_Marketplace\\_2009-10.pdf](http://www.loc.gov/bibliographic-future/news/MARC_Record_Marketplace_2009-10.pdf) (accessed Aug. 31, 2010).
  63. *Ibid.*, 27.
  64. *Ibid.*, appendix B.
  65. Swank, "Cataloging Cost Factors," 312–13.

# Open Access Literature Review 2008–9

## A Serials Perspective

**Maria Collins**

*Stemming from a previously published serials literature review by Library Resources and Technical Services (LRTS), this paper provides a review of a subset of the serials literature published in 2008 and 2009 focusing on open access (OA). The broader scope of the serials literature sets the stage for a culture of openness receptive to the OA movement. Catalysts to this movement, such as the National Institutes of Health (NIH) mandate, university OA policies, and increased emphasis on self-archiving in institutional repositories (IRs), are of interest to serials and electronic resource professionals who steward academic research collections. This interest is exemplified by the significant number of open-access-related articles intertwined throughout the serials literature of 2008 and 2009. Topics covered in this article include the NIH mandate, universities' responses to the NIH mandate, overviews of OA, the IR as a model of OA, strategies for supporting IRs, and evaluation of the effects of OA on scholarly communication.*

In a separate review of the 2008 and 2009 serials literature published by *Library Resources and Technical Services (LRTS)*, the author excluded the topic of open access (OA) because of its appearance in a preponderance of articles. The extensive treatment of OA across the 2008 and 2009 serials literature merited separate investigation and provides the context for this paper. The literature reveals numerous reasons for serials professionals' enamored views of OA. Given the background of the economic recession that began in 2007, OA has been touted as a possible saving grace to the crisis in scholarly communication and a potentially viable solution to the ever-increasing, unsupportable inflation of library subscription prices. In addition, the Internet has helped to resolve concerns about the feasibility of distributing OA content. Librarians also have embraced a culture of openness across the information environment—supporting open systems, new industry standards for interoperability, data exchange, and Web 2.0 concepts of connectivity. The library environment of 2008 and 2009 was ripe for change, and the plethora of articles written on OA during this time reveals the profession's fascination with the possibilities of the OA movement.

A quick search for “open access” in the Library and Information Science Abstracts (LISA) database resulted in more than 500 hits for 2008 and 2009. An exhaustive review of the literature on this topic is far beyond the scope of this paper; instead, this review focuses on the open-access-related conversations embedded within a serials context or OA discussions in publications targeting serials or electronic resource librarians. Topics covered by the review include the National Institutes of Health (NIH) mandate, universities' responses to the NIH mandate, overviews of OA, the institutional repository (IR) as a model of

**Maria Collins** (maria\_collins@ncsu.edu) is Associated Head of Acquisitions, North Carolina State University Libraries, Raleigh, North Carolina.

Submitted November 5, 2010, and tentatively accepted, pending modest revision; revision submitted January 26, 2011, and accepted for publication February 5, 2011.

I wish to acknowledge the contributions of my husband, Leonard Collins, for writing the reference notes and providing thoughtful edits to this paper (as well as looking after our three children). He has been wonderfully supportive and patient during the many hours spent on this project. I also would like to thank Liz Burnette, Head of Acquisitions at NCSU Libraries, for her continued support of my research and professional service.

# Books as Expressions of Global Cultural Diversity

## Data Mining for National Collection Analysis

Timothy J. Dickey

*A number of bodies have been jointly interested in book publication data as measures of cultural diversity. The United Nations Educational, Scientific, and Cultural Organization (UNESCO) Institute for Statistics is especially interested in global patterns in book publication as expressions of cultural diversity and heritage. Such data, however, are not widely collected by national publishing organizations and library statistics agencies. The increasingly global reach of the WorldCat database, on the other hand, makes it an obvious source for data mining. This paper presents results from an OCLC Research project that produced a rich data portrait of global book publishing, with emphasis on collection analysis by country. Researchers were able to compare the annual publishing for every country of the world (as reflected in WorldCat), the libraries that collect and import a country's works, the monographs their libraries import from other countries, and the proportion of publications in various official and native languages. The results provide a global overview of book publishing and a wealth of case studies in single countries' practices in book publishing and the preservation of their literary heritage. The present paper compares the book publishing and book collections in libraries in six countries around the world and demonstrates the power of data mining within this sphere.*

**Timothy J. Dickey** (tdickey1@kent.edu) is an adjunct faculty member at Drexel University, Philadelphia; Kent State University, Kent, Ohio; and San Jose State University, San Jose, California.

Submitted November 11, 2010; tentatively accept for publication December 13, 2010, pending modest revision; revision submitted February 16, 2011; accepted for publication March 15, 2011.

This research was conducted when the author was a postdoctoral researcher at OCLC Research in Dublin, Ohio. Outcomes from the research project have been previously reported in an OCLC Research Webinar (Sept. 16, 2010), at the Library History Seminar XXII: Libraries in the History of Print Culture (Madison, Wisc., Sept. 11, 2010), at the XXIX annual Charleston Conference (Charleston, S.C., Nov. 5, 2009), and at the 4th International Conference on the Arts in Society (Venice, Italy, 28–31 July 2009, not presented in person). The author wishes gratefully to acknowledge the helpful comments and contributions of Lynn Silipigni Connaway, Jeremy Browning, Karen Smith-Yoshimura, Eric Childress, and the anonymous reviewers in the preparation of this paper.

“The limits of my language mean the limits of my world.”

—Ludwig Wittgenstein, *Tractatus Logico-Philosophicus*, 5.6

Globally and nationally, book publishing represents a central kind of cultural heritage. A number of bodies have found themselves jointly interested in any statistics to measure book publication as a measure of cultural diversity. The United Nations Educational, Scientific, and Cultural Organization (UNESCO) Institute for Statistics has been exploring library statistics for worldwide book consumption and helped found the European Expert Meeting on Book and Library Statistics.<sup>1</sup> These bodies, as well as the International Federation of Library Associations (IFLA), are especially interested in any global patterns in the publication world as expressions of cultural diversity and heritage. Such data, however, are not widely collected by any national publishing organizations or library statistics agencies. The increasingly global reach of library records in the WorldCat database, on the other hand, makes it an obvious source for data mining. OCLC's bibliographic database represents more than 200 million items, with 1.6 billion copies held by libraries worldwide.<sup>2</sup> It is well suited to serve as a global aggregate library collection, to be mined for data on national publication and library collection profiles.

This paper presents the fruits of an OCLC Research project that aimed to compare national publishing profiles and to determine whether WorldCat provides an adequately diverse bibliographic collection to allow comparison worldwide.<sup>3</sup> The project used the method of data mining to research specific aspects of the global literary arts, with emphasis on collection analysis by country and region. Researchers specifically attempted to profile the annual publishing for every country of the world (as reflected in the WorldCat database), the libraries that collect and import a country's works, the monographs their libraries import from other countries, and the proportion of publications in various official and native languages, as well as data on translated works. The results provide a global overview of the publishing arts and a wealth of case studies in single countries' practices in both literary publishing and the preservation of their literary heritage.

This paper compares case studies of book publishing and library book collections from six countries: Bolivia, Chile, Germany, Poland, South Africa, and Thailand. As a set of test cases, the six were chosen to highlight non-English works and non-English cataloging, and to reflect a mix of continents across the world, of development levels, and of OCLC member libraries' contributions to the database. All data presented are limited by librarians' contributions to WorldCat. However, the six case studies amply demonstrate the strength of the data-mining methodology within this sphere and the richness of the data possible to mine for historical profiles of national bibliographies.

### Literature Review

The UNESCO Institute for Statistics has been exploring library statistics for worldwide book consumption, actively promoting the 2005 Convention on the Protection and Promotion of the Diversity of Cultural Expressions.<sup>4</sup> Specifically, one tenet of the expected results is "Linguistic diversity [being] promoted through publishing and translation."<sup>5</sup> A number of bodies—UNESCO, IFLA, the International Publishers' Association (IPA), and the International Booksellers' Federation (IBF)—met at the European Expert Meeting on Book and Library Statistics in November 2008 to explore potential pilot projects to collect national bibliographic statistics; some of the individuals involved include Michael Heaney, executive secretary of Oxford University Library Services and chair of the ILFA Committee for Statistics and Evaluation; Simon Ellis, head of Cultural Expressions and Creative Industries, UNESCO Institute for Statistics (Montreal); and Mauro Rossi, UNESCO Division of Cultural Expressions (Paris).<sup>6</sup> Similar meetings occurred at the request of IFLA in June and September of 2008, resulting in a consultant's report on

book statistics, though funding for further steps at that time was not available; contact with OCLC Research offered one option for further research.<sup>7</sup>

One specialized expression of the interest these bodies have is tracking any available statistics on global book publishing in the Index Translationum, an "international bibliography of translations" published by the UNESCO Division of Cultural Expressions in print since 1932 and in digital form since 1979.<sup>8</sup> This resource tracks the translations of a culture's books into other languages, and the online database contains more than 2 million records. However, this excellent tool is somewhat limited in scope because it is dependent on data donated annually by participating national libraries and uses a proprietary data format—it has, for instance, no ISBN field for linking its data to other bibliographic tools. In addition, neither book publishing organizations, nor national libraries as a whole, nor library statistics agencies have been collecting data on book publishing on any global scale.<sup>9</sup>

For countries around the world, some literature addresses national bibliographies. Recent examples include articles relating to the national bibliographies of Canada, Korea, a collection of Eastern European nations, and Mauritius.<sup>10</sup> Since 2004, IFLA has published the journal *International Cataloging and Bibliographic Control*; this publication has presented much valuable research, especially in the area of standards for international cooperation in cataloging. However, a noticeable gap remains in global bibliographic control and in comparative studies.

OCLC Research had recently developed one prototype service, the OCLC WorldMap, for visualization of global library data.<sup>11</sup> OCLC Researchers Lynn Connaway, Timothy Dickey, and Jeremy Browning, with OCLC librarian Lawrence Olszewski, developed the prototype service to mine, collect, and compare library data from both WorldCat and standard print reference sources and to compare bibliographic information of different countries. The WorldMap graphically portrays and compares library and bibliographic data, including titles published in a country, holdings worldwide of those titles, libraries in a country and their type, national expenditures on those libraries, and concentrations of archives, museums, and other cultural heritage institutions.<sup>12</sup> UNESCO interest in this particular prototype and in OCLC Research experience in the method of data mining led to the research project reported here.

### Research Method

The method selected was data mining. OCLC's Office of Research has invested significant effort in the area of data mining.<sup>13</sup> Data mining is the computer-aided analysis of databases and other large datasets (such as records of

website hits, logs of transactions within an automated system such as a library catalog or online retailer, or electronic stores of demographic data) to expose new information derived from the aggregate of the data. The technique first appeared as a tool for business intelligence, only later to be adopted by libraries; the success of Google and Amazon has taught the library field that greater value exists within bibliographic data as well. Libraries have made huge investments in creating and maintaining rich, structured information describing the resources in their collections. These data already embody considerable value by supporting basic local access and inventory control. They also represent potential value in terms of knowing more about the characteristics of library collections. OCLC Research sees data mining as an effort to get increased return on libraries' investment in this bibliographic data.

Specifically, research projects have demonstrated the value of the WorldCat database as an "aggregate collection" of bibliographic data.<sup>14</sup> As a global-scale dataset of potential value, it can "not only provide librarians data for decision-making for collection and service development, but also provide users with enhanced discovery and access methods."<sup>15</sup> The WorldCat database is an increasingly global and increasingly comprehensive source of bibliographic data and remains strongest in its data on books. WorldCat contains more than 200 million records, with more than 1.6 billion holdings of those resources; approximately 54 percent are non-English catalog records, illustrating the increasingly global reach of the "aggregate collection."<sup>16</sup> Its member libraries are located in more than 100 countries, and the data go beyond those countries to include works from countries that are collected in other OCLC member libraries. Perhaps most importantly for this project, WorldCat contains publications in more than 470 languages; language was a central part of the definition UNESCO and OCLC considered while developing the OCLC Research project "Books as an Expression of Cultural Diversity." WorldCat, although somewhat limited by its Anglo-centric roots, offers a globally aggregated source of bibliographic data to examine the question at hand.

The basic objective of the project, then, is to mine WorldCat's overwhelmingly monographic records for data on global book publication and collection patterns. Researchers parsed the monographic data by country of publication, year of publication, and language use as a measure of cultural diversity. An axiomatic concept on the importance of language exists in cognitive anthropology that (in the oft-quoted thoughts of Benjamin Lee Whorf) "language shapes the way we think, and determines what we can think about."<sup>17</sup> The so-called Sapir-Whorf Hypothesis offers the concept that language functions not only as a framework for communication between speakers, but also as a framework for our basic comprehension of the world; it has been tempered

somewhat in regards to linguistic relativity starting in the 1980s, but remains an influential interpretation.<sup>18</sup> The language or languages spoken by a culture help determine that culture's perception of the world and its expression of itself within that world; thus language data remain important to tracking books as expressions of any culture.

Within this method of data mining, OCLC researchers set specific data limits. To filter only textual materials, the MARC leader field "Type of Record" must carry the value T06 = a (for books), and leader field must be 07 = a or m (indicating a record for either a monograph component part or a complete monograph); this procedure excluded serials, theses, and dissertations, but the data mining otherwise cast a deliberately wide net. Publication dates in the catalog records (also taken from the fixed field elements where possible) had to be numeric and less than or equal to 2010; this filtered out works coded with publication dates of, for instance, "19xx," which could not be folded reliably into the rest of the data. Dates as early as 1000 AD were included, but these records tended to be outliers in the data for any particular country.

Technical staff extracted all bibliographic records that passed these filter into databases, which were subject to internal analysis and are available for further work by outside researchers. Each bibliographic record was counted in Functional Requirements for Bibliographic Records (FRBR) terms, so manifestations were counted rather than works because of the assumption that any new edition (be it of Shakespeare or the Bhagavad-Gita) is a fresh cultural artifact and pertinent to the cultural collection. Library holdings were counted for each manifestation—both worldwide holdings of the nation's cultural heritage and, where possible, the balance between worldwide holdings and holdings within the country of publication, a measure of how foreign libraries collect and value that cultural heritage. Researchers collected all language data possible from the fixed-field data elements, as well as from the 041 (language code) field, which contains not only three-letter codes for the language of the item being cataloged, but also subsidiary codes for the original language or languages if translation of the original has occurred.<sup>19</sup> The language data thus was able to represent how both "official" and "indigenous" languages are represented in a country's bibliographic heritage, and also how multiple linguistic content appears in the national collection.

This paper covers a sample of results, primarily case studies in the data from six countries (Bolivia, Chile, Germany, Poland, South Africa, and Thailand) in an effort to answer the following questions:

- How do these national bibliographic collections (as reflected in WorldCat) compare to one another?
- Does WorldCat provide an adequately diverse *global*

bibliographic collection to allow such comparisons worldwide? In other words, are the national bibliographies, as reflected in WorldCat, distinct enough from one another, and distinct in culturally and historically justifiable ways, to support the WorldCat data as globally reliable?

The six countries (Bolivia, Chile, Germany, Poland, South Africa, and Thailand) addressed in this paper deliberately highlight non-English works and non-English cataloging. They include some mix of continents across the world, as well as a mix of development levels, and a mix of OCLC member libraries' contribution to the database. South Africa has eleven official languages; the other countries have single dominant languages and a handful of minority tongues. Data have been extracted from WorldCat (at the time of writing) for all non-U.S. countries, totaling more than 66 million bibliographic records and 450 million holdings. The findings from the six profiles tend to support the integrity of the worldwide dataset mined from the data in the WorldCat database in the distinctiveness of the profile they give of each national bibliography.

### Limitations

All of the data discussed below are subject to the caveat that a national bibliographic profile is being constructed as reflected in WorldCat. This means that it is subject to what libraries—and specifically libraries that participate in WorldCat—*have* collected and cataloged. For some countries in the present comparison, such as Thailand and Bolivia, very few libraries in each country have been OCLC members, and thus the data will tend to reflect more of what other Anglo-American libraries and other major national libraries participating in WorldCat have collected of their publications.

In addition, several issues of Anglo-American cataloging practice and other cataloging standards as translated into MARC 21 affect the following profiles. For example, different cataloging standards may have different (or changing) concepts of what will be coded as a book. Even more difficult for the purpose of this research is the definition of a country. The German data below, for example, reflect generations of different catalogers' judgments of several centuries of shifting boundaries that have finally coalesced into what we in 2011 call Germany. The national bibliographic profiles of the current Balkan nations may be completely corrupted by the shifting assignments of MARC country codes in the region. Furthermore, cataloging practice in the area of date of publication may vary, especially with reprintings of prominent historical works. Finally, none of the historical profiles below should be taken as an assertion of historical causality on national book publication and collection.

## Findings and Discussion

### Book Publishing Profiles: Basic Publication Data

The increasingly global and comprehensive nature of the WorldCat database should yield ample and relatively trustworthy representations of the six national bibliographic profiles being considered. In addition, the six basic historical publication profiles offer remarkably distinct and lucid images of each country's evolving publishing history and how books and library collections tend to reflect the history of each.

For the six case study countries, table 1 shows the scope of the basic dataset, with the ongoing caveat that these are the records represented in WorldCat. This gives an idea of the richness of the data possible to mine in WorldCat, country by country. German publications, not unsurprisingly, are the richest subset of these data; the records from Germany include, pertinently, the catalogs of the Deutsche Nationalbibliothek (DNB), the Bayerische Staatsbibliothek, and Hessische Bibliotheksinformationssystem (HeBIS), the consortium of the largest university libraries in the Federal Republic.<sup>20</sup> In addition, this country has spearheaded national efforts at bibliographic control under the leadership of the DNB.<sup>21</sup> South Africa has a relatively high rate of holdings for its national publications, perhaps because of the participation of the University of Cape Town and the University of South Africa (Unisa) as cataloging partners in WorldCat.<sup>22</sup> Of the six countries in this study, Bolivia has the fewest records in WorldCat; one can presume that their national bibliography is larger than 58,000 titles, but the lesser participation of Bolivian libraries in WorldCat to date somewhat weakens the publishing profile presented here.

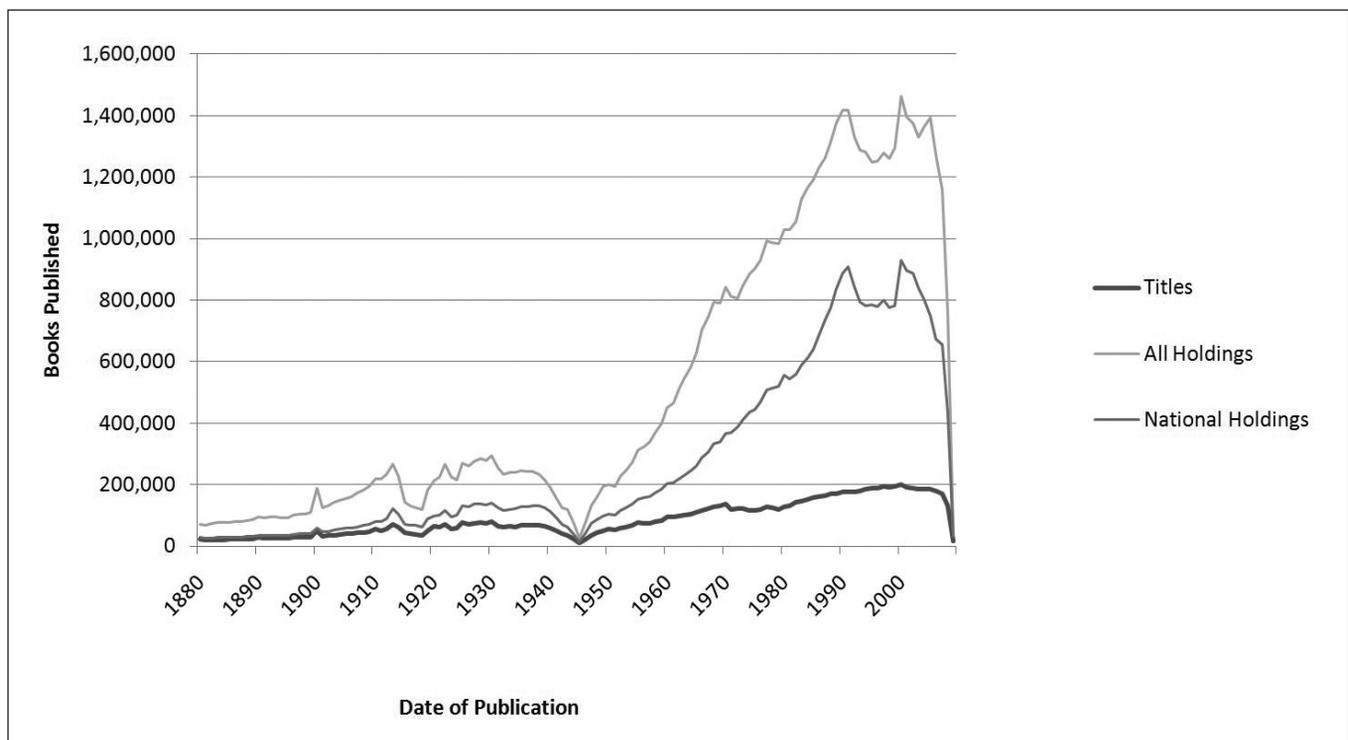
Table 2 provides a comparative sample of several other countries' datasets within WorldCat. It appears that for many countries (depending somewhat, but not entirely, on libraries' contributions), the data mining will be able to profile a large set of publications throughout their history. Even in the case of a problematic country like Ukraine (a former Soviet republic), enough catalogers participating in WorldCat even during the Cold War have paid attention to specifying that Soviet publications from Kiev should receive the country code for the Ukrainian Soviet Socialist Republic (unr) rather than the blanket code for the Soviet Union (ur). Thus Ukraine's bibliographic heritage can remain separable from the Soviet Union in the bibliographic data. In the case of the national bibliography as seen in WorldCat for China, the data was mined in August of 2009 and will certainly offer a much more complete picture once bibliographic data from the National Library of China are fully integrated into the database.<sup>23</sup> The example of China highlights the importance of national libraries' contribution to the aggregate bibliographic data available for data mining in projects

**Table 1.** The Six Case Studies

Country	OCLC Member Libraries	Publications	Holdings	Translations
Bolivia	1	58,838	302,309	780
Chile	387	265,948	900,330	7,351
Germany	378	12,843,605	70,341,725	381,141
Poland	27	1,225,446	3,108,677	174,738
South Africa	1,082	299,574	2,014,327	10,063
Thailand	39	181,003	463,104	5,700

**Table 2.** A Sample of Other National Bibliographies

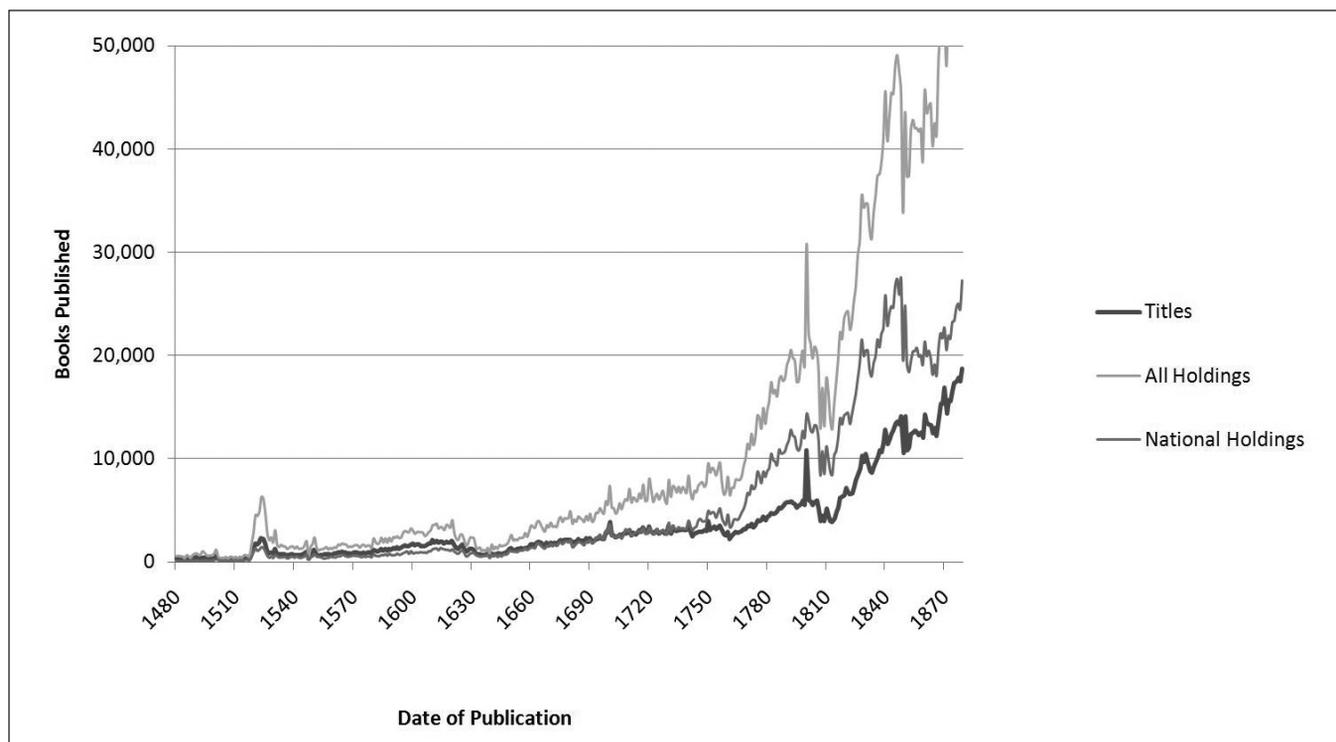
Country	OCLC Member Libraries	Publications	Holdings	Translations
China	1,160	3,208,114	8,160,062	238,315
Finland	71	905,252	1,476,734	122,150
France	253	4,948,620	29,127,570	211,341
India	37	1,121,311	6,636,530	52,107
Russia	45	1,883,418	8,151,097	94,319
Ukraine	3	236,166	710,372	9,061

**Figure 1.** Book Publication in Germany 1880–2010, as Reflected in WorldCat

of this type: whereas acquisition of a country's publications by American research libraries can offer the beginnings of a robust dataset, the bibliographic universe represented in WorldCat, Google, or any such data source is highly fortified by including the contents of national libraries.

A graphical representation of the data *within* a subset can at times be even more revealing. The robust data on book publication and library collection from Germany in the twentieth and twenty-first centuries (figure 1) reflect the historical dips in book publication from 1914 to 1919 and the complete collapse of the industry in 1945 and 1946. (No necessary historical correlation is being asserted, but rather the overlap of historical events and changes in the publication profile as

shown in the data mining.) After German unification in the 1990s, German book publication itself may not have waned, but the merging of libraries and university systems within the now-unified nation may have led to fewer copies being held (thus being represented graphically in a parallel dip in both national and international holdings of German books). From a peak of more than 900,000 indigenous library copies of German publications in 1991 (the bulk in these data of almost 1.5 million copies worldwide), by the middle of the 1990s, the German libraries' holdings of their own publications had dipped to around 780,000. The trough in "All Holdings" worldwide in figure 1 corresponds to the apparent culling of duplicate copies in German libraries.



**Figure 2.** Book Publication 1480–1879 in Germany, as Reflected in WorldCat

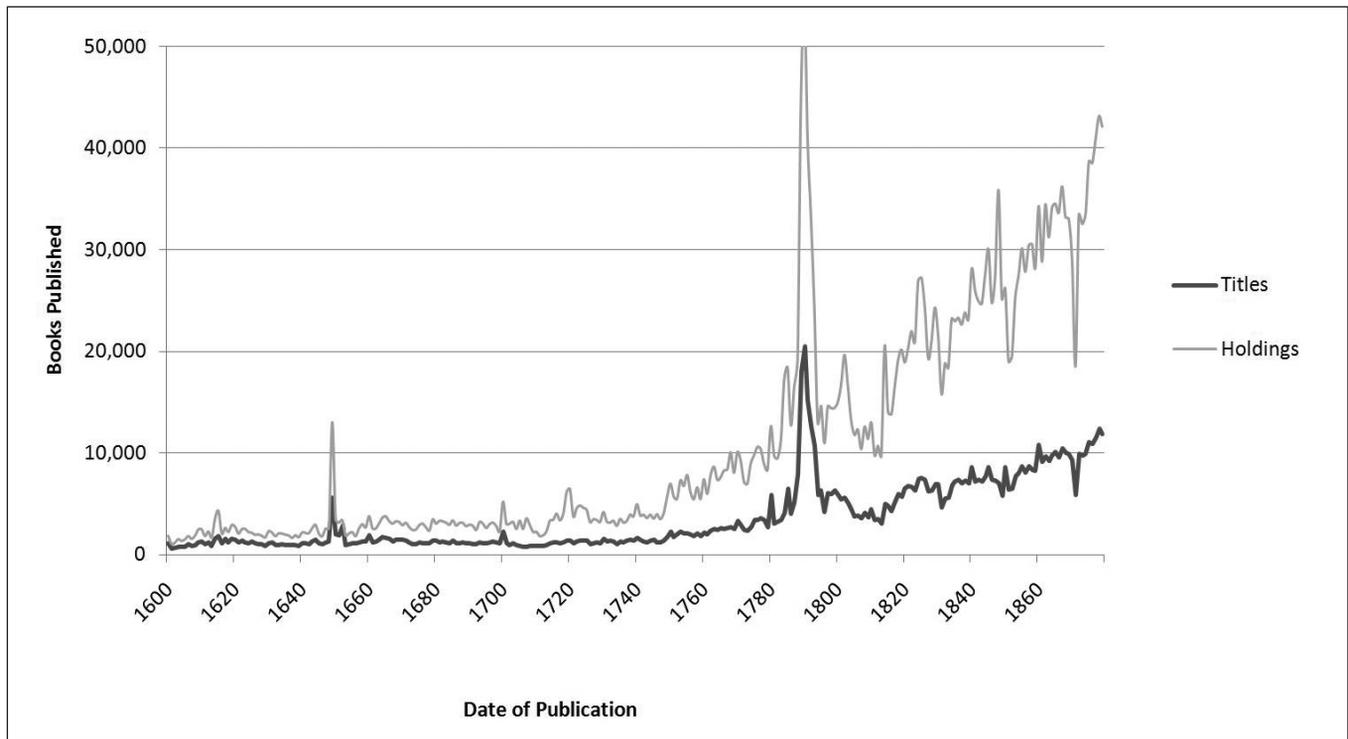
The robust German dataset within WorldCat similarly documents some interesting overlaps with events earlier in the history of the German-speaking lands. In this earlier dataset (see figure 2), one can note several changes in recorded book publication that correspond to important historical developments. These include dips in both book printing and holdings during the Napoleonic Wars (1800s to the early 1810s); publications peaked at more than 4,500 in 1808 (with nearly 17,000 holdings) but would not return to similar numbers until 1815. During the revolutionary period and into the 1840s, total publications surged from around 9,000 books to 14,093 in the year 1848 (with a corresponding growth in libraries holding copies), immediately dropping nearly a third to 10,550 in the year after the tumultuous the 1848 revolutions. The WorldCat data even present a smaller but palpable falloff in book publishing activity much earlier, during the Thirty Years' War: a publication rate of almost 2,000 books a year from 1617 to 1620 falls to only 974 in 1627, to an abysmal 517 in 1639, and no real recovery until well after the war's end in 1648.

German data also show publication rates that may correspond to positive historical effects on the national literary scene. During the reign of Bismarck, the rates of publication within German-speaking lands surge from 14,404 in 1871 (Bismarck's ascension to Imperial Chancellor) to 28,440 in 1890 (the year of his resignation). The German publication

data even experience an early peak during the Protestant Reformation that began in 1517. Only 327 publications survive from 1517, but that leaps immediately to 777 in 1518 and crests at 2,341 in 1523. This peak also could reflect the historical importance of Reformation materials, including the explosion of warring theological pamphlets, which libraries would tend to heavily collect and preserve.<sup>24</sup>

The unique nature of the German data also is revealed in comparison to the historical data from France (figure 3). The French data also contain a drop in publication rates in the revolutionary year of 1848, from 7,262 publications in 1847 to 5,815 in 1849. Even more pertinently, in 1871, the year the Prussians occupied Paris, the publication record crashes from 9,269 in the previous year to fewer than 6,000. The years 1789 and 1790, around the time of the French Revolution, see the most dramatic spike in both French national print publication and, perhaps tellingly, in libraries' collection and preservation of materials, visible in figure 3 as a quadrupling of printed publications collected by libraries to more than 20,000 titles.

Alternatively, one can compare the twentieth-century data from Germany (figure 1) to the twentieth-century data from neighboring Poland (figure 4). The Polish data, as might be expected, do show a slump during their occupation by Nazi and Soviet forces, as well as a surprising dip in general worldwide holdings in the late 1970s, followed by a surge



**Figure 3.** Book Publication in France 1600–1870, as Reflected in WorldCat

in publishing after the fall of Communism, a period that also corresponds to a very volatile period in the Polish publishing industry as the vibrant underground publishing community from the 1980s suddenly experienced the collapse of state censorship as well as turbulent economic conditions.<sup>25</sup> Also, one can note that the national holdings for Poland—holdings of Polish publications in Polish libraries—are in every year consistently lower than the total number of titles, an indicator of fewer data from libraries within Poland.

For a non-European country like South Africa, on the other hand, a comparable slump in publication activity during World War II would be less expected, and indeed it is less evident in the data. One major turning point in the South African data profile (see figure 5) instead follows the establishment of a national legal library deposit system in 1959.<sup>26</sup> South African publications, as reflected in the WorldCat data, exploded from 1,695 publications in 1959 to more than 9,500 in the year 1990. However, international pressure in the 1970s and 80s against Apartheid could be reflected in a tension between flatter growth in South African books within library collections outside of South Africa: over the same span of years, South African books discovered by this research in South African libraries increased more than tenfold, from 3,104 South African library holdings in 1959 to 41,234 in 1990, while non-South African library holdings of the same publications only increased from 7,000 to 20,000.

The numbers are even more striking in proportion to the overall surge in publication rate. In 1959, each book published in South Africa was collected on average by at least four other libraries worldwide; by 1990, the proportion was fewer than two libraries per publication.

#### Book Publishing Profiles: Language Data

Reliable data on language of publication (see appendix A) also emerged from most of the sample national datasets and in most cases reflected appropriate and expected differences between them. Although the dominant language of publication in books from Bolivia was expectedly Spanish, note the presence in WorldCat of books published in indigenous languages, such as Aymara, Quechua, and Guarani. The data from Chile can be treated with even greater trust knowing that the country enjoys a national shared library network.<sup>27</sup> English-language publishing is apparently stronger in Chile than in Bolivia, despite the lesser dependence of the Chilean data on libraries external to Chile and perhaps can be explained by the dominant strain of minority English speakers in Chile.

For the larger European countries, conversely, the book publishing seems to be more Eurocentric, with less emphasis on a single dominant national language, greater emphasis on publishing in other European languages, and

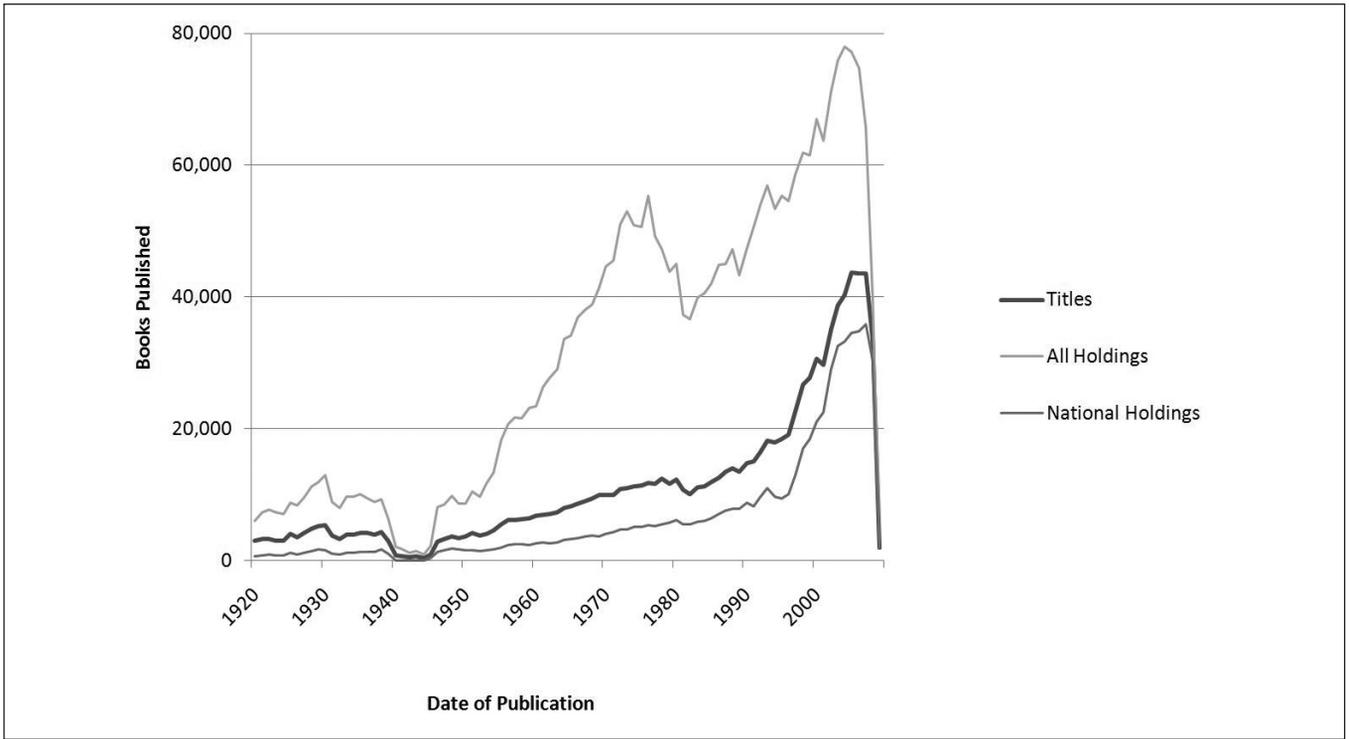


Figure 4. Book Publication in Poland 1920–2010, as Reflected in WorldCat

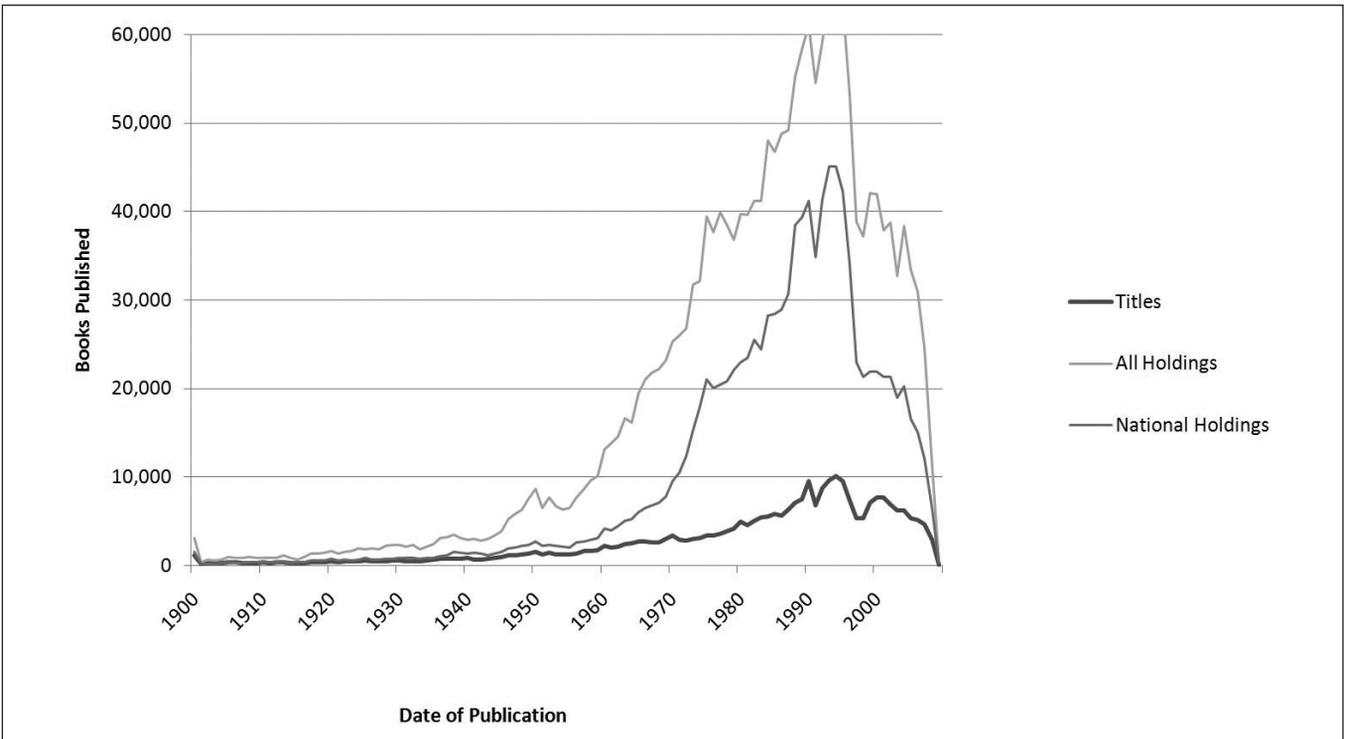
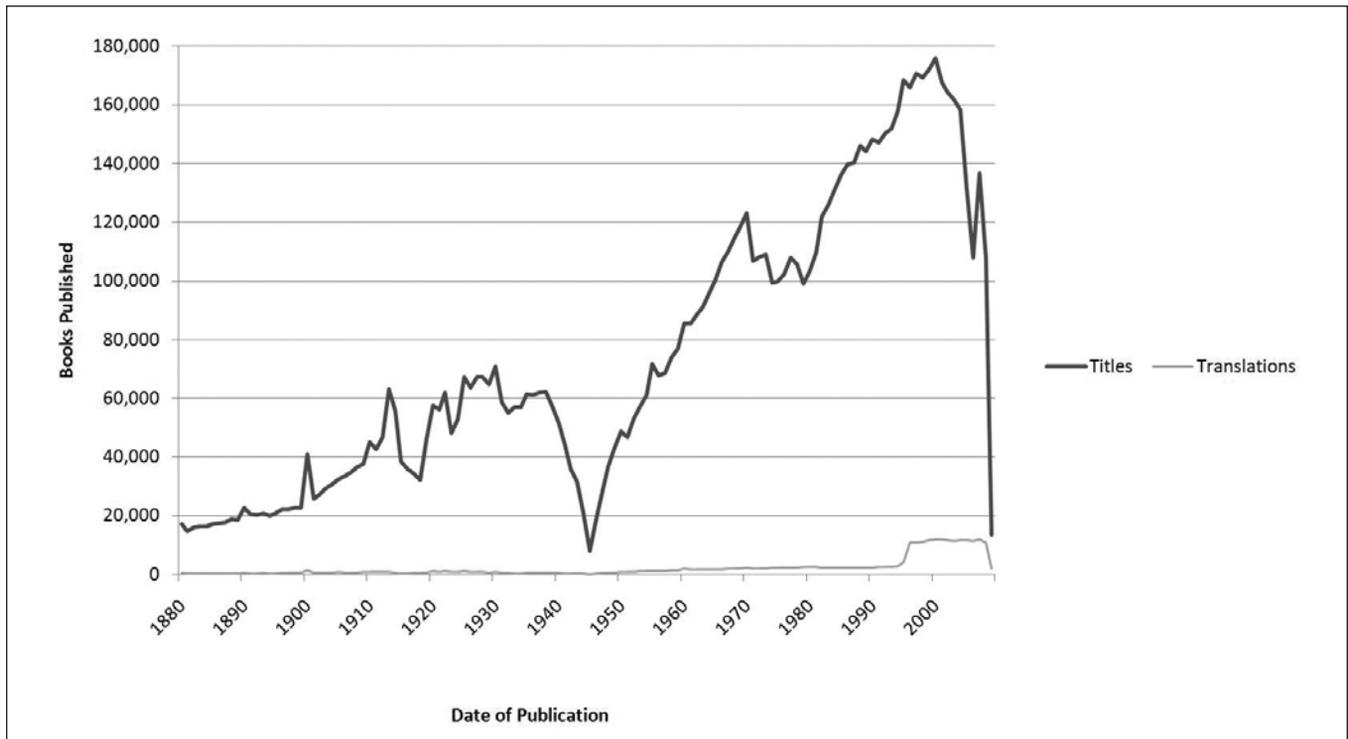


Figure 5. Book Publication in South Africa 1900–2010, as Reflected in WorldCat



**Figure 6.** German-Language Publishing in Germany 1880–2010, as Reflected in WorldCat

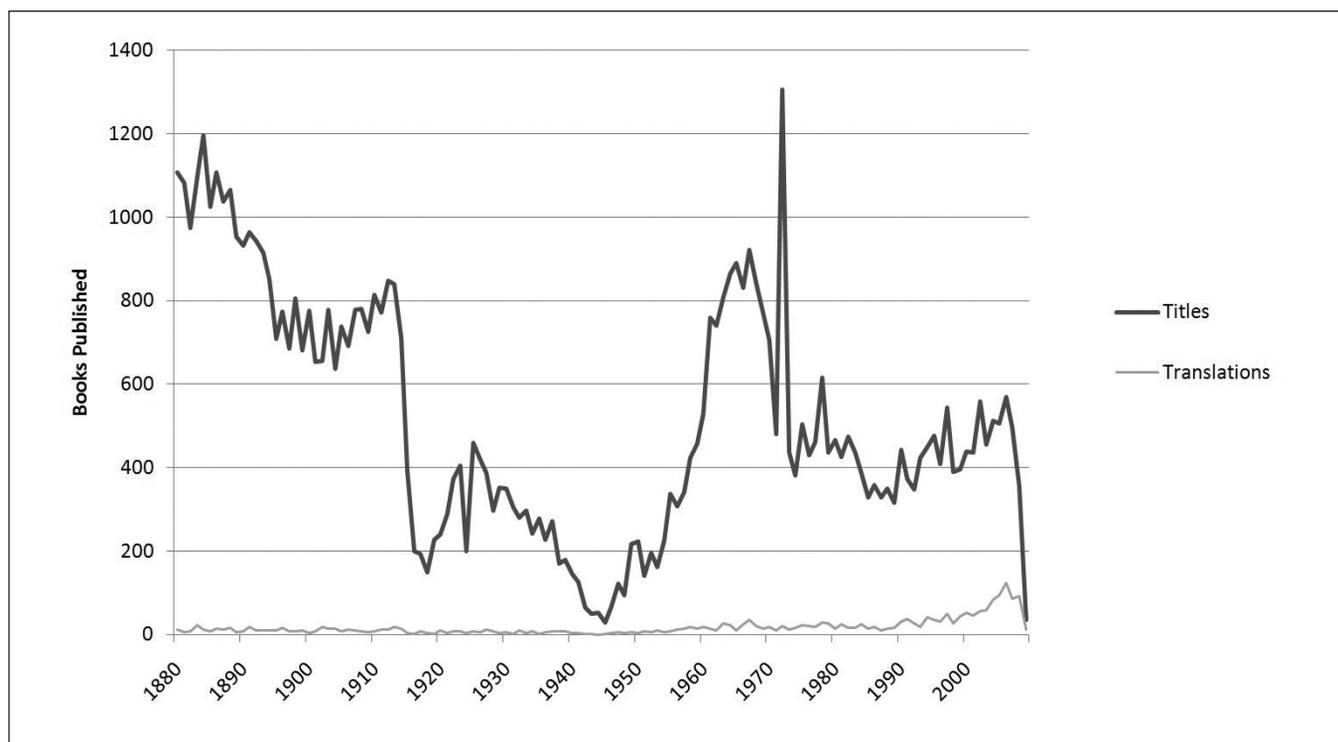
fewer appearances of minority but native languages such as Danish, Frisian, and Sorbian in the German publications, and Lithuanian in those from Poland. On the other hand, these data include a strong presence of books published in Latin in the data from Germany and Poland—a concrete reflection of Western libraries' function as historical memory institutions. Not shown in appendix A but definitely present in the Polish and German data are works in Greek, Hebrew, and historical languages such as Low German and Middle High and Old High German, again reinforcing the function of the library collection in preserving these aspects of the historical culture. The data from South African book publishing, on the other hand, appropriately reflect the more complex linguistic heritage of that country—two dominant languages of the ex-colonial powers (and likely founders of much of the South African publishing industry) with a very healthy dose of both translations into, and works original to, a variety of indigenous languages.

The language data also can be parsed out by year across the historical span of a dataset. German-language publications in Germany, for instance (see figure 6), display interesting nodes around a spike in German-language publications in 1913 (leaping to more than 60,000 publications in the national language) as the country was gearing up for what would become the First World War, and a surprising dip through the 1970s. Compare that to the very different

graphical shape of Germany's Latin-language publications (figure 7). Germany's Latin-language publishing first spiked in 1517 (with 245 Latin-language publications collected and preserved by libraries at the outset of the Reformation and appearing in WorldCat), and remained generally strong (around 1,000 titles per year published in Latin) up to the nineteenth century. At the start of the twentieth century, however, Latin-language publishing in Germany fell off to numbers around 800 titles per year; in the 1930s, during the Nazi era, this fell even more dramatically, and 1968 (the year the Second Vatican Council liberalized many Roman Catholic practices) ushered in another decline in Latin-language publication. (The spike in the year 1972 could be an outlier data point within the trajectory of the data; a random sample of Latin-language publications from Germany in 1972 revealed no explanation for such an increase in Latin-language publication in this year.)

#### Book Publishing Profiles: Translation Data

Data mining also was able to reveal patterns in book translations (see appendix B). Such data, allowing a view of the interactions between languages used by a culture, is among the most valuable to UNESCO as it seeks to measure cultural statistics. Once again, the data for Poland and Germany reflect a more Eurocentric vision, with translations from a



**Figure 7.** Latin-Language Publishing in Germany 1880–2010, as Reflected in WorldCat

variety of major European languages into German and Polish responsible for the majority of the translations; although in these data as well, translations into English from German and Polish also figure prominently. Latin and Greek translations also are represented in the WorldCat data from these two countries. The data on these translations were, however, somewhat less prominent. Greek to German translations were the tenth most prevalent, at 6,178 books; in Poland, Latin to Polish translations (2,687 books) were somewhat less common than Polish to German, and Greek to Polish lagged behind, although still recording 1,072 translations. The data on translations for Bolivia, Chile, and Thailand give a similar picture of both the predominant languages spoken in each country (Spanish and Thai), as well as the interaction of English and other languages with these dominant languages.

The data for the more culturally diverse country of South Africa, on the other hand, presents a much more varied tapestry of translation data, including translations between English, German, Afrikaans, and a number of tribal languages. Even more interesting historical patterns are revealed by plotting these data on a similar historical graph (figure 8). Whereas book publications from South Africa (figure 5) were more avidly collected after the 1959 deposit library system, the dominant type of translated works during the early decades of the Republic (1961 until the mid-1980s) are translations between English and Afrikaans,

varying between around 25 and 150 per year. After the Zulu and Xhosa languages were decriminalized in 1991, however, translations from them enjoy a surge to more than 50 translations per year.

A different set of shifting cultural interactions may be viewed in the historical patterns of translated works from Poland (figure 9). The most overt pattern is the large increase in translations from English, French, and German after the fall of Communism; English to Polish translations, particularly, rocketed from fewer than 400 in 1989 to more than 6,000, literally off the chart, after 2000. But from the 1950s to the 1980s, the most important pair of languages tended to be Russian and Polish. In a lesser-populated dataset, such as the publication and collection profile from Thailand (figure 10), the spiky nature of the graph suggests that data mining in this instance is nearing the functional end of its reliability. The Thai data on translations also are the only instance in which translations into English lead the statistics (as reflected in WorldCat).

### Future Research

These findings suggest several avenues for further research. First, data mining that compares copies of a country's national bibliography in foreign libraries to the holdings of

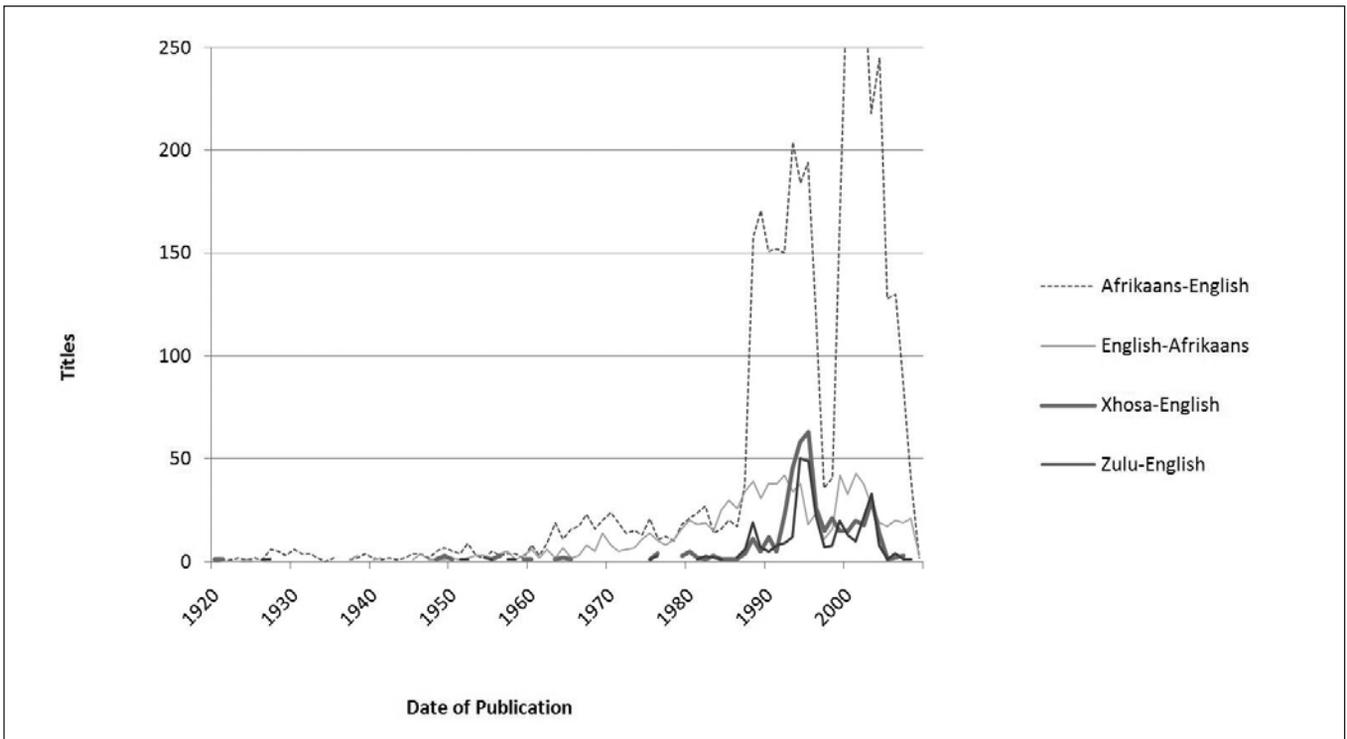


Figure 8. Translated Work from South Africa 1920–2010, as Reflected in WorldCat

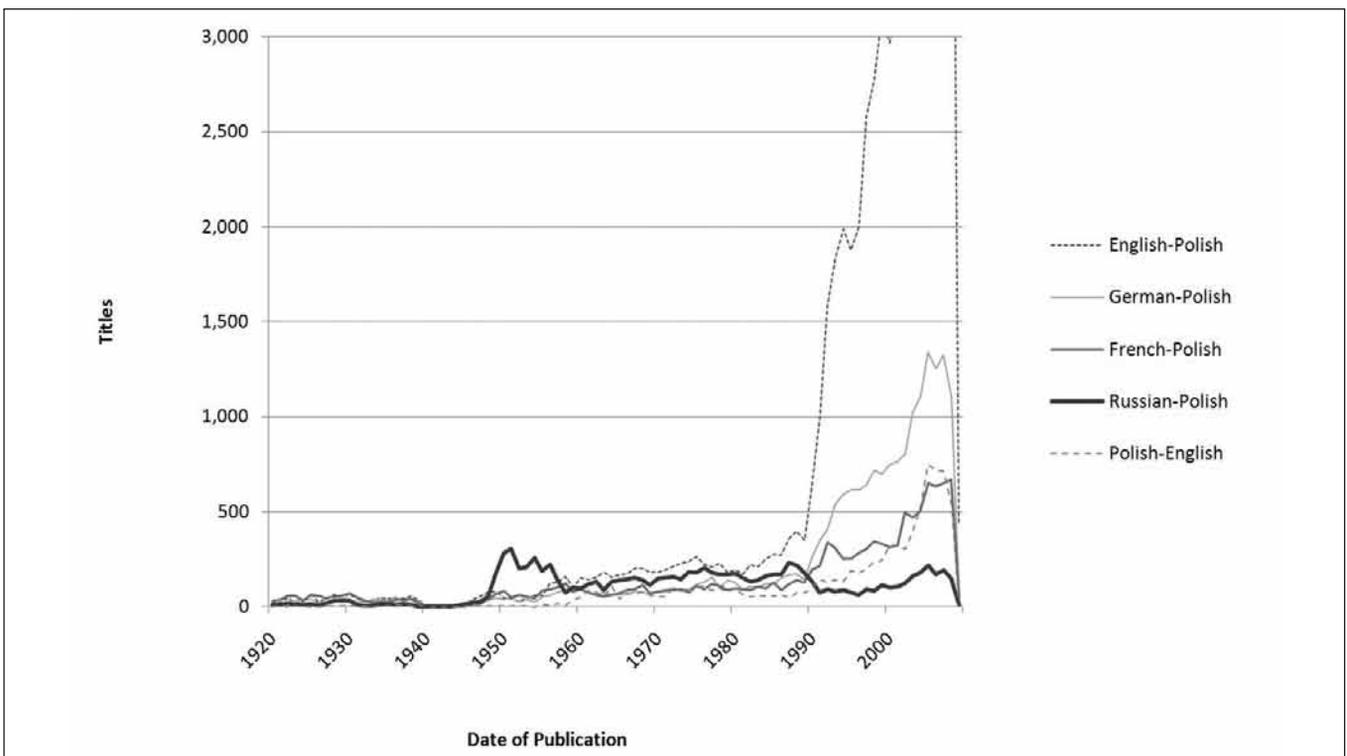
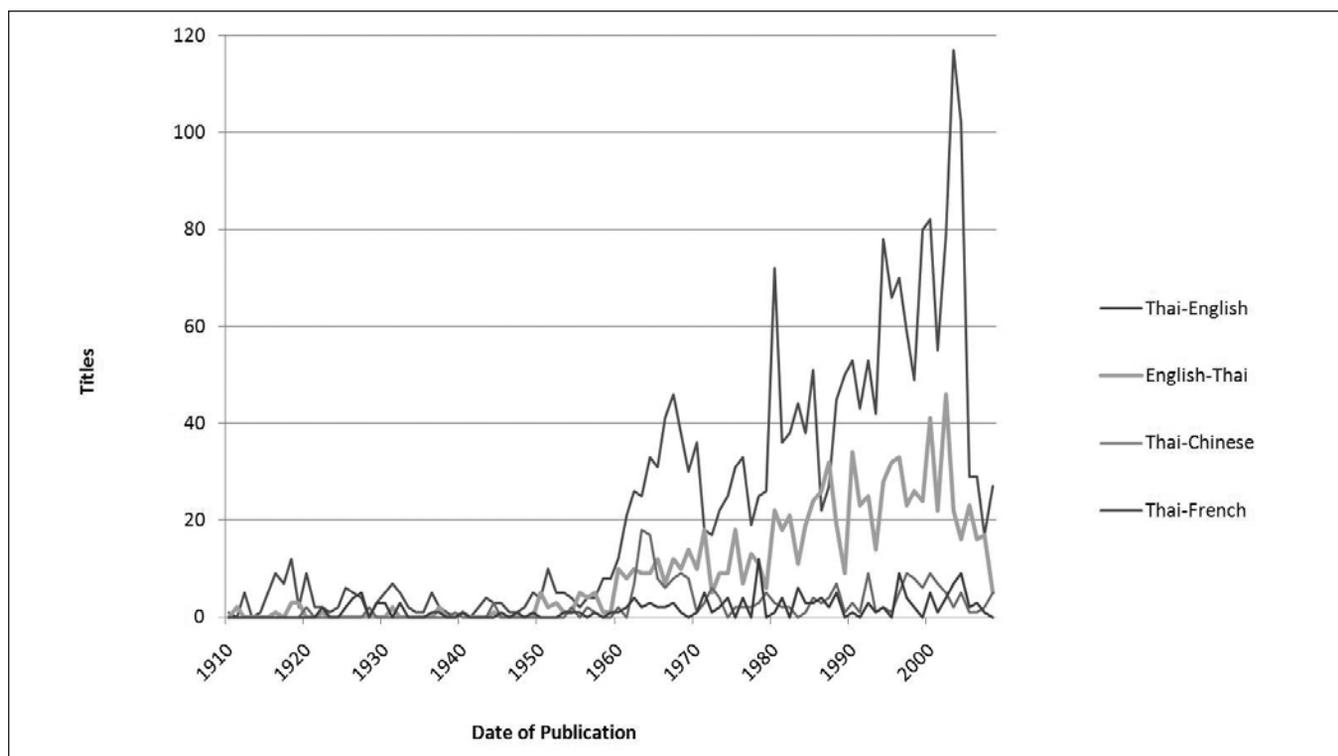


Figure 9. Translated Work from Poland 1920–2010, as Reflected in WorldCat



**Figure 10.** Translated Work from Thailand 1910–2010, as Reflected in WorldCat

libraries within the country should provide a different take on cultural production. How often do the books published as a reflection of the country's culture travel to other countries and end up in the collections of other libraries? What language materials and translations are being collected by others? Second, the potential exists to conduct longitudinal studies on books as an expression of cultural diversity: rechecking these global data every few years (as more libraries join WorldCat) to seek better data and trends of globalization in book collecting. Such longitudinal comparison also could offer UNESCO valuable evidence on progress toward its goals of preserving global linguistic diversity. Finally, as each country's national bibliographies are added more widely into the aggregate collection, the type of national publication profiles explored here can empower, via a data source, specialists to interpret historical trends in a country's literatures or publication history.

### Conclusions

The research presented above began as an attempt to compare national publication profiles for countries around the world. Researchers used the techniques of data mining within the aggregate bibliographic database, WorldCat, to extract data for collection analysis, with emphasis on

differentiation by country and region. The project successfully extracted data from each country (as their publication record is reflected in WorldCat) and parsed it according to publication patterns, publication languages, and data on translated works in those publication patterns. To demonstrate the validity of this approach, data from six countries were compared in some detail in this paper. Despite the limitations of data mining in a database with roots in Anglo-American traditions, the six countries' case studies appear robust and are sharply delineated from one another.

The data on book publication and library collections from each country, and the evidence of different patterns in language use and translations, offers different portraits of the literary arts in each country. Specifically, these patterns differ in ways that correlate to each country's historical trajectory. Different experiences during the Second World War (as well as earlier conflicts) correspond to different patterns in book production; political movements as different as German nationalism and the end of Apartheid in South Africa produce different patterns of linguistic publication; even global religious upheavals, such as the Protestant Reformation and the Second Vatican Council, may affect the data on book publication. In the countries whose publishing footprint in the WorldCat database is smaller, the data on book translations are weakest, yet even they offer strong indications of the interactions between languages

one might expect. In the case of a more complete national bibliography, as is likely represented by the data mined for Germany, even events deep in the country's history coincide with changes in the mined publication data and the presence of dead languages. This testifies to the importance of library collections as custodians of historical culture. These six case studies present a brief glimpse into the richness of information about library collections that can be mined from the catalog data already available.

### References and Notes

1. See, for example, UNESCO, Culture, Cultural Diversity, [http://portal.unesco.org/culture/en/ev.php-URL\\_ID=34321&URL\\_DO=DO\\_TOPIC&URL\\_SECTION=201.html](http://portal.unesco.org/culture/en/ev.php-URL_ID=34321&URL_DO=DO_TOPIC&URL_SECTION=201.html) (accessed Nov. 8, 2010). The European Expert Meeting took place on Nov. 7, 2008, as a collaborative effort of publishers, librarians, and booksellers, and as part of Amsterdam's work to become labeled as World Book Capital City.
2. OCLC, WorldCat Facts and Statistics, [www.oclc.org/us/en/worldcat/statistics/default.htm](http://www.oclc.org/us/en/worldcat/statistics/default.htm) (accessed Nov. 8, 2010).
3. OCLC Research, Books as an Expression of Cultural Diversity, [www.oclc.org/research/activities/globalbooks/default.htm](http://www.oclc.org/research/activities/globalbooks/default.htm) (accessed Nov. 8, 2010).
4. UNESCO, *Convention on the Protection and Promotion of the Diversity of Cultural Expressions: Paris 20 October 2005* (London: Stationer's Office, 2007), [www.unesco.org/new/en/unesco/themes/2005-convention/the-convention/convention-text](http://www.unesco.org/new/en/unesco/themes/2005-convention/the-convention/convention-text) (accessed Nov. 8, 2010).
5. UNESCO, Programme—Section of Creative Industries for Development, Major Programme 2010–2011 (Project 35 C/5) IV, Culture, [http://portal.unesco.org/culture/en/ev.php-URL\\_ID=35864&URL\\_DO=DO\\_TOPIC&URL\\_SECTION=201.html](http://portal.unesco.org/culture/en/ev.php-URL_ID=35864&URL_DO=DO_TOPIC&URL_SECTION=201.html), (accessed Nov. 8, 2010).
6. Michael Heaney, e-mail to the author, Dec. 4, 2008.
7. IFLA Statistics and Evaluation Section, "Minutes of the First Standing Committee Meeting" (Milan, Italy, Aug. 22, 2009), [www.ifla.org/files/statistics-and-evaluation/minutes/august-2009.pdf](http://www.ifla.org/files/statistics-and-evaluation/minutes/august-2009.pdf), (accessed Jan. 22, 2011).
8. UNESCO, *Index Translationum: World Bibliography of Translation* (Paris: UNESCO, 1932–), [http://portal.unesco.org/culture/en/ev.php-URL\\_ID=7810&URL\\_DO=DO\\_TOPIC&URL\\_SECTION=201.html](http://portal.unesco.org/culture/en/ev.php-URL_ID=7810&URL_DO=DO_TOPIC&URL_SECTION=201.html) (accessed Nov. 8, 2010).
9. Heaney, e-mail to the author.
10. Liz McKeen, "Canadiana: The National Bibliography for Canada, in the Digital Age," *International Cataloging & Bibliographic Control* 38, no. 2 (Apr. 2009): 19–22; Yeon-Kyoung Chung, "National Bibliographies Past, Present, and Future: The Korean Experience," presentation, World Library and Information Congress: 72nd IFLA General Conference and Council, August 2004, 2006, Seoul, Korea, <http://archive.ifla.org/IV/ifla72/papers/109-Chung-en.pdf> (accessed April 14, 2011): 1–16; Daniel M. Pennell, "The Fate of Book Chambers National Bibliographies in Belarus, Ukraine, and Moldova since 1991," *Slavic & East European Information Resources* 11, no. 1 (Jan. 2010): 10–20; Anna Katuna Chelidze and Janet Zmroczek, "National Bibliography of Georgia: Past, Present, and Future," *Slavic & East European Information Resources* 11, no. 1 (Jan. 2010): 41–45; Heghine Hakobyan, "National Bibliography in Armenia," *Slavic & East European Information Resources* 11, no. 1 (Jan. 2010): 46–53; Ibrhaim Ramjuan, "National Bibliographic Control in Mauritius: Issues and Challenges," *Information Development* 25, no. 4 (Nov. 2009): 296–303.
11. OCLC Research, WorldMap, [www.oclc.org/research/activities/worldmap/default.htm](http://www.oclc.org/research/activities/worldmap/default.htm) (accessed Nov. 8, 2010). See also OCLC, Global Library Statistics, [www.oclc.org/global-librarystats/default.htm](http://www.oclc.org/global-librarystats/default.htm) (accessed Jan. 22, 2011).
12. A full list of the reference sources consulted for the data in the OCLC WorldMap may be found at OCLC, Global Library Statistics, [www.oclc.org/global-librarystats/sources.htm](http://www.oclc.org/global-librarystats/sources.htm) (accessed Jan. 22, 2011).
13. See OCLC Research, Data Mining Research Area, [www.oclc.org/research/activities/past/orprojects/mining/default.htm](http://www.oclc.org/research/activities/past/orprojects/mining/default.htm) (accessed Nov. 8, 2010), for a partial bibliography of studies.
14. Brian Lavoie, Lynn Silipigni Connaway, and Edward T. O'Neill, "Mapping WorldCat's Digital Landscape," *Library Resources & Technical Services* 51, no. 2 (2007): 106–15.
15. Lynn Silipigni Connaway and Timothy J. Dickey, "Beyond Data Mining: Delivering the Next Generation of Service from Library Data," part of the panel presentation "Transforming Data into Services: Delivering the Next Generation of User-Oriented Collections and Services," *Proceedings of the American Society for Information Science & Technology* 45, no. 1 (2008): 1062.
16. OCLC, WorldCat Facts and Statistics; Jay Jordan, "OCLC Update Breakfast," presentation, American Library Association Annual Conference, Washington D.C., June 27, 2010, <http://vidego.multicastmedia.com/player.php?p=mal7k7dn> (accessed Nov. 8, 2010).
17. Benjamin Lee Whorf, "Thinking in Primitive Communities," in *Language, Thought, and Reality*, ed. John B. Carroll, 65–87 (Cambridge, Mass.: MIT Pr., 1964).
18. For a summary of the basic tenets, see Harry Hoijer, "The Sapir-Whorf Hypothesis," in *Language in Culture: Conference on the Interrelations of Language and Other Aspects of Culture*, ed. Harry Hoijer, 92–105 (Chicago: Univ. of Chicago Pr., 1954); for later challenges, see George Lakoff, *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind* (Chicago: Univ. of Chicago Pr., 1987); and John J. Gumperz and Steven C. Levinson, eds., *Rethinking Linguistic Relativity* (Cambridge, England: Cambridge Univ. Pr., 1996).
19. A special algorithm was developed to interpret the various ways multiple linguistic context has been encoded in that field's subfield \$a and \$h, which are both repeatable and which underwent a change in MARC coding instructions in 1981; this necessitated also mining the record creation date from each record.
20. OCLC, Brief History of OCLC Activities with National Libraries outside the U.S., [www.oclc.org/us/en/worldcat/catalog/national/timeline/default.htm](http://www.oclc.org/us/en/worldcat/catalog/national/timeline/default.htm) (accessed Nov. 8, 2010).
21. Claudia Fabian, Bibliographic Control in Germany, *Bollettino AB* 44, no. 1 (Mar. 2004): 18.
22. Ina Fourie and Marlene Burger, "Bibliographic Control in South Africa," *International Cataloging & Bibliographic*

- Control* 37, no.1 (Jan. 2008): 3–8.
23. For a timeline of national libraries' participation in WorldCat, to date involving forty-one countries worldwide, see OCLC, Brief History of OCLC Activities with National Libraries Outside the U.S., [www.oclc.org/us/en/worldcat/catalog/national/timeline/default.htm](http://www.oclc.org/us/en/worldcat/catalog/national/timeline/default.htm) (accessed Nov. 8, 2010).
24. For a recent historical view of this era of post-Reformation printing history, see Andrew Pettegree, *The Book in the Renaissance* (New Haven, Conn.: Yale University Press, 2010).
25. Izabella Tomljanovich, "Recent Publishing Trends and Developments in Poland," *Slavic & East European Information Resources* 1, no. 1 (2000): 83–96.
26. Peter Lor, "From a Trickle to a Torrent: Bibliographic Control of Books in South Africa, 1796 to 1996," *Mousaion* 23, no. 1 (2005): 19–38.
27. Elizabeth N. Steinhagen, "Leaders of Bibliographic Control: The Chilean Experience," *Cataloging & Classification Quarterly* 31, no. 1 (2000): 3–14.

### Appendix A. Language Data from the Six Countries

Country	Publications	Language	Titles	Translations	Language Distribution (%)
Bolivia	58,838	Spanish (Official)*	56,326	529	95.73
		Quechua (Official)	202	46	0.34
		Aymara (Official)	228	28	0.39
		Guarani	28	10	0.05
		Other Amerindian (English)	269	64	0.46
			887	69	1.51
Chile	265,948	Spanish (Official)	248,935	6,505	93.60
		English	10,965	533	4.12
		German	772	81	0.29
		Mapudungun (French)	127	33	0.05
			935	55	0.35
Germany	12,843,605	German	10,558,895	317,448	82.21
		Danish	1,712	320	0.01
		Frisian	401	51	0.00
		Sorbian	83	22	0.00
		(English)	580,149	29,073	4.52
		(Latin)	378,698	8,041	2.95
		(French)	92,516	6,919	0.72
Poland	1,225,446	Polish	917,229	150,061	74.85
		German	85,304	5,041	6.96
		Ukrainian	2,105	282	0.17
		Lithuanian	327	81	0.03
		Belarusian	691	92	0.06
		(English)	63,137	10,916	5.15
		(Latin)	23,877	882	1.95
		(Russian)	9,927	912	0.81
South Africa	299,574	English	193,504	2,082	64.59
		Afrikaans	77,449	5,145	25.85
		Zulu	2,649	487	0.88
		Xhosa	2,364	553	0.79
		Sesotho	1,368	298	0.46
		Sepedi	1,344	177	0.45
		Setswana	1,290	154	0.43
		Xitsonga	729	132	0.24
		Venda	670	115	0.22
		Ndebele	167	31	0.06
Thailand	181,003	Thai	129,461	3,806	71.52
		English	44,233	1,514	24.44
		Chinese	1,224	65	0.68
		(French)	991	91	0.55
		(Pali)	920	47	0.51

\* Languages whose names are in parentheses are not technically native to the country.

### Appendix B. Translation Data from the Six Countries

Country	Publications	Translations	Original	Translation	Number	Language Distribution (%)
Bolivia	58,838	780	English	Spanish	211	27.05
			French	Spanish	61	7.82
			Spanish	English	58	7.44
			German	Spanish	50	6.41
			Aymara	Spanish	32	4.10
Chile	265,948	7,351	English	Spanish	2,571	34.97
			French	Spanish	1,365	18.57
			German	Spanish	575	7.82
			Spanish	English	442	6.01
			Italian	Spanish	410	5.58
Germany	12,843,605	381,141	English	German	149,318	39.18
			French	German	41,076	10.78
			Russian	German	14,254	3.74
			German	English	19,861	5.21
			Latin	German	12,811	3.36
			Italian	German	12,103	3.18
			Swedish	German	12,006	3.15
Poland	1,225,446	174,738	English	Polish	72,638	41.57
			German	Polish	20,626	11.80
			French	Polish	13,778	7.88
			Russian	Polish	9,828	5.62
			Polish	English	8,673	4.96
South Africa	299,574	10,063	English	Afrikaans	4,077	40.51
			Afrikaans	English	1,008	10.02
			English	Xhosa	453	4.50
			English	Zulu	357	3.55
			Dutch	Afrikaans	292	2.90
			English	Sesotho	219	2.18
			German	Afrikaans	209	2.08
			English	Tswana	204	2.03
			German	English	165	1.64
			English	Sepedi	109	1.08
			English	Venda	80	0.79
			Zulu	English	54	0.54
			Afrikaans	Zulu	40	0.40
Xhosa	English	31	0.31			
Thailand	181,003	5,700	Thai	English	2,322	40.74
			English	Thai	983	17.25
			Chinese	Thai	237	4.16
			Pali	Thai	178	3.12
			French	Thai	165	2.89

# Notes on Operations

## PCC Training for Copy Catalogers

### Is It Worth the Investment? The Columbia University Libraries Experience

Natalia Gelber and Irina Kandarasheva

*This paper presents a case study program implemented at Columbia University Libraries that trains copy catalogers in Program for Cooperative Cataloging (PCC) and Name Authority Cooperative Program (NACO) principles. Under the guidance of the NACO coordinator and cataloging supervisors, copy catalogers create PCC-level records for belles lettres materials, construct name and uniform title headings, and submit classification number proposals for literary authors in a variety of languages in both roman and nonroman scripts. The benefits of the program include timely and efficient processing of materials, high-quality bibliographic records, better collocation of belles lettres, and upgrade of vendor or member records utilized by the shared cataloging community. The program serves as an important step in the professional development of copy catalogers and contributes to the growth of training and mentoring skills of professional catalogers. The skills acquired by copy catalogers can be used in future cataloging projects such as metadata creation and name disambiguation.*

The past two decades have witnessed a significant change in functions and daily operations of technical services departments in academic libraries. Rapid technological developments, shrinking budgets, merging of cataloging and acquisitions departments, and increased use of vendor records have contributed to reevaluation of workflows, reassignment of personnel, and creation of new in-house training programs. The current economic and information environment suggests that technical services will continue to experience these changes in the near future. As libraries continue to move from ownership-based to access-based models, technical services departments will face the challenge of providing access to new types of resources while still processing materials in more traditional formats.

These transformations will influence staffing changes and blend job responsibilities in technical services departments in academic libraries. The shift of professional tasks from professional librarians to paraprofessional staff in various aspects of technical services operations is a reality widely acknowledged by library literature and practitioners in the field.<sup>1</sup>

The area of cataloging in particular has much potential for training and developing copy catalogers' skills to perform more complex and motivating tasks. The research to date has largely focused on successful development of copy cataloging skills. The specifics of training paraprofessionals in original

**Natalia Gelber** (Natalia.Gelber@liu.edu) is Technical Services Librarian, Assistant Professor, B. David Schwartz Memorial Library, C.W. Post Campus, Long Island University, Brookville, New York, and **Irina Kandarasheva** (Ink2@columbia.edu) is Copy Cataloging Unit Librarian, Columbia University Libraries, New York.

Submitted December 15, 2010; tentatively accepted January 27, 2011, pending modest revision; revision submitted February 15, 2011, and accepted for publication March 16, 2011.

cataloging have rarely been a focus of professional literature.<sup>2</sup> In addition, the authors found no research describing the process of training copy catalogers in Name Authority Cooperative Program (NACO) of the Program for Cooperative Cataloging (PCC) principles, cataloging to PCC level, and cataloging of belles lettres by paraprofessionals.

The present study describes a successful program developed at Columbia University Libraries (CUL) in which the combination of copy catalogers' skills and abilities and the institutional training potential resulted in a rewarding outcome. This paper begins by reviewing existing literature on paraprofessionals' participation in original cataloging and principals of PCC and NACO. Then the authors discuss the training and ongoing workflows for paraprofessionals cataloging belles lettres material to PCC level, including creating authority records and submitting call number proposals to be included in the name authority record for a literary author. Finally, the authors argue that competencies obtained by paraprofessional staff in this program have the potential for application in a new emerging cataloging and metadata environment.

## Literature Review

### Paraprofessionals and Original Cataloging

Several studies have discussed the topic of paraprofessionals' involvement in original cataloging. In 1992, Benaud, reflecting on trends in cataloging departments, appealed to library managers with the suggestion to take better advantage of people's skills and abilities by actively involving paraprofessionals in original cataloging.<sup>3</sup> In a separate article, Younger argued for the same approach, stating that libraries need to recognize the fact that some professional tasks can

be performed by support staff.<sup>4</sup> She further underscored the fact that this practice would benefit library users. Nevins advised library supervisors to consider whether future hires in copy cataloging units would be involved in original cataloging.<sup>5</sup> She pointed out the importance of teaching cataloging theory and practice and use of cataloging applications to new paraprofessionals. Her list of initial qualifications for copy cataloging positions includes foreign language proficiency, basic knowledge of cataloging standards, and solid general educational background.

Citing changes in the copy cataloging workflows triggered by development of vendor and contract cataloging services and the availability of better cataloging copy in WorldCat, Rider encouraged greater involvement of paraprofessionals in original cataloging.<sup>6</sup> In particular, she delineated a range of library materials that can benefit from in-house processing by support staff, such as foreign language monographs, rare books, and graphic and audiovisual materials. Rider elaborated on possible avenues for development of paraprofessional skills and suggested call number and subject heading assignments, original cataloging of literary works, and cataloging of variant editions as possible training areas. She mentioned solid problem-solving skills and extensive copy cataloging experience as prerequisites for the support staff participation and predicted that paraprofessionals would benefit from better professional opportunities as a result of the training.

Researchers actively used surveys to provide important insights into the role of paraprofessionals in original cataloging. Eskoz surveyed cataloging staff in academic libraries on two occasions and examined the level of cataloging delegated to paraprofessionals.<sup>7</sup> Eskoz's findings indicated a perceived tendency toward assignment of complex cataloging tasks to paraprofessionals. The level of participation of

paraprofessionals in the original cataloging was further investigated by Mohr and Schuneman.<sup>8</sup> Their survey data revealed that 77.1 percent of the respondents assigned original cataloging responsibilities to paraprofessionals. Mohr and Schuneman noted that the participation of support staff in original cataloging was most visible in the larger cataloging departments, confirming Eskoz's findings. Mohr and Schuneman highlighted various reasons why cataloging managers delegated professional tasks to support staff. Respondents cited the large volume of materials to be processed, the lack of professional catalogers, paraprofessionals' language and subject expertise, and cost-saving and support-staff professional development reasons. Buttlar and Garcha surveyed catalogers in academic libraries: 75.1 percent of the participants indicated a trend toward paraprofessionals' involvement in original cataloging.<sup>9</sup> Responding to the survey conducted by Bordienu and Seiser, 67 percent of academic research library administrators confirmed assigning professional duties to paraprofessionals.<sup>10</sup> Finally, Cox and Myers provided the most recent snapshot on the state of professionals' and paraprofessionals' participation in cataloging in Association of Research Libraries (ARL) cataloging departments.<sup>11</sup> The 2010 study confirmed the positive attitude of professional catalogers toward paraprofessionals' involvement in original cataloging. As library budgets decrease, support staff's participation in original cataloging helps reduce backlogs, makes possible the redeployment of professional catalogers to more complex work or service needs, and expands work skills of paraprofessionals.

Several authors have observed advantages, disadvantages, and certain concerns related to paraprofessionals' involvement in complex cataloging tasks. The advantages include better job outlook and opportunities for promotion, reduced backlogs

and increased statistics, and teaching opportunities for professional catalogers.<sup>12</sup> Time-consuming training processes, delegation of managerial duties to professional catalogers, and copy catalogers' lack of theoretical background are reasons for concern.<sup>13</sup> Adequate compensation and revision of job descriptions are additional important considerations related to assignment of original cataloging tasks to paraprofessionals.<sup>14</sup>

### Training Programs and Case Studies

In contrast to the number of studies focused on paraprofessionals' involvement in original cataloging, the body of literature concerned with the analysis of specific training programs is relatively modest. Robare described a formal training program developed by cataloging professionals at the University of Oregon to teach paraprofessionals to assign call numbers and subject headings to WorldCat member copy.<sup>15</sup> The study noted a number of beneficial outcomes for the library and highly skilled and enthusiastic copy catalogers, including the reduction of the library cataloging backlog, increased productivity, and a more positive outlook of paraprofessionals regarding the future of their positions.

El-Sherbini described the process of a technical services reorganization at Ohio State University and the subsequent changes in copy cataloger job functions.<sup>16</sup> Paraprofessionals underwent a training program developed by professional catalogers. After training, new duties assigned to paraprofessional staff included descriptive cataloging, subject analysis, call number assignment, original cataloging of literary works and variant editions, and authority record creation for the Library of Congress Name Authority File (LC NAF).<sup>17</sup> The program allowed professionals to dedicate more time to supervisory and training duties, eliminated backlogs, and created new challenges and work opportunities for

paraprofessionals. This cost-effective program resulted in improved productivity, a streamlined cataloging workflow, and better use staff's skills and abilities.

Training of paraprofessionals in subject analysis of theses and dissertations in Oregon State University was a focus of a case study conducted by Sapon-White.<sup>18</sup> The reduction of professional cataloging staff and the need to provide training to staff responsible for the description of digital resources were motivating factors in the program development. The study emphasized that original cataloging by paraprofessionals requires a significant investment of time for training and supervision. Sapon-White also drew attention to the fact that changing job responsibilities from copy to original cataloging involves an important transformation of the work model. He stated that "changing from one work paradigm to another necessitates retraining in both perspective and specific knowledge."<sup>19</sup> Paraprofessionals received ongoing training and support from professional catalogers and were given a certain degree of independence following the initial training. Professionals and paraprofessionals alike benefited from the teamwork environment, continued professional growth, and the development of mentor-mentee relationships.

### Paraprofessionals, NACO, and PCC

The quantitative measurement of paraprofessionals' contribution to NACO work was first mentioned by Mohr and Schuneman in their analyses of 1997 survey data. Precisely 57.8 percent of the responding libraries stated that paraprofessionals participated in the establishment of names and uniform titles headings.<sup>20</sup> The authors further speculated that well-documented rules and paraprofessionals' experience with authority work in copy cataloging were important factors contributing to this figure. The authority-work studies

by Wells, Wolverton, and Burke and Shorten mentioned the number of paraprofessional staff involved in authority work in a range from 42 to 78 percent, but did not provide specific data on how many support staff are trained in NACO principles and submit headings to NAF.<sup>21</sup>

In the case study of the NACO program in the University of Florida Libraries, Simpson and Williams addressed staff issues related to authority work. In particular, the authors mentioned that paraprofessionals' participation in NACO workflow in the past caused some concerns among professional catalogers regarding deprofessionalization.<sup>22</sup> The authors reported that after the duties of professional librarians were better defined, this attitude changed and senior level cataloging staff became actively involved in the creation of NACO headings. Simpson and Williams noted the creation of guidelines specifying participation of staff of all levels in the NACO program.

Describing the management of pre-professionals at Kent State University Libraries, Lisius and colleagues mentioned that "some experimentation has taken place training pre-professional catalogers to create name authority records (NARs), all under the review of a NACO cataloger."<sup>23</sup> In her historical overview of the Indiana Libraries experience with the PCC program, Charbonneau stated that both paraprofessionals and professionals have been trained in NACO principles.<sup>24</sup>

The "White Paper on PCC Role in Continuing Education for Catalogers," prepared for the PCC Policy Committee, recognized significant contributions made to original cataloging by paraprofessionals.<sup>25</sup> The paper cited results of the informal survey of Monographic Bibliographic Record Program (BIBCO) institutions, acknowledging that 73 percent of the libraries employ paraprofessionals in their BIBCO operations.

Arguing about the future of the PCC program itself, Banush stated in his recent article that “the [PCC] program no longer measures its success by the number of bibliographic records it produces each year, but instead by the extent of its outreach and training, its ability to influence standards development and application, and its expertise in automated systems.”<sup>26</sup>

## Background

CUL have a long standing history of training professional staff to catalog materials to PCC level. It started soon after the inception of the PCC in 1995. CUL were one of the early program partners, and many CUL staff participated in the three standing committees initially formed by the program. CUL have been a NACO library since 1989 and implemented series NACO in 1995. In 1996, CUL became a fully accredited BIBCO library, contributing name, series, and uniform title headings to the NAF and later participating in Subject Authority Cooperative Program (SACO) of the PCC (1996) and Cooperative Serials Program (CONSER) of the PCC (1997). Since then bibliographic and authority records have been produced in a variety of roman and nonroman languages, including Chinese, Japanese, and Korean. Numerous professional cataloging staff have been trained to create BIBCO records and name, series, conference, and uniform title authority records, and to submit subject heading proposals to the Library of Congress (LC). All components of the PCC program have been very successful and prolific over the years. A recent report by R2 Consulting for the LC, *Study of the North American MARC Records Marketplace*, shows that CUL were the fifth most active contributors to the BIBCO program in 2008.<sup>27</sup>

One little-known aspect of the CUL PCC training program is that it

soon expanded to include some copy catalogers who started contributing PCC-level records for belles lettres in a variety of languages. In the workflow of technical services at CUL, “belles lettres” are defined as works of fiction, poetry, and drama. This paper discusses the process and efforts involved in training paraprofessionals to catalog to PCC level, create authority records, and submit call number proposals for belles lettres. It also shows that the skills acquired by copy catalogers have potential future use for broader authority control projects and initiatives.

CUL historically have had a remarkable partnership between trainers and trainees. Expert professional catalogers and PCC trainers are capable and willing to train in PCC-level cataloging and authority work. The copy cataloging department also has employed highly motivated copy catalogers with excellent cataloging skills and valuable language expertise. From the program’s inception and through its subsequent development, crucial institutional and administrative support was provided, including managerial assistance to free resources for training and support and to establish workflows to facilitate this level of cataloging.

## Staffing Overview

The initial training was conducted by the NACO coordinator in two cataloging divisions (called departments until 2008): Monographs Processing Services (MPS) and Original and Special Materials Cataloging (OSMC). Monographs Processing Services is divided into three units: LC Copy Cataloging, Database Maintenance, and Precat (a precataloging backlog of circulating materials lacking a standard record on receipt). The responsibilities of the MPS copy cataloging staff include creating preliminary records, editing standard records according to

national and in-house cataloging rules, verifying personal name and series headings and call numbers assigned by agencies other than the LC, and various database maintenance tasks. In addition, all staff in MPS perform some original cataloging, e.g., “variant edition” cataloging based on an existing standard record for a different edition. The Original and Special Materials Cataloging Division performs most of the original cataloging at CUL, including monographs, serials, and rare books. They also support various digital and metadata initiatives throughout the libraries. One of OSMC’s goals has been to produce a large number of BIBCO records. Thus the division and its members are perfectly situated to provide training and consultation regarding BIBCO and NACO standards. The NACO coordinator was chosen to train copy catalogers in the principles of PCC cataloging and familiarize them with the NACO manual and creation of authority records for personal names and uniform titles.

The MPS and OSMC staff involved in the training process are the NACO training coordinator and the metadata coordinator in OSMC, several copy catalogers, and two professional librarians, who are supervisors in MPS. Staff from both divisions have an excellent long-standing working relationship that contributes to the success of the training and continuing development of the program.

## Training Program

In 1997 the NACO coordinator in OSMC started the PCC training program for paraprofessionals by training Slavic copy catalogers to contribute names and uniform titles authority headings to the NAF. The decision to train copy catalogers to create records for belles lettres was made under the rationale that most belles lettres do not require subject analysis and the

classification is more limited and easier to learn. Slavic languages materials were chosen initially for training purposes for several reasons. First, a steady flow of Russian and other Slavic belles lettres required immediate cataloging to avoid creating a backlog. Second, experienced and motivated copy catalogers were capable and willing to learn new workflows, an important factor that is crucial in a unionized environment. An additional contributing factor was that the NACO coordinator is also the principal Slavic cataloger, so the training and later revisions were facilitated by her language expertise.

The initial training took several weeks of one-on-one sessions familiarizing the trainees with the *NACO Participants Manual*, the appropriate Anglo-American cataloging rules, the *Library of Congress Rule Interpretations*, and WorldCat bibliographic searches needed to perform authority work.<sup>28</sup> Teaching the trainees to catalog directly in WorldCat also was imperative; prior to training, all copy cataloging work was done in the local integrated library system.

The workflow was and still is very straightforward. The copy cataloger creates a PCC-level record for the monograph in WorldCat, verifies all headings, creates authority records for new name and uniform title headings as appropriate, saves them in the online authority save file, assigns an LC call number, and finishes processing. The fully cataloged book with the attached printout of the proposed authority heading is then routed to the reviewer, who reviews the bibliographic and authority records and discusses any needed edits with the copy cataloger. After the corrections are made, the authority records are produced by the reviewer in WorldCat.

Figure 1 shows a bibliographic record created by a copy cataloger in WorldCat for a Greek title (the example includes only the variable MARC fields of the WorldCat record).

040	\$a ZCU \$c ZCU \$d ZCU
066	\$c (S
020	9789600350401
020	960035040X
042	pcc
050	4 PA5638.23.O99 \$b P45 2010
049	ZCUA
100	1 Μουζουράκης, Κώστας, \$d 1974-
100	1 Mouzourakeṣ, Koṣtas, \$d 1974-
245	1 0 Φίδια στο σκορπιό : \$b μυθιστόρημα / \$c Κώστας Μουζουράκης.
245	1 0 Phidia sto skorpio : \$b mythistoreīma / \$c Koṣtas Mouzourakeṣ.
250	1η. εκδ.
250	1e.ṯ. ekd.
260	Αθήνα : \$b Εκδόσεις Καστανιώτη, \$c 2010.
260	Atheṅa : \$b Ekdoseis Kastaniote,ṯ, \$c 2010.
300	307 p. ; \$c 22 cm.
655	0 Greek fiction, Modern.

Figure 1. PCC Level Bibliographic Record for a Greek Title

Here the 040 (cataloging source) field is ZCU—the CUL WorldCat code—and the 066 (character sets present) field indicates presence of vernacular scripts, in this case Greek. The record is coded “pcc” in the 042 (authentication code) field and the call number is assigned according to the LC guidelines for creating call numbers for literary authors. The 100 (main entry—personal name), 245 (title statement), 250 (edition statement), and 260 (imprint) fields have parallel fields in Greek; providing the vernacular script for nonroman languages is one of the strengths of the program. The name of the author is constructed according to the NACO principles, and an authority record is submitted to the LC NAF.

Figure 2 shows an example of the name authority record in the LC NAF for the personal author that appears in figure 1 (variable fields only). The name authority record in figure 2 uses the birth date of the author as part of the heading; it also provides the transliterated and Greek form of the name and cites the bibliographic information as found in the book.

The trainers implemented some modifications to the initial ambitious plan. For example, although contributing series authority records (SARs) by paraprofessional staff had been envisioned, it never fully materialized because of the additional extensive training and oversight required. For a period, some copy catalogers created SARs for Slavic materials with the

010	no2010188131
040	NNC \$b eng \$c NNC
053	PA5638.23.O99
100 1	Mouzourakēs, Kōstas, \$d 1974-
400 1	Μουζουράκης, Κώστας, \$d 1974-
667	Non-Latin script reference not evaluated.
670	Phidia sto skorpio, 2010: \$b t.p. (Κώστας Μουζουράκης = Kōstas Mouzourakēs) inside front flap (b. 1974 in Athens, Greece)

**Figure 2.** Authority Record for a Personal Author Submitted to the Library of Congress NAF

Local LC Literary Author (053)	
<i>Today's date:</i>	
11/16/2010	
<i>Your name and e-mail address:</i>	
[copy cataloger's name and e-mail address]	
<i>053: Suggested Classification Number (\$a only):</i>	
PA5638.23.O99	
<i>100: Author's name as formulated in the NAR (do not supply tag number of subfield \$a, however, supply other subfield codes with indicators when needed):</i>	
Mouzourak(macron)es, K(macron)ostas, Sd 1974-	
<i>952: Brief information about author, e. g. author's nationality or place of residence, language used for works, etc.</i>	
Greek author. His first work was published in 2010. Writes in Greek.	
<i>952: Additional information or comments:</i>	
Resides in Athens, Greece	

**Figure 3.** Sample Library of Congress Classification Proposal for Literary Author

assistance of the MPS cataloging manager. The complex and time-consuming process of formulating, correctly structuring, and revising SARs proved to be an impediment to an otherwise smooth and efficient process of cataloging to PCC level. After the departure of the paraprofessional cataloger who produced most of the SARs, MPS managers decided that this process was too time-consuming to continue.

In 2006, the LC ceased producing series authority records for new series; therefore the CUL decision not to trace most of the series on the belles lettres materials cataloged to PCC level produced by copy catalogers was the most practical one. Currently, the CUL guidelines for copy catalogers mandate recording new series lacking SARs in the NAF in the 490 (series statement) field, second indicator 0, of

the bibliographic record.

Since 2003, the workflow includes an additional component that facilitates cataloging subsequent works of the same literary author. Because classification numbers for current literary authors are generally not printed in the LC classification schedules, the class number for a literary author is included in the 053 field of the author's name authority record to keep a consistent shelflist. CUL copy catalogers are trained to assign the call number for the work in the bibliographic record and recommend a class number to be included in the 053 (LC classification number) field of the authority record. This step required additional training and involved further searching in both local and LC shelflists.

The copy cataloger uses a local Library of Congress classification proposal form (figure 3) to propose an LC literary author call number to be included in the authority record's 053 field. When this is finalized, it is used to populate the request sent to the LC. This local form also includes a date, the copy cataloger's name and e-mail address, the author's name as formulated for the name authority record, and brief information about the author's nationality or place of residence, language, occupation, and so forth.

Over time, some nonroman scripts (mostly Slavic and Modern Greek) have been added to both bibliographic and authority records, requiring further training for creating name and uniform titles authority records. For example, copy catalogers needed to learn the differences in fixed fields of authority records with and without vernacular scripts as well as the order of the transliterated fields or positions versus those in the vernacular script.

The program uses locally developed and WorldCat macros for complex cataloging tasks. The in-house macros are used in the 053 proposal form for supplying diacritics in authors' names. This step is especially

important because of the variety of vernacular scripts used in the 053 proposal forms. The WorldCat macros include creating authority records to be contributed to the NAF and adding nonroman scripts to the bibliographic record in WorldCat. Some macros, such as those to create name and uniform title authority records, were first tested in the workflows of the OSMC Division; others (e.g., adding the vernacular script to the bibliographic record in WorldCat) were initially implemented in MPS. All macros facilitate quicker cataloging of materials and result in an accurate and typo-free environment crucial for cataloging materials to PCC level.

Since the inception of the program, seven copy catalogers have been trained to create PCC-level records for belles lettres materials in a variety of languages. Most recently, Greek, English, Spanish, and some African languages have been added to the Slavic and Eastern European languages that inaugurated the program. Only seven copy catalogers were chosen to participate in the program because of a combination of the availability of belles lettres materials in certain languages that match the copy catalogers' language proficiency and a need for a high degree of cataloging judgment and a good understanding of national cataloging principles in addition to a thorough understanding of CUL's local procedures. MPS contributes a significant number of PCC-level cataloging to the CUL BIBCO and NACO production. The recent PCC production statistics for 2008–10 indicates that 21.6 percent of CUL PCC output is created by copy catalogers.

## Discussion

The process of painstakingly producing PCC-level records, and especially creating authority records for name and uniform titles, requires significantly more complex skills than does

basic copy cataloging. It is a time-consuming process involving considerable intellectual effort. The skills required for this type of high-level cataloging work can be taught to copy catalogers, but ongoing support is necessary to promote their continuing development. To identify the correct form of a personal name and appropriately construct it, researching personal names and performing additional bibliographic and authority record searches to identify the proper form of the author's name are essential. Paraprofessional staff need to continually practice establishing authority records on a regular basis to maintain these valuable skills.

Also worth mentioning are potential bargaining union issues associated with assigning original PCC-level cataloging to copy cataloging staff. Copy catalogers, unlike professional librarians at CUL, are unionized, and CUL has established a special task for PCC cataloging of belles lettres materials in its union classification system. This requires a high degree of cataloging judgment, very good understanding of national cataloging principles in addition to a thorough knowledge of CUL's local procedures, and often a substantial knowledge of specified foreign languages. Assigning this task to a copy cataloger has meant upgrading a union position to the highest level in the union classification system. This practice has been successful in the CUL environment primarily because specialized language skills are needed to perform these tasks. However, libraries lacking a steady volume of belles lettres acquisitions may not fully benefit from implementing similar programs.

Using a WorldCat search combining the PCC marker, language of cataloging, and imprint date, the authors determined that many U.S. academic and some U.S. public libraries contribute PCC-level records for belles lettres materials. Whether any of them employ paraprofessional catalogers to

create PCC-level records is unknown. In this respect, sharing this program experience with technical services managers and practitioners is important because, despite initial training expenditures and efforts to establish workflows, it is an efficient process that brings high-quality records to WorldCat to be shared by libraries and cataloging agencies.

This program has many positive results for catalogers and users alike. The success of this training and resulting contributions to BIBCO and NACO can be seen in the absence of a belles lettres backlog at CUL. Aside from revisions and corrections, professional catalogers' time is not used to process these materials. The belles lettres are available to patrons in a timely manner, thus contributing to comprehensive collection development. If the work is in a nonroman language, the record provides access to the original script, facilitating users' discovery of these materials. CUL enjoys a well-organized shelf list in the literary authors' area because of the extra attention to the classification and 053 proposals. The establishment of a literary author class number leads to a better collocation of materials for works of the same author, allowing for better user browsing capabilities. CUL PCC records for belles lettres are immediately available in WorldCat, which gives other libraries an opportunity to use them for copy cataloging faster and increases the cooperative value of the program. The contributions of the PCC-level records for belles lettres by paraprofessional staff amounts to an average of 1,060 bibliographic records per year.

In addition, CUL's NACO and BIBCO training program and steady production of name authority records by copy catalogers paves the way for future expansion, especially in developing new tools and potential opportunities for catalogers to employ their knowledge and skills. The skills acquired by copy catalogers also could

be used in other projects such as name disambiguation, metadata creation, and contributing to any type of future authority file. Recently, a survey conducted by Veve and Feltner-Reichert revealed that 71 percent of copy catalogers would like to have a broader variety of job duties and would like to be on the “cutting edge of cataloging.”<sup>29</sup> In this survey, professional and nonprofessional catalogers alike perceived creating non-MARC metadata is a “natural extension of catalogers’ responsibilities.”<sup>30</sup>

Furthermore, Banush underscored the importance of “consistent naming conventions” in the current information environment, and he believed that the PCC—and especially its NACO component—have the potential to preserve vital principles of authority control.<sup>31</sup> Contributions of name authority headings to the WorldCat Cooperative Identities Hub ([www.worldcat.org/identities](http://www.worldcat.org/identities)) or to the Virtual International Authority File (VIAF) (<http://viaf.org>), which were proposed in the 2009 “Report of the PCC Task Group on the Internationalization of the Authority Files,” can serve as examples of potential libraries participation.<sup>32</sup> Name authority control becomes crucial to ensure discoverability in the new networking environment. Copy catalogers well versed in PCC cataloging and creation of name authority records could be trained to participate in name disambiguation projects or the identification and distinguishing of names in a metadata context.

## Conclusion

The CUL PCC training program for copy catalogers has proven to be a valuable investment of cataloging resources. The program benefits not only CUL but the entire cooperative cataloging community. It provides opportunities for professional development and has great potential for the future.

The program for training paraprofessional catalogers to catalog belles lettres materials to PCC level, create name and uniform titles authority records, and propose class numbers for literary authors serves the immediate needs of CUL for more efficient and practical solutions to cataloging these materials. The belles lettres materials are available to the users in a timely manner, they are not backlogged, and, because of the authority work and a careful revision process, they represent high-quality bibliographic records. The program for training copy catalogers to create PCC records for belles lettres materials not only has a potential to benefit libraries with extensive collections of foreign poetry, drama, and fiction, but provides a model for cataloging more commonly held English language belles lettres. Moreover, the upgrade of vendor records found in WorldCat to PCC level represents an additional benefit to shared cataloging. In addition to providing bibliographic records, which can be shared immediately with peer institutions, the program trained paraprofessionals whose professional development was very much influenced by the assigned task. Of the seven copy catalogers trained in the CUL PCC program, four graduated from library schools and became professional catalogers. The success of the program also suggests other potential benefits. Currently, when the international potential of the PCC program is tested, expanding the pool of catalogers who can expertly contribute headings to either the NAF or the VIAF becomes an important consideration.<sup>33</sup> As the bibliographic universe becomes increasingly complex, the need for quality metadata, consistent author numbers, and author disambiguation will only increase in importance.

## References

1. Lori P. Robare, “Training Paraprofessionals in Classification and Subject Analysis,” *Technical Services*

- Quarterly* 14, no. 2 (1996): 49–63; Claire-Lise Benaud, “The Academic Paraprofessional Cataloger: Underappreciated?” *Cataloging & Classification Quarterly* 15, no. 3 (1992): 81–92; Jennifer A. Younger, “Support Staff and Librarians in Cataloging,” *Cataloging & Classification Quarterly* 23, no. 1 (1996): 27–47; Susanne Nevins, “Recruiting and Training the Paraprofessional Cataloger,” *College & Research Libraries* 4, no. 2 (1997): 65–92; Mary M. Rider, “Developing New Roles for Paraprofessionals in Cataloging,” *Journal of Academic Librarianship* 22, no. 1 (1996): 28–32; Deborah A. Mohr and Anita Schuneman, “Changing Roles: Original Cataloging by Paraprofessionals in ARL Libraries,” *Library Resources & Technical Services* 41, no. 3 (1997): 205–18; Madga El-Sherbini, “Copy Cataloguers and Their Changing Roles at the Ohio State University Library: A Case Study,” *Library Management* 22, no. 1/2 (2001): 80–85.
2. Robare, “Training Paraprofessionals”; El-Sherbini, “Copy Cataloguers”; Richard Sapon-White, “Subject Analysis Training for Cataloging Paraprofessionals: A Model for Ongoing Learning and Support,” *Technical Services Quarterly* 26, no. 3 (2009): 183–93.
3. Benaud, “The Academic Paraprofessional Cataloger.”
4. Younger, “Support Staff.”
5. Nevins, “Recruiting and Training.”
6. Rider, “Developing New Roles.”
7. Patricia A. Eskoz, “The Catalog Librarian: Change or Status Quo? Survey of Academic Libraries,” *Library Resources & Technical Services* 34, no. 3 (1990): 380–92.
8. Mohr and Schuneman, “Changing Roles.”
9. Lois Buttlar and Rajinder Garcha, “Catalogers in Academic Libraries: Their Evolving and Expanding Roles,” *College & Research Libraries* 59, no. 4 (1998): 311–21.
10. Sever Bordeianu and Virginia Seiser, “Paraprofessional Catalogers in ARL Libraries,” *College & Research Libraries* 60, no. 6 (1999): 532–40.
11. Elizabeth J. Cox and Ann K. D. Myers, “What is a Professional

- Cataloger? Perception Differences Between Professionals and Paraprofessionals," *Library Resources & Technical Services* 54, no. 4 (2010): 212–26.
12. Rider, "Developing New Roles"; Robare, "Training Paraprofessionals"; Mohr and Schuneman, "Changing Roles."
  13. Mohr and Schuneman, "Changing Roles."
  14. Robare, "Training Paraprofessionals"; Mohr and Schuneman "Changing Roles."
  15. Robare, "Training Paraprofessionals."
  16. El-Sherbini, "Copy Cataloguers."
  17. Library of Congress Name Authority File, <http://authorities.loc.gov> (accessed Feb. 2, 2011).
  18. Sapon-White, "Subject Analysis."
  19. *Ibid.*, 184.
  20. Mohr and Schuneman, "Changing Roles."
  21. Kathleen L. Wells, "Authority Control in Mississippi Public and Academic Libraries," *Technical Services Quarterly* 18, no. 2 (2001): 1–14; Robert E. Wolverson Jr., "Authority Control in Academic Libraries in the United States: A Survey," *Cataloging & Classification Quarterly* 41, no. 1 (2005): 111–31; Susan K. Burke and Jay Shorten, "Name Authority Work Today: A Comparison of Academic Libraries," *Library Resources & Technical Services* 54, no. 1 (2010): 4–20.
  22. Betsy Simpson and Priscilla Williams, "Growing a NACO Program: Ingredients for Success," *Cataloging & Classification Quarterly* 40, no. 1 (2005): 123–32.
  23. Peter H. Lisius et al., "Integrating Enhance and NACO Work into Pre-Professional Experience: A Successful Strategy for All," *Cataloging & Classification Quarterly* 48, no. 6/7 (2010): 623.
  24. Mechael D. Charbonneau, "Program for Cooperative Cataloging: The Indiana Experience," *Cataloging & Classification Quarterly* 48, no. 2/3 (2010): 113–25.
  25. Jean Hirons and Carlo Hixson, "White Paper on PCC Role in Continuing Education for Catalogers," prepared for PCC Policy Committee Discussion, Nov. 2001, [www.loc.gov/catdir/pcc/whitepapertmg.html](http://www.loc.gov/catdir/pcc/whitepapertmg.html) (accessed Nov. 22, 2010).
  26. David Banush, "Cooperative Cataloging at the Intersection of Tradition and Transformation: Possible Futures for the Program for Cooperative Cataloging," *Cataloging & Classification Quarterly* 48, no. 2/3 (2010): 254–57.
  27. Ruth Fisher and Rick Lugg, *Study of the North American MARC Records Marketplace*, Oct. 2009, [www.loc.gov/bibliographic-future/news/MARC\\_Record\\_Marketplace\\_2009-10.pdf](http://www.loc.gov/bibliographic-future/news/MARC_Record_Marketplace_2009-10.pdf) (accessed Nov. 22, 2010).
  28. Patricia Van Ryn and William L. Starck, eds., *NACO Participants Manual*, 3rd ed. (Washington, D.C.: Library of Congress, Cooperative Cataloging Team, 2005); Library of Congress, Cataloging Distribution Service, *Library of Congress Rule Interpretations*, 2nd ed. (Washington D.C.: Library of Congress, 1989–).
  29. Mariel Veve and Melanie Feltner-Reichert, "Integrating Non-MARC Metadata Duties into the Workflow of Traditional Catalogers: A Survey of Trends and Perceptions among Catalogers in Four Discussion Lists," *Technical Services Quarterly* 27, no. 2 (2010): 199.
  30. *Ibid.*, 205.
  31. Banush, "Cooperative Cataloging," 255.
  32. Joan Schuitema et al., "Report of the PCC Task Group on the Internationalization of the Authority Files," Aug. 15, 2009, [www.loc.gov/catdir/pcc/scs/IAFReport.pdf](http://www.loc.gov/catdir/pcc/scs/IAFReport.pdf) (accessed Nov. 22, 2010).
  33. *Ibid.*

# Book Reviews

Norm Medeiros

***Making a Collection Count: A Holistic Approach to Library Collection Management.*** By Holly Hibner and Mary Kelly. Oxford: Chandos, 2010. 172 p. \$75.00 paper (ISBN 978-1-9433-4606-7).

This book is part of a new series by Chandos “designed to provide easy-to-read and (most importantly) practical coverage of topics that are of interest to librarians” (series t.p.). Collection management is certainly a topic of interest, and Hibner and Kelly succeed in providing an easy-to-read volume with a practical bent. Their tone is straightforward and occasionally even chatty. The focus is on collection management, not to be confused with the narrower topic of collection development. Both authors work at public libraries, and *Making a Collection Count*, while not explicitly aimed at public libraries, seems most appropriate as an introduction to the subject of collection management from that perspective.

Hibner and Kelly divide their work into eight chapters: “Life Cycle of a Collection,” “Understanding Your Workflow,” “Collection Audit and Using Statistics,” “Physical Inventory,” “Creating Collection Objectives and Benchmarks,” “Collection Organization,” “Making the Most of a Library Collection Budget,” and “Everything is Connected.” Most chapters include a brief, current bibliography. The authors also include two appendixes: a sample public library collection management policy and strategic content negotiation for the small library.

The first chapter, on the life cycle of a collection, aptly introduces collection management. Hibner and Kelly emphasize that the collection-management policy is the basis for all work

connected with the collection, and they stress the need for staff collaboration during all phases of the collection life cycle. The authors enumerate the stages of a collection’s life cycle this way: selection, acquisition, processing, shelving, checkout, use, check-in, re-shelving, repair and maintenance, and weeding.

In their next chapter, Hibner and Kelly suggest that workflow analysis is a fundamental way to improve the quality of collection management by understanding the particulars of the associated processes. They examine the details of each of their collection life cycles and provide examples for performing the workflow analysis. However, considering how time and labor intensive workflow analysis can be, the authors did not make a clear case that the benefit is worth the cost for all situations or libraries.

Hibner and Kelly’s use of the term “collection quality” was initially confusing. They are largely concerned about the quality of information about the collection (Does the book label match the catalog record? Is the book where the catalog says it is? Is the catalog status correct?) rather than the quality of the content (Is it up to date? Is it reliable?). Both kinds of quality are important and their intended meaning became clear, but the semantics were challenging at first. However, in chapters 3 and 4, Hibner and Kelly provide good suggestions for how a collection audit and physical inventory can help not only to assess the quality of information being provided about the collection to staff and patrons, but also to identify problems, both isolated and systemic.

In their chapter on collection objectives and benchmarks, the authors provide examples for establishing these

performance criteria and suggestions for evaluation of them. In this age of assessment, having as many strategies available as possible is important, and new ideas and perspectives are welcome. Hibner and Kelly next provide a concise introduction to the issues of physical space and the collection when they discuss collection organization. They point out some of the potential usage implications of subdividing collections, collection location, the choice of classification systems, special displays, and signage.

The authors contend, “It is well worth the time and effort to enhance a library collection to go through the public’s generous donations in pursuit of useful items!” (119). There may indeed be gems in library donations, and it may be important to accept gifts from a public-relations standpoint, but Hibner and Kelly are not convincing in their argument that gifts are worth the time and effort required to sort, evaluate, and process them. The most useful suggestion in the budget chapter may be for library staff to remind users that library services and collections are not free.

In the final chapter, Hibner and Kelly go beyond collection management to discuss holistic library service, including the role of the collection. The most interesting aspect is the discussion of holistic budgeting, which may provide an attractive alternative for some libraries to explore.

*Making a Collection Count* provides a good introduction to the many aspects of collection management. The examples are drawn largely from public libraries, but the framework is applicable to most library collections.—Karen Greever ([greeverk@kenyon.edu](mailto:greeverk@kenyon.edu)), Kenyon College, Gambier, Ohio

***Structures for Organizing Knowledge: Exploring Taxonomies, Ontologies, and Other Schemas.***

By June Abbas. New York: Neal-Schuman, 2010. 249 p. \$85.00 soft-cover (ISBN 978-1-5557-0699-9).

The stated goals of *Structures for Organizing Knowledge: Exploring Taxonomies, Ontologies, and Other Schemas* are to examine how people organize information in personal and professional contexts; to explore the roles of categories, taxonomies, and other structures in that work; and to understand the human organizing behaviors that should guide the design of useful information structures (xv–xvi). The book is intended both for students and scholars studying information organization and for information professionals seeking inspiration for and insight into the design of organizational structures. In keeping with its potential use as a textbook, each chapter begins with a summary statement of themes and concludes with a set of “thought exercises.” Along the way, readers are directed to other sources for more in-depth treatment of particular topics.

The book is divided into three parts. Part 1 studies the traditional structures used to organize information, including both structures used to describe particular resources (such as the MARC record) and those used to organize concepts ranging from the very general to the highly specific. Abbas’s account of the familiar tools of library science—subject lists, controlled vocabularies, classification systems, etc.—is followed by an analysis of classification systems developed in the natural sciences and by a discussion of cognitive science’s reformulation of categories as structures based on the observers’ variable perceptions of “family resemblances,” and not simply on a binary logic of sameness and difference. Part 1 also covers the standards environment responsible for developing and maintaining many of the traditional structures and the

importance of guidelines for application of standards-based structures. Part 1 accounts for roughly 60 percent of the text, which reflects the diversity, depth, and complexity of traditional organizing practices in the various disciplines examined.

Part 2 examines the ways individuals organize information in personal and professional settings. The author reviews research in both areas, noting such interesting findings as the limited effort people devote to organizing their own information space, the tendency of that space to be occupied by unintentional accumulations, and the importance of intended use and frequency of use to personal classification decisions. Abbas notes a preponderance of studies of how successfully people adapt to new organizational design prototypes, but a lack of research “into how people organize their paper-based and digital information” (167) and the absence of such supplied structures. Research into “the structure(s) we employ for organizing personal information in both work and nonwork environments may reflect how we conceptually think about and structure the routines, tasks, projects, etc., that we are engaged in on a daily basis and over a lifetime” (162). This quest for a deeper understanding of the “natural” organizational impulses of human beings is a running theme of the book. Yet “natural” behavior is perhaps not a major driver for some organizational behavior. More might have been said about the additional factors that come into play when structures for information organization are intended to be built and shared by communities.

Part 3 turns to the emerging study of socially based knowledge-organization tools, including LibraryThing, Flickr, and YouTube. The structures examined in this section include user tags, social bookmarking, folksonomies, and others, deployed in different combinations by different Web 2.0 applications. As user choices accrete

in these environments, they drive the development of community preferences for particular models of terminology and socially based ranking. Yet the advantages of flexibility can be offset by the disadvantages of inconsistency and lack of consensus. Abbas observes that while the studies in part 3 “provide us with a window into how people naturally will organize objects in their personal knowledge spaces,” they do not reveal “the structures people prefer for organizing objects” within these spaces (200).

In the last two chapters, Abbas explores how insights from the three areas discussed might be combined to provide a basis for designing more flexible and useful organizing structures. She notes that while the tools of socially constructed knowledge organization have been introduced into such traditional structures as the online catalog, little empirical evidence has emerged to prove that users find this combination useful and that further research into the management of tags and folksonomies is needed. She also argues for giving more attention to a fourth “thread” in the pattern being woven, highlighting not just the structures of socially based information, but their participatory aspect and value as users’ voices engage with providing new contexts for understanding information objects.

In her Preface, Abbas writes,

The book is not meant to be a “how-to” guide for developing, applying, or implementing structures for organizing knowledge; rather, it is designed to present a conceptual discourse and to inspire thinking about taxonomic behavior, or how and why people organize knowledge, in various contexts. (xix)

*Structures for Organizing Knowledge* succeeds on these terms, providing both a thoughtful survey of

research findings and a stimulating complex of issues to ponder. Inevitably, some topics are slighted; for example, relatively little is said about the potential of the Semantic Web and open linked data as a new structural model for organizing information. This does not detract from the valuable contribution Abbas makes in this book to the study of knowledge organization.—*Stephen Hearn (s-hear@umn.edu), University of Minnesota, Minneapolis*

**Starting, Strengthening and Managing Institutional Repositories: A How-To-Do-It Manual.** By Jonathan A. Nabe. New York: Neal-Schuman, 2009. 169 p. \$85.00 softcover (ISBN 978-1-5557-0689-0). How-To-Do-It Manuals.

*Starting, Strengthening and Managing Institutional Repositories: A How-To-Do-It Manual* is a practical guide that combines lessons and expertise of early institutional repository (IR) implementers. The book can be read straight through or consulted chapter-by-chapter as needed. Each chapter is clearly written, contains a bibliography, and can stand on its own.

The book has two parts. The first seven chapters, written by Nabe, cover all aspects of IRs, from why libraries should adopt them through assessment. The second part of the book consists of seven chapters by authors who provide their own perspectives on IR management. Both parts work well together to form a cohesive, strongly written whole.

In part 1, the introduction and first chapter define IRs and explain why libraries should adopt them. These sections provide useful background for people who are not familiar with IRs.

Chapter 2 covers planning. One of the recurring themes in the book—avoiding too much planning—is introduced on page 13:

Inability to address these issues should not stymie all progress, and overplanning

can lead to frustration and gridlock. Furthermore, there is no demonstrable correlation between the resources committed to these ancillary activities and the success . . . of an IR.

Chapter 3 covers the major IR platforms and provides criteria for evaluation. Even if one's institution already has an IR, understanding the benefits and limitations of the software options is helpful to better understand the structure of other IRs. Because software features continually improve, this chapter should be used in conjunction with websites about each of the products.

Chapter 4 provides an overview of policies that should be in place for an IR. It gives examples from several institutions, encouraging new repository administrators to learn from the early adopters. Adapting another institution's policy is a great time-saver when starting an IR program.

The next chapters cover marketing, recruitment of content, and collection development. The author gives a realistic picture of the difficulty in recruiting content. He also provides good information regarding how to communicate with faculty and researchers about how the repository can fulfill their needs.

The last chapter of part 1 covers use and assessment. Increasing the use of one's IR relies on making content discoverable in search engines, often through the use of interoperable metadata. Some of the information conveyed in this chapter is slightly out-of-date (e.g., OAIster is now part of WorldCat), but the concepts are still valid.

Part 3 provides expert views on topics, as well as a variety of perspectives on specific software products. While many of the issues the authors raise echo points made earlier in the book, this collection of essays also offers fresh ideas. For instance, it is

instructive to learn from these IR managers why they selected their systems. It is also helpful that the authors represent a diverse set of institutions. Most importantly, since each of the case studies represents a successful IR, readers can learn from multiple people and find lessons relevant to their own institutions.

The book only touches lightly on archiving datasets and using an IR as a publishing platform, despite the emerging importance of these issues in the last two years. The lack of this information is a minor criticism; the nature of a monograph is to capture the state of a topic at a given point in time.

*Starting, Strengthening and Managing Institutional Repositories* is a very useful collection of information for managers of existing repositories. It would have been extremely helpful when we were beginning our repository; I recommend it to colleagues embarking on such an endeavor.—*Wendy C. Robertson (wendy-robertson@uiowa.edu), University of Iowa, Iowa City*

**Collection-Level Cataloging: Bound-With Books.** By Jain Fletcher. Third Millennium Cataloging. Santa Barbara, Calif.: Libraries Unlimited, 2010. 97 p. \$45.00 softcover (ISBN 978-1-5915-8543-5).

Recognizing the disparity between time-intensive cataloging of bound-with books according to current code and guidelines, and the quick processing advocated by the "core record" movement and "hidden collections" initiative, Fletcher offers a reasonable middle ground in her book: collection-level cataloging. In this accessible and comprehensive manual, Fletcher shares her expertise while encouraging readers to tackle the challenge of bound-with cataloging. To that end, Fletcher's discourse fits neatly into the *Third Millennium Cataloging* series, which "provides an ongoing set of guides to problems of contemporary cataloging, and clarifies issues,

research findings and a stimulating complex of issues to ponder. Inevitably, some topics are slighted; for example, relatively little is said about the potential of the Semantic Web and open linked data as a new structural model for organizing information. This does not detract from the valuable contribution Abbas makes in this book to the study of knowledge organization.—*Stephen Hearn (s-hear@umn.edu), University of Minnesota, Minneapolis*

**Starting, Strengthening and Managing Institutional Repositories: A How-To-Do-It Manual.** By Jonathan A. Nabe. New York: Neal-Schuman, 2009. 169 p. \$85.00 softcover (ISBN 978-1-5557-0689-0). How-To-Do-It Manuals.

*Starting, Strengthening and Managing Institutional Repositories: A How-To-Do-It Manual* is a practical guide that combines lessons and expertise of early institutional repository (IR) implementers. The book can be read straight through or consulted chapter-by-chapter as needed. Each chapter is clearly written, contains a bibliography, and can stand on its own.

The book has two parts. The first seven chapters, written by Nabe, cover all aspects of IRs, from why libraries should adopt them through assessment. The second part of the book consists of seven chapters by authors who provide their own perspectives on IR management. Both parts work well together to form a cohesive, strongly written whole.

In part 1, the introduction and first chapter define IRs and explain why libraries should adopt them. These sections provide useful background for people who are not familiar with IRs.

Chapter 2 covers planning. One of the recurring themes in the book—avoiding too much planning—is introduced on page 13:

Inability to address these issues should not stymie all progress, and overplanning

can lead to frustration and gridlock. Furthermore, there is no demonstrable correlation between the resources committed to these ancillary activities and the success . . . of an IR.

Chapter 3 covers the major IR platforms and provides criteria for evaluation. Even if one's institution already has an IR, understanding the benefits and limitations of the software options is helpful to better understand the structure of other IRs. Because software features continually improve, this chapter should be used in conjunction with websites about each of the products.

Chapter 4 provides an overview of policies that should be in place for an IR. It gives examples from several institutions, encouraging new repository administrators to learn from the early adopters. Adapting another institution's policy is a great time-saver when starting an IR program.

The next chapters cover marketing, recruitment of content, and collection development. The author gives a realistic picture of the difficulty in recruiting content. He also provides good information regarding how to communicate with faculty and researchers about how the repository can fulfill their needs.

The last chapter of part 1 covers use and assessment. Increasing the use of one's IR relies on making content discoverable in search engines, often through the use of interoperable metadata. Some of the information conveyed in this chapter is slightly out-of-date (e.g., OAISTER is now part of WorldCat), but the concepts are still valid.

Part 3 provides expert views on topics, as well as a variety of perspectives on specific software products. While many of the issues the authors raise echo points made earlier in the book, this collection of essays also offers fresh ideas. For instance, it is

instructive to learn from these IR managers why they selected their systems. It is also helpful that the authors represent a diverse set of institutions. Most importantly, since each of the case studies represents a successful IR, readers can learn from multiple people and find lessons relevant to their own institutions.

The book only touches lightly on archiving datasets and using an IR as a publishing platform, despite the emerging importance of these issues in the last two years. The lack of this information is a minor criticism; the nature of a monograph is to capture the state of a topic at a given point in time.

*Starting, Strengthening and Managing Institutional Repositories* is a very useful collection of information for managers of existing repositories. It would have been extremely helpful when we were beginning our repository; I recommend it to colleagues embarking on such an endeavor.—*Wendy C. Robertson (wendy-robertson@uiowa.edu), University of Iowa, Iowa City*

**Collection-Level Cataloging: Bound-With Books.** By Jain Fletcher. Third Millennium Cataloging. Santa Barbara, Calif.: Libraries Unlimited, 2010. 97 p. \$45.00 softcover (ISBN 978-1-5915-8543-5).

Recognizing the disparity between time-intensive cataloging of bound-with books according to current code and guidelines, and the quick processing advocated by the "core record" movement and "hidden collections" initiative, Fletcher offers a reasonable middle ground in her book: collection-level cataloging. In this accessible and comprehensive manual, Fletcher shares her expertise while encouraging readers to tackle the challenge of bound-with cataloging. To that end, Fletcher's discourse fits neatly into the *Third Millennium Cataloging* series, which "provides an ongoing set of guides to problems of contemporary cataloging, and clarifies issues,

research findings and a stimulating complex of issues to ponder. Inevitably, some topics are slighted; for example, relatively little is said about the potential of the Semantic Web and open linked data as a new structural model for organizing information. This does not detract from the valuable contribution Abbas makes in this book to the study of knowledge organization.—*Stephen Hearn (s-hear@umn.edu), University of Minnesota, Minneapolis*

**Starting, Strengthening and Managing Institutional Repositories: A How-To-Do-It Manual.** By Jonathan A. Nabe. New York: Neal-Schuman, 2009. 169 p. \$85.00 softcover (ISBN 978-1-5557-0689-0). How-To-Do-It Manuals.

*Starting, Strengthening and Managing Institutional Repositories: A How-To-Do-It Manual* is a practical guide that combines lessons and expertise of early institutional repository (IR) implementers. The book can be read straight through or consulted chapter-by-chapter as needed. Each chapter is clearly written, contains a bibliography, and can stand on its own.

The book has two parts. The first seven chapters, written by Nabe, cover all aspects of IRs, from why libraries should adopt them through assessment. The second part of the book consists of seven chapters by authors who provide their own perspectives on IR management. Both parts work well together to form a cohesive, strongly written whole.

In part 1, the introduction and first chapter define IRs and explain why libraries should adopt them. These sections provide useful background for people who are not familiar with IRs.

Chapter 2 covers planning. One of the recurring themes in the book—avoiding too much planning—is introduced on page 13:

Inability to address these issues should not stymie all progress, and overplanning

can lead to frustration and gridlock. Furthermore, there is no demonstrable correlation between the resources committed to these ancillary activities and the success . . . of an IR.

Chapter 3 covers the major IR platforms and provides criteria for evaluation. Even if one's institution already has an IR, understanding the benefits and limitations of the software options is helpful to better understand the structure of other IRs. Because software features continually improve, this chapter should be used in conjunction with websites about each of the products.

Chapter 4 provides an overview of policies that should be in place for an IR. It gives examples from several institutions, encouraging new repository administrators to learn from the early adopters. Adapting another institution's policy is a great time-saver when starting an IR program.

The next chapters cover marketing, recruitment of content, and collection development. The author gives a realistic picture of the difficulty in recruiting content. He also provides good information regarding how to communicate with faculty and researchers about how the repository can fulfill their needs.

The last chapter of part 1 covers use and assessment. Increasing the use of one's IR relies on making content discoverable in search engines, often through the use of interoperable metadata. Some of the information conveyed in this chapter is slightly out-of-date (e.g., OAISTER is now part of WorldCat), but the concepts are still valid.

Part 3 provides expert views on topics, as well as a variety of perspectives on specific software products. While many of the issues the authors raise echo points made earlier in the book, this collection of essays also offers fresh ideas. For instance, it is

instructive to learn from these IR managers why they selected their systems. It is also helpful that the authors represent a diverse set of institutions. Most importantly, since each of the case studies represents a successful IR, readers can learn from multiple people and find lessons relevant to their own institutions.

The book only touches lightly on archiving datasets and using an IR as a publishing platform, despite the emerging importance of these issues in the last two years. The lack of this information is a minor criticism; the nature of a monograph is to capture the state of a topic at a given point in time.

*Starting, Strengthening and Managing Institutional Repositories* is a very useful collection of information for managers of existing repositories. It would have been extremely helpful when we were beginning our repository; I recommend it to colleagues embarking on such an endeavor.—*Wendy C. Robertson (wendy-robertson@uiowa.edu), University of Iowa, Iowa City*

**Collection-Level Cataloging: Bound-With Books.** By Jain Fletcher. Third Millennium Cataloging. Santa Barbara, Calif.: Libraries Unlimited, 2010. 97 p. \$45.00 softcover (ISBN 978-1-5915-8543-5).

Recognizing the disparity between time-intensive cataloging of bound-with books according to current code and guidelines, and the quick processing advocated by the "core record" movement and "hidden collections" initiative, Fletcher offers a reasonable middle ground in her book: collection-level cataloging. In this accessible and comprehensive manual, Fletcher shares her expertise while encouraging readers to tackle the challenge of bound-with cataloging. To that end, Fletcher's discourse fits neatly into the *Third Millennium Cataloging* series, which "provides an ongoing set of guides to problems of contemporary cataloging, and clarifies issues,

primarily in the electronic environment.”<sup>1</sup> Still, the author’s choice of the title *Collection-Level Cataloging: Bound-With Books* is peculiar because it suggests a limited focus on collection-level treatment when this text actually serves as an exhaustive manual on the cataloging of bound-with books, both complete and collective treatment alike.

The book consists of an introduction, seven chapters, sources cited, and an index. Structurally, the chapters can be divided into two sections: (1) a historical overview of bound-with books (chapters 1–3), including general background, early cataloging treatment, and an analytical survey of existing cataloging guidelines; and (2) the author’s approach to cataloging bound-with books (chapters 4–7), including precataloging identification and preparation, a best practices approach to complete cataloging treatment, guidance on collection-level cataloging, and concluding thoughts.

The first chapter, “Background and Overview,” is particularly informative for those unfamiliar with bound-with books. The author begins by clarifying terminology, providing description (complete with photographic illustration of bindings and contents), and contextualizing the historical practice of binding separately issued materials. In the following chapter, Fletcher discusses early cataloging treatment of bound-with books in both book and card catalogs, and successfully conveys how the often-inadequate descriptions kept the contents hidden from researchers. The third chapter offers an objective analysis and comparison of existing rules, guidelines, and relevant practices for cataloging bound-with books, primarily focusing on *Rules for Descriptive Cataloging*

*in the Library of Congress; Anglo-American Cataloguing Rules*, 2nd ed., rev.; *Descriptive Cataloging of Rare Materials (Books)*; and International Standard Bibliographic Description (ISBD).<sup>2</sup> Fletcher’s analysis is effective in navigating the opacity of the various sets of guidelines, revealing where discrepancies exist and demonstrating how the lines are often blurred between volumes compiled by a publisher and volumes compiled by an individual.

Catalogers seeking practical guidance on describing bound-with books can skip directly to chapters 4–6. Fletcher’s vast knowledge and experience in handling these problematic materials are reflected in a compendium of cataloging options. In these chapters, Fletcher presents clues on distinguishing between unique and publisher-issued compilations, provides instruction on precataloging physical preparation, provides a synthesis of best practices to aid in complete cataloging treatment, presents three alternatives for collection-level treatment, and reveals the decision-making process behind selecting the appropriate course of action. These chapters contain supporting examples that elegantly illustrate the options described; Fletcher goes further by offering guidance to a few ambiguous situations that catalogers may confront.

Chapter 6, “Collection-Level Cataloging for Bound-With Books,” contains the author’s core contribution in providing a feasible solution to efficiently catalog bound-with books. Fletcher contends that uniquely compiled bound-with books are, in fact, small collections, and should therefore be treated as such in describing them. Her argument is compelling; however, despite the title of the book and

the detailed instruction given in this chapter, Fletcher gives equal justification for complete cataloging treatment, stating that it “should always be considered as viable; indeed, it should be considered as the first option” (52).

Chapter 7 concludes with the author’s final thoughts, in which she issues a challenge to catalogers: seek out bound-with books, whether currently on the shelves with less-than-full cataloging records or collecting dust in the backlog, and prioritize them for cataloging. Given Fletcher’s contribution to providing guidance on the cataloging of bound-with books and her justification of the importance of unlocking the potential research value within them, she is not only encouraging, but convincing.—*Sandy Rodriguez (rodriguezsan@umkc.edu), University of Missouri–Kansas City*

## References

1. ABC-CLIO, “ABC-CLIO—Series, Third Millennium Cataloging,” [www.abc-clio.com/series.aspx?id=51798](http://www.abc-clio.com/series.aspx?id=51798) (accessed Mar. 18, 2011).
2. Library of Congress. Descriptive Cataloging Division, *Rules for Descriptive Cataloging in the Library of Congress* (Washington, D.C.: Library of Congress, Descriptive Cataloging Division: for sale by the Supt. of Docs., U.S. G.P.O., 1949); *Anglo-American Cataloguing Rules*, 2nd ed., 2004 rev. (Ottawa: Canadian Library Assn.; London: Library Assn.: Chicago, 2004+); Bibliographic Standards Committee, Rare Books and Manuscripts Section, Association of College and Research Libraries, in collaboration with the Cataloging Policy and Support Office of the Library of Congress, *Descriptive Cataloging of Rare Materials (Books)* (Washington, D.C.: Cataloging Distribution Service, Library of Congress, 2007).