Library Resources & Technical Services

ISSN 2159-9610
July 2017
Volume 61, No. 3

ALCTS
Association for Library Collections & Technical Services

OJS Is the New Host of *LRTS*

http://journals.ala.org/lrts

Library Resources & Technical Services (*LRTS*), which went completely online in January 2015, is now available at http://journals.ala.org/lrts on the Open Journal Systems (OJS) platform hosted by the American Library Association (ALA) and is updated and maintained internally by in-house production staff. Content is no longer available on the Metapress site.

ACCESSING YOUR CONTENT

Every *LRTS* user (members and subscribers are both considered *LRTS* users) can login to journals.ala.org/lrts using the same credentials used for other ALA websites, including www.ala.org, ALA conference registration, and ALA Connect. Users will be notified in advance by ALCTS staff before the OJS login requirements take effect.

Be sure to visit http://journals.ala.org to enjoy other ALA digital journals and newsletters, including *Library Technology Reports, Reference & User Services Quarterly, Childen & Libraries,* and the *Smart Libraries Newsletter.*

For Technical Questions and Support

Please contact journals@ala.org.

Membership or Subscription Questions

To receive free access to *LRTS*, join ALA and ALCTS by visiting www.ala.org/membership/joinala or call ALA's Member and Customer Service (MACS) department at 1-800-545-2433 and press 5. Contact MACS with membership questions, too.

To subscribe to *LRTS* or to ask questions about your existing subscription, email subscriptions@ala.org or call ALA's MACS department at 1-800-545-2433 and press 5.

If you have any general questions about *LRTS*, please contact Brooke Morris (bmorris@ala.org) in the ALCTS Office.

Library Resources & Technical Services, journals.ala.org/lrts (ISSN 2159-9610) is published quarterly by the American Library Association, 50 E. Huron St., Chicago, IL 60611. It is the official publication of the Association for Library Collections & Technical Services, a division of the American Library Association, and provided as a benefit to members. Subscription price to nonmembers $100. Individual articles can be purchased for $15. Business Manager: Keri Cascio, Executive Director, Association for Library Collections & Technical Services, a division of the American Library Association. Submit manuscripts using the online system: http://www.editorialmanager.com/lrts. Mary Beth Weber, Editor, *Library Resources & Technical Services*; e-mail: mbfecko@rulmail.rutgers.edu. Advertising: ALCTS, 50 E. Huron St., Chicago, IL 60611; 312-280-5038; fax: 312-280-5033; alcts@ala.org. ALA Production Services: Chris Keech, Tim Clifford, Lauren Ehle, and Hannah Gribetz. Members may update contact information online by logging in to the ALA website (http://www.ala.org) or by contacting the ALA Member and Customer Services Department—*Library Resources & Technical Services*, 50 E. Huron St., Chicago, IL 60611; 1-800-545-2433. Nonmember subscribers: Subscriptions, orders, changes of address, and inquiries should be sent to *Library Resources & Technical Services*, Subscription Department, American Library Association, 50 E. Huron St., Chicago, IL 60611; 1-800-545-2433; fax: (312) 944-2641; subscriptions@ala.org.

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

Instructions for authors appear on the *Library Resources & Technical Services* webpage at http://www.ala.org/alcts/resources/lrts. Copies of books for review should be addressed to Elyssa M. Gould, University of Michigan Law Library, 801 Monroe St, Ann Arbor, MI 48109; e-mail: lrtsbookreviews@lists.ala.org.

Association for Library Collections & Technical Services (ALCTS)

For current news and reports on ALCTS activities, see the ALCTS News at http://www.ala.org/alctsnews.

LRTS was available in print (ISSN 0024-2527) from 1957 through 2014. Single print issues from volume 38 through volume 58 can be purchased for $30 each. Contact alcts@ala.org with purchase requests.

Library Resources & Technical Services

ISSN 2159-9610 **July 2017** **Volume 61, No. 3**

Cover image: "Perched, 2016," photo by John Brennan.

Association for Library Collections & Technical Services (ALCTS)
Visit *LRTS* online at http://www.ala.org/alcts/lrts.
For current news and reports on ALCTS activities, see the *ALCTS News* at http://www.ala.org/alctsnews.

Editorial

I am delighted to announce that Kelly Thompson's paper "More than a Name: a Content Analysis of Name Authority Records for Authors Who Self-identify as Trans" is the winner of the 2017 Edward Swanson Memorial Best of *LRTS* Award. The award jury selected this paper in recognition of how it addresses timely and relevant issues in our profession, and because it makes a powerful case that our name authority initiatives should respect a person's agency. I concur, and believe that the topics it raises are just the beginning of a long process of change, and one for the greater good. It was my pleasure to work with Kelly, and the honor is well deserved. Kelly will be presented with the award during the ALCTS Awards Ceremony during the 2017 ALA Annual Conference in Chicago.

Presentations and publications are natural outgrowths of our work and provide the opportunity to share our experiences (successes and failures) to benefit our colleagues. We work in a very collaborative profession where information is readily exchanged. I find solutions to challenges through ALCTS e-forums, discussion lists, reports in the *ALCTS News* and of course, *LRTS* papers. Consider submitting a paper to *LRTS*. Share your experience and knowledge. Information on the journal and submissions guidelines are available at http://www.ala.org/alcts/resources/lrts. Scroll down to the "For authors" section, which also includes a FAQ, resources for authors, a helpful guide on how to turn a presentation into a paper, and other helpful information. Potential authors are also encouraged to contact me directly to discuss ideas for submissions.

When papers are submitted to *LRTS*, I assign them to two reviewers from the editorial board, based on their expertise. The papers undergo a double blind review, meaning that the reviewers' identities are unknown, even to each other. The same is true for the author's identity. Papers are evaluated on criteria including relevance to the journal's scope, documentation and sources of background information, research methods, and presentation. After reviews are complete, the author receives from me a summary of both reviewers' feedback and a copy of the paper showing the reviewers' feedback using Word's "track changes" feature. Authors have an opportunity to revise and resubmit papers. In some cases, a paper might undergo a second round of review. In these cases, one of the original reviewers and a new reviewer are assigned to the paper.

Published paper are not limited to mainstream issues in technical services. Topics of published papers include revising cataloging standards to meet the needs of people with disabilities (volume 57, no. 1, 2013), using genealogist's tools to identify the long dead and little known (volume 60, no. 4, 2016), and the use of digital images in North American dental schools (volume 52, no. 3, 2008).

In closing, I will provide a preview of the contents of this issue of *LRTS*:

- John Pell and Megan Huppuch detail their assessment of the information management practices of a large non-for-profit organization in the field of reproductive health and the development and implementation of an information management pilot for that organization in their paper "IGAPS: A Taxonomy and Facet Classification System."
- Divergent practices in a shared bibliographic database can generate unexpected display issues that affect the user experience. This issue can be

compounded when databases from multiple institutions are merged. In "Using Automation and Batch Processing to Remediate Duplicate Series Data in a Shared Bibliographic Catalog," Elaine Dong, Margaret Anne Glerum, and Ethan Fenichel share their experience with the application of automation tools during a large scale series remediation project.

- In "Repositories at Master's Institutions: A Census and Analysis," Deborah B. Henry and Tina M. Neville used a population of Carnegie-designated Master's institutions to attempt to quantify the existence of digital repositories at those institutions. Pathways of discovery were also noted.
- RDA implementation typically means that an organization will no longer include General Material Designations (GMD) in their resource description. James Kalwara, Melody Dale, and Marty Coleman detail how libraries may benefit from retaining GMDs in their catalog to support user tasks. In their paper "GMD or No GMD: RDA Implementation for a Consortial Catalog," they detail the challenges the Mississippi State Libraries encountered when leading RDA enrichment for the Mississippi Library Partnership consortium.
- Book reviews courtesy of *LRTS* Book Review Editor Elyssa.

Repositories at Master's Institutions

A Census and Analysis

Deborah B. Henry and Tina M. Neville

Using a population of Carnegie-designated master's institutions, this study attempted to quantify the existence of digital repositories at those institutions. A content analysis of repositories containing some type of faculty content was conducted. Pathways of discovery of these collections—including open web searching, inclusion in repository directories, and access through an institution's website—were also noted. Approximately 20 percent of the master's colleges and universities maintain repositories containing faculty scholarship plus many other types of student productivity and university documents.

Since Lynch and Lippincott published a comprehensive census of institutional repositories (IR) in 2005, numerous studies have examined topics relating to the growth, development, and content of academic repositories.[1] Subsequent investigations often focused on repositories at major research institutions, particularly members of the Association of Research Libraries (ARL) since these institutions were early adopters of IRs.[2] Much of the IR literature is survey- or interview-based, soliciting information and experience from librarians, repository administrators, faculty, and students about the maintenance of the repository or user awareness of it.[3] Other researchers conducted content analyses of repositories, but many of those projects are dated or considered as a subset of operating repositories in the United States.[4] Investigators indicated a need for more research on IRs at smaller academic institutions, analyses comparing faculty and student content, and assessments of scholarly and non-scholarly content.[5]

Master's-level colleges and universities provide a unique contrast between institutions that focus primarily on teaching undergraduates and those with a dominant research agenda. The majority of repository content at smaller and teaching-oriented institutions may consist of student research.[6] Faculty at master's institutions often have larger teaching assignments yet still have a strong interest in and an obligation to conduct research. As at research-focused universities, faculty at master's-level institutions may be very interested in promoting their research accomplishments through an IR.

The main purpose of this study was to conduct a thorough census of institutional repositories supported by Carnegie-classified master's colleges and universities (small, medium, and large programs), thus providing a comprehensive and updated inventory of master's repositories.[7] In addition to documenting the existence of these repositories, this project sought to investigate the type of content that they contained. Considering research expectations at master's institutions, the study focused primarily on determining the percent of repositories that contained some type of faculty content but also recorded other types of content to compare results with previously published studies on academic repositories. A

Deborah B. Henry (henry@mail.usf.edu) and **Tina M. Neville** (neville@mail.usf.edu) are both Librarians at Nelson Poynter Memorial Library, University of Florida St. Petersburg.

Manuscript submitted June 13, 2016; returned to authors for revision September 2, 2016; revised manuscript submitted October 11, 2016; manuscript returned to authors for minor revision January 3, 2017; revised manuscript submitted January 19, 2017; accepted for publication March 10, 2017.

third goal of the study was to analyze discoverability using these possible pathways: entry for the IR in an established directory (Registry of Open Access Repositories (ROAR) or the Directory of Open Access Repositories (OpenDOAR)), tracking discoverability through the open web, and through the home organization's webpages.[8]

Literature Review

Censuses

Several authors have attempted to define the number and growth of institutional repositories throughout the United States. Lynch and Lippincott conducted the first major study in 2005. Their analysis focused on Coalition for Networked Information (CNI) members, a joint project of ARL and Educause. Survey respondents were consortial members from ninety-seven doctoral-granting institutions and thirty-five liberal arts colleges. At the time of the survey, 40 percent of the CNI members had an IR in place and 88 percent of the remainder planned to implement one. Only two of the liberal arts institutions, however, had a working repository at that time.[9]

As a follow-up to the 2005 census, McDowell broadened the potential study pool by using ROAR and membership lists from DSpace and bepress' Digital Commons repository software. She also conducted Google searches of all doctoral-granting institutions and the top ranked liberal arts colleges to locate as many repositories as possible regardless of institution size or focus. This study revealed that the IR movement was not limited to ARL or large doctoral-granting institutions. By late 2006, more than half of the repositories in the United States were at institutions with enrollments below 15,000 students and 53 percent of the seventy-three repositories were at non-ARL institutions.[10] A 2006 survey of academic library directors at four-year institutions found that 10.8 percent of the respondents (n = 446) had an established IR, and an additional 15.7 percent were actively planning to launch a repository.[11] The Bishoff Group in 2014 re-examined non-ARL institutions, noting that 81 percent of the respondents were collecting digital content, including some faculty and student research.[12]

Navigational Studies

Although many institutions register their repositories with directories such as ROAR or OpenDOAR, not all repositories are included in these directories and, even when they are, searchers may not be aware of them. As Crow commented in his early SPARC position paper, "For the repository to provide access to the broader research community, users outside the university must be able to find and retrieve information from the repository."[13] Coates used Google Analytics data to investigate how users were finding electronic dissertations at Auburn University. She compared navigational paths for local and out-of-state researchers. Local users found the dissertations using a variety of methods: links on the university's website, open search engines, or direct access to the dissertation. External users, however, discovered the dissertations mostly by using open search engines. This finding emphasizes the need for repositories to make their content as accessible as possible to web crawlers.[14]

Jantz and Wilson found that forty of sixty-three institutions that they examined provided a link to the repository from the library's website. Of those that included a link, the most common path was via the "scholarly communication" page, with the "for faculty" page as the next most common navigational path.[15] Mercer reported that although some libraries linked directly to the repository from their home page, many navigational paths require two to four links to reach the repository.[16] St. Jean's 2011 study used semi-structured interviews to understand how repository users located the site. Although respondents mentioned several discovery methods, the most common method reported was a direct link to the repository from the library's homepage, with a Google search being the second most common method. When asked why researchers might not use the repository, nearly two-thirds of the respondents noted the resource's lack of visibility. In fact, one respondent considered the IR to be a "well-kept secret."[17]

Repository Size and Platform

Lynch and Lippincott noted the difficulties in comparing repository sizes (number of items) since "it is clear that no two institutions are counting the same things."[18] This is especially true when comparing IRs using different software platforms; however, it has not prevented researchers from attempting size comparisons. In 2005, McDowell discovered a correlation between Carnegie classification and content size of the repository in an analysis of seventy-three repositories. Only the institutions with the highest research classification held more than 500 items in their entire collection.[19] By 2009, however, Nykanen located fourteen baccalaureate or master's institutions with repository counts greater than 500 items.[20]

There is general consensus that DSpace and Digital Commons are the two most frequently used platforms at American institutions. In studies where researchers reported software platform usage, DSpace installations ranged from 43 to 58 percent with Digital Commons implementations ranging from 21 to 27.8 percent of all platforms identified.[21]

Repository Content

Detailed analyses of IR content are sometimes hampered by platform interface differences and the institution's desire to organize and present its content in ways that reflect its organizational needs. Investigators have analyzed the type of faculty content, the percentage of faculty content compared to the repository as a whole, and faculty participation rates.[22] In addition to scholarly publications, non-research content such as teaching materials, university governance documents, campus history, etc. has also been considered.[23]

Studies have examined the size and variety of student content, particularly at institutions where teaching and student research are a priority. Some authors have conjectured that student scholarship provides visibility for undergraduate research and helps with repository growth.[24] Student contributions may include electronic theses, capstone projects, student research journals, undergraduate research presentations and posters, and specific course papers and projects.

Hertenstein discussed the effect that repository submissions may have on students' later attempts to get their scholarship accepted by traditional publishers.[25] Presenters at an Association of College and Research Libraries (ACRL) Conference shared comments from faculty mentors regarding student postings of preliminary research, and whether that preempts faculty from publishing final results in peer-reviewed journals. Faculty also questioned if repositories clearly differentiate between student and faculty authors.[26]

Master's-level Institutions

As previously noted, several studies have attempted large-scale investigations of repositories at non-ARL institutions. Many of these analyses include master's-level institutions but do not provide detailed breakdowns of size or content by institution type.[27] Case studies examining implementation at one specific institution are also available.[28] While individual studies are useful exemplars for others who are considering building or increasing the size of a repository and the larger census studies give a general idea of the status of repositories at non-research-intensive universities, none of them provides the details or context needed to consider the unique conflicts between teaching and research found at many master's-level institutions.

Method

The authors obtained a list of small, medium, and large master's-level institutions from the *Carnegie Classification of Institutions of Higher Education* and downloaded it into an Excel spreadsheet.[29] They created the list on March 6, 2015, and work began to ascertain how many of those institutions have an IR. Various definitions of repositories are found in the literature. The most regularly cited definition comes from Lynch's 2003 article introducing the concept of institutional repositories:

> A university-based institutional repository is a set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members. It is most essentially an organizational commitment to the stewardship of these digital materials, including long-term preservation where appropriate, as well as organization and access or distribution . . . a mature and fully realized institutional repository will contain the intellectual works of faculty and students—both research and teaching materials—and also documentation of the activities of the institution itself in the form of records of events and performance and of the ongoing intellectual life of the institution.[30]

To conduct analyses of comparable collections and using Lynch's definition as a guide, the authors created the following definition to direct the focus of this study:

> An online, institution-wide or consortial, multidisciplinary repository that includes scholarly works of faculty and students and may also include institutional history and documentation, institution-sponsored publications or partnerships, and other local digitized collections. Only those institutions showing a clear intent to include the scholarly pursuits of faculty and students are included.

Size was not necessarily a factor in the review if the repository included the criteria listed above. Many library websites provide a description of their physical or print collections or provide digitized finding aids to these collections. These were not included in the analysis since the collections themselves were not digitized. A large number of institutions have digitized special collections of images or text that are very narrow in scope and often related to local history or prominent local dignitaries. Although of value to the larger research community, these collections were not included in this analysis since they do not relate to the institution's scholarly output or administration.

The first review of all institutions was completed on April 29, 2015. Each institution on the list was examined to determine the existence of a repository that fit the authors' definition. A navigational analysis was performed based on the methods described by Jantz and Wilson.[31] The same procedure was followed to search for each repository and the

results of each step were recorded on the master spreadsheet. First, a Google advanced search was performed using the search strategy: "exact word" (institutional name) AND "any of these words" (repository archive). A well-cited research study by van Deursen and van Dijk reported that 91 percent of Google searchers do not go past the first page of results.[32] Based on that fact and the need to keep the navigation portion of this study manageable, only the first page of results was examined. Next, the OpenDOAR and ROAR directories were searched. Finally, the authors examined each institution's main site and the institution's library homepage to see if there were links to the repository. In addition to searching for a link on the main institutional webpage, the authors examined other institutional pages aimed at faculty, research, or general academics, plus an A to Z list or site index. If no repository was found using any of these steps, the final action was to conduct a keyword search of the entire institution's website for the terms "repository" or "archive." Again, only the first screen of results was examined.

Repositories were considered as discoverable in Google if they could be reached using no more than one link from Google. Sources that could not be located within one click of the initial Google search were excluded. Broken links on Google were not included in the discovery search for IRs. The repository name (if applicable) and its URL were recorded. The study found some independent institutions participating in what appeared to be a consortial or shared repository where each was able to present collections unique to their organization. Similarly, some of the master's institutions that are part of a multi-campus system shared the same platform, each with its own discrete collection of materials. These collections were included in the final analysis as long as the institution's collection could be accessed independently from the larger group.

If the steps described above failed to identify any semblance of a repository, the institution was recorded as lacking an IR. If an institution had a website or collection that required further investigation, this was recorded and a second review was conducted to carefully determine if the established criteria for this study were met. URLs that failed to open or resulted in the display of an error message after several attempts were not counted in the final analysis. Institutions located outside of the fifty United States, entities that had gone out of business, or those that appeared to have changed from a master's institution to another Carnegie classification were excluded from final consideration. The initial review of the 137 repositories that met the study's definition gathered basic descriptive information such as the software platform and a count of the total number of items in the collection, if it could be determined.

A navigational analysis of each library's website was conducted to locate links to the IR. When available, the following pages were examined: "about the library," "for faculty," scholarly communications, collections or resource lists, an A to Z list, digital collections, special collections, news and events, "finding information," and any discovery tools. Direct links including those from a pull-down menu, a persistent toolbar, or on the main page of a LibGuide were counted.

A more detailed qualitative content analysis of each repository was also conducted. Content types defined by earlier studies were employed in the analysis.[33] As software platform features may vary considerably, it can be difficult to determine if a particular type of content was included in the repository, much less quantify how many of a certain type of item were in the repository. For this reason, a qualitative approach seemed more practical. Therefore, if the authors found one faculty-authored journal article or if one student presentation or thesis was identified, the IR was marked as including that type of content. In addition to peer-reviewed papers, faculty content consisted of books, book chapters, conference presentations, reports, working papers, and data sets. Syllabi or other course-related teaching materials such as learning objects or assignments were also found in a few IRs. Student-generated content included theses (both honors and masters), capstone or class projects, poster sessions, and student journals.

Results

The total number of master's institutions downloaded from Carnegie was 724. Of these, institutions that were located in US territories or foreign countries (n = 15) were excluded, as were institutions that appeared to be out of business or whose Carnegie classification could not be verified (n = 7), resulting in twenty-two organizations eliminated from the initial download. Additionally, four institutions appeared to have an IR but the URL could not be opened after repeated attempts, making the final population equal to 698.

The search for IRs was conducted for the remaining 698 universities and colleges. The number of institutions with a working repository that met some of the study's criteria for a repository, was 190 or 27 percent of all of the institutions examined (190/698). Of the total IRs, however, 28 percent (53/190) lacked any type of faculty scholarship, which was the focus of this study. The final total of qualifying repositories that met the authors' definition of an IR and included faculty content numbered 137 (20 percent of 698). Table 1 illustrates the distribution of master's institutions according to type for the final set of repositories. Table 2 provides a comparative breakdown by student enrollment.[34]

This study also investigated how discoverable these IRs were using four possible avenues: Google, OpenDOAR, ROAR, and the institution's main website (see table 3). Overall, Google and ROAR provided the most access. In a cross-comparison of the IRs with faculty content, OpenDOAR

registered only one unique IR, i.e. one not discoverable by either a Google search or listed in ROAR. ROAR listed five unique IRs whereas the Google search discovered thirty-one unique IRs. One IR was only found by searching its institutional website. It follows that 72 percent (99/137) of the IRs could be found by more than one method. As illustrated in table 4, navigation to the IR on the library website is often more obvious, with most libraries including a link to the repository not only on their library homepage but also providing access through other library pages.

Table 1. Number of institutions with repositories

Carnegie type	Total number of master's institutions (*N* = 698)	Institutions with IRs having faculty content (*n* = 137)
Master's Large	405 (58%)	96 (70%)
Master's Medium	176 (25%)	30 (22%)
Master's Small	117 (17%)	11 (8%)
Total	698	137

Source: The Carnegie Classification. Basic Classification Methodology.

Table 2. IRs by institutional enrollment

Enrollment[i]	IRs with faculty content (*n* = 135)
Student population 0-5000	38 (28%)
Student population 5001-10000	46 (34%)
Student population 10001-15000	23 (17%)
Student population 15001-20000	13 (10%)
Student population over 20000	15 (11%)
Total[ii]	135

i. Source: Enrollment figures taken from National Center for Education Statistics.
ii. Two institutions did not provide student enrollment figures.

Table 3. Discoverability of IRs

Path Source	IRs with faculty content (*n* = 137)
Google	112 (82%)
ROAR	86 (63%)
OpenDOAR	50 (36%)
Campus website	18 (13%)

Table 4. Library website analysis of IRs with faculty content

Type of library webpage	Number of libraries with page type	IR link found on page
Library homepage	137	62% (85/137)
Digital Projects or Digital Collections page	70	60% (42/70)
Scholarly Communications page	38	58% (22/38)
Collections & Resources page or Database list	136	51% (70/136)
Special Collections page	115	48% (55/115)
"For Faculty" page	108	45% (49/108)
Services page	119	29% (35/119)
"About the library" page	132	25% (33/132)
News & Events page	124	23% (29/124)
"Finding Information" page or Discovery tool	134	20% (27/134)

Ten repository software platforms are represented in the set of 137 IRs containing faculty content, with Digital Commons being the most popular platform (80 or 58.4 percent). DSpace was the second most heavily used repository software (36 or 26.3 percent) with 15.3 percent using other repository solutions. Nineteen colleges and universities share platforms (13.9 percent) while 86.1 percent (118) maintain their own repository platform. Of those that share, eleven are master's large (57.9 percent), seven medium (36.8 percent), and one small (5.3 percent). IRs with faculty content are more likely to use Digital Commons than those lacking faculty content. Digital Commons offers a sophisticated software module expressly designed to store and display faculty profiles and content, which may account for this preference. The types of platforms used to support archives are summarized in table 5.

The total number of items in the repositories containing faculty scholarship ranged from 7 to 57,649. To be consistent, the total number of items provided by the software platform was recorded rather than a manual count of items. Five repositories' platforms did not generate item totals and are not included in the data presented. The most common type of scholarship found was the journal article, followed by presentations, books or book chapters, and reports. Although finding raw research data was more difficult to locate in collections, thirteen IRs contained obvious data files (see table 6). In reviewing types of student scholarship, theses and dissertations were the most common type of content. Capstone or class projects, distinct from theses and dissertations, were the second most common, followed by student journals and presentations (see table 7).

Other types of content, including syllabi, other course-related materials, and library working documents were also noted. University materials, such as minutes, policies, and guidelines, were defined as governance related. Newsletters, catalogs, yearbooks, reports, and other types of university publications were classified separately. Any type of media collection (e.g. images, photos,

Table 5. Software platform comparison

Software platform	All IRs (*n* = 190)	IRs with faculty content (*n* = 137)[i]
Digital Commons	86 (45%)	80 (58.4%)
DSpace	59 (31%)	36 (26.3%)
Web-based program	8 (4%)	8 (5.8%)
ContentDM	26 (14%)	7 (5%)
Islandora	3 (2%)	1 (0.73%)
Ebrary	1 (0.5%)	1 (0.73%)
Omeka	1 (0.5%)	1 (0.73%)
SelectedWorks (bepress)	1 (0.5%)	1 (0.73%)
Open Repository	1 (0.5%)	1 (0.73%)
ArchivalWare	1 (0.5%)	1 (0.73%)
Eprint3	1 (0.5%)	---
ContentPro	1 (0.5%)	---
Irplus	1 (0.5%)	---

i. Totals may not equal 100% because of rounding.

maps, or audio files) was recorded. Each repository was examined to see if it hosted one or more external journals (see table 8).

Discussion

Census

In one of the earliest censuses, only two of the liberal arts consortial members of the CNI group had an established IR in 2005.[35] A broader study in 2006, however, discovered that 19 percent of the master's-level institutions sampled had already implemented an IR and 32 percent were in the process of implementing one.[36] In the current study of all master's institutions, 27 percent (190/698) had a working IR of any type and 20 percent (137/698) had an IR with faculty content.

McDowell used ROAR and open web searches, along with directories from the major IR software vendors, to compile a list of active IRs. Her 2006 search located seventy-three active IRs with 47 percent of those coming from ARL institutions. McDowell also noted that more than half of the IRs were located at academic institutions with student enrollments below 15,000.[37] This project discovered that 79 percent of the IRs with faculty content were supported by institutions with student populations below 15,000.

In this study, the collection sizes ranged from a low of seven items to a high of 57,649, with a mean collection size of 4,538 and a median of 1,822. This appears to be commensurate with numbers and averages reported in the literature.

Table 6. Faculty content by type (*n* = 137)

Type of faculty scholarship	At least one record in IR
Journal article	126 (92%)
Presentations, etc.	108 (79%)
Book or book chapters	95 (69%)
Reports	90 (66%)
Data	13 (9%)

Table 7. Student content by type (*n* = 125)

Type of student content	At least one record in IR
Theses	116 (93%)
Projects	79 (63%)
Student journal	64 (51%)
Presentations	60 (48%)

Table 8. Other Content (*n* = 137)

Content type	At least one record in IR
Course syllabi	17 (12%)
Other course materials	35 (25.5%)
Library-related documents	66 (48%)
University governance	67 (49%)
University publications	87 (63.5%)
Media collections	89 (65%)
Hosted external journals	51 (37%)

For example, Nykanen's 2009 study showed an average of 2,968 items.[38] Xia and Opperman in 2009, examining master's and baccalaureate institutions, saw a range from four to 7,573 items.[39] Mercer's review of faculty content at ARL institutions found a wide variety in size with a range of eleven to 46,823 items.[40]

Location and Navigation

A perceived lack of discoverability was noted by Davis and Connelly in interviews with several Cornell faculty who saw the IR as "a single island completely isolated from other institutional repositories."[41] Good metadata and navigational links allow users from any location to find IR content. The current study indicates that IRs are more visible when links are provided on a variety of library webpages, including the homepage. Scholarly communications, faculty, and collections pages continue to be popular gateways to the IR, but more libraries are now adding links on general library pages such as those devoted to services, news, or "about the library." See table 4 for more information.

Table 9. Software platform comparison

	Current study IRs with faculty content (*n* = 137, 2015 data)	Current study all IRs (*n* = 190, 2015 data)	Mercer (*n* = 72, 2009 data)	Nykanen (*n* = 14, 2007 data)	Rieh, (*n* = 446, 2006 data)	Lynch & Lippincott (*n* = 38, 2005 data)
Digital Commons	58%	45%	27.8%	50%	26.8%	21%
DSpace	26%	31%	56.9%	43%	46.4%	58%
Web-based	6%	4%	---	---	---	---
ContentDM	5%	14%	4.2%	---	4.9%	---
Islandora	1%	2%	---	---	---	---
Other	4%	4%	11.1%	7%	21.9%	---

Sources: Mercer, et al., "Structure, Features, and Faculty Content," 335; Nykanen, "Institutional Repositories at Small Institutions," 11; Rieh, et al., "Census of Institutional Repositories," 9–10. (implemented IRs); Lynch and Lippincott, "Institutional Repository Deployment," 6.

Platform

In Hertenstein's 2013 survey (n = 36) of institutions with established IRs, 43 percent were using DSpace.[42] Jantz and Wilson's 2009 study reported DSpace as the most common platform with bepress as the second choice.[43] Xia and Opperman's 2009 study of fifty IRs at master's and baccalaureate institutions also found that DSpace was used most often, followed by Digital Commons.[44] In contrast, this study found the Digital Commons software (a bepress product) to be much more heavily used than DSpace confirming that Digital Commons and DSpace continue to dominate IR software implementations. Additional studies by Mercer, Nykanen, Rieh, and Lynch allowed direct comparisons to the current study of platform use (see table 9).

Content: Faculty

This study provides a qualitative review of the types of faculty content in 137 master's IRs (see table 6), and is similar in nature to the overall content of faculty collections described in other studies. Because of the size of the population, quantitative data on the number of items of each faculty content type were not collected here; therefore, the data is not directly comparable to the quantitative data included in some smaller studies.[45] A future study of a small, randomly selected subset of master's IRs would enable counts of faculty items, thus providing comparable data.

Content: Student

Rozum's 2014 survey of librarians working with IRs that contain student content concluded that "libraries are somewhat passive collectors of student research," willing to take student content but not seeking it in the same way that they push for faculty content.[46] While this may be true, other studies have reported that student contributions at master's and baccalaureate repositories account for a large percentage of the overall content.[47] In 2013, Hertenstein's survey discovered that 92 percent of the institutions with IRs included student content.[48]

Although the content analysis of this study was limited to IRs that contained faculty scholarship, like Hertenstein, some type of student content was present in 91 percent (125/137) of the IRs. The largest category of student content was theses (93 percent). Fifty-one percent of the IRs hosted some type of student research journal. The results of this study are similar to those found in a 2013 study of student content. In the earlier report, 85 percent of the IRs contained theses or dissertations and 45 percent provided access to student presentations or posters. There appears to be a slight increase in the inclusion of student class papers and projects with 63 percent (79/125) of the current IRs containing these materials compared to 39 percent of the IRs examined in 2013.[49]

Content: Other

McDowell found that 4.5 percent of the IRs in her study consisted of non-scholarly content including marketing materials and university governance documents.[50] These materials were a larger part of IRs at institutions with less than 10,000 students, comprising 16.9 percent of the content.[51] In this study, over 48 percent (66/137) of the IRs contained library materials and 49 percent (67/137) had some sort of university-related governance materials. Syllabi were included in 12 percent (17/137) of the current IRs and course-related materials were present in 26 percent (35/137).

Conclusion

This study benchmarks IR development in Carnegie-designated master's institutions. Since no other research

published to date has examined this exact population, speculating on the growth of IRs in this segment of the academic community is difficult. Rieh's early study of 446 four-year institutions found that 118 respondents either had or were actively planning IRs.[52] In 2014, Bishoff and Smith reported that 117 (81 percent) of the two-year and four-year master's and doctorate institutions in their study maintained IRs.[53] Rather than looking at a sample, this project investigated all Carnegie-designated master's institutions. Within this population of 698 institutions, the 137 IRs with faculty content and 190 total IRs seem to indicate at least some kind of growth over the last ten years.

The nature of the content appears very similar to that found in other study populations, whether at teaching or research institutions. In general, it appears that faculty scholarship, primarily journal articles and presentations, continues to represent an important part of most repositories. Student content is still primarily theses; other types of student productivity, however, such as student projects and presentations, are also included. This study indicates there may be an increasing interest in content beyond faculty peer-reviewed books and articles. In the current review, 66 percent of the IRs contained faculty working papers and technical reports.

A 2009 study of fifty master's and baccalaureate institutions was unable to locate much in the way of teaching materials and found just one IR that contained syllabi.[54] In this analysis, course syllabi were included in 12 percent of the IRs and nearly 26 percent had other kinds of course-related materials. Nykanen's examination of the content in ten repositories found that 16.9 percent of the overall content was devoted to university documentation and marketing materials, much of which was produced by the library.[55] Some degree of university governance documents and library materials appeared in nearly half of the IRs in this study.

Examining the discoverability of IRs with faculty content, Google searching appears to be the most successful way to discover IRs and produced the most unique number of IRs, i.e., those not found elsewhere. The ROAR directory consistently included more repositories than OpenDOAR, and had a larger number of unique entries than OpenDOAR. IR visibility also appears to be increasing on library webpages with 62 percent (85/137) of the libraries in this study including a link to the IR on their library homepage as compared to only four (n = 40) libraries of those analyzed in 2006.[56]

The current study represents a snapshot in time and the creation and development of IRs is continually changing. Different platforms and even IR organizational structure make direct comparisons on size and content difficult. That said, additional analyses of content, such as full-text versus bibliographic content, comparisons by discipline, etc., would be useful.

In one of the earliest papers describing the potential of IRs, Lynch commented, "Not every higher education institution will need or want to run an institutional repository, though I think ultimately almost every such institution will want to offer some institutional repository services to its community."[57] This report offers some quantitative and qualitative evidence that less than 20 percent of the master's institutions in the United States have established repositories with faculty content, but those that do, contain content similar to those other types of institutions previously examined.

References

1. Clifford A. Lynch and Joan K. Lippincott, "Institutional Repository Deployment in the United States as of Early 2005," *D-Lib Magazine* 11, no. 9 (2005): 1–10, accessed February 21, 2015, http://webdoc.sub.gwdg.de/edoc/aw/dlib/.
2. Philip M. Davis and Matthew J.L. Connolly, "Institutional Repositories: Evaluating the Reasons for Non-use of Cornell University's Installation of DSpace," *D-Lib Magazine* 13, no. 3–4 (2007): 1–17, accessed September 1, 2014, http://www.dlib.org/dlib/march07/davis/03davis.html; Ronald C. Jantz and Myoung C. Wilson,"Institutional Repositories: Faculty Deposits, Marketing, and the Reform of Scholarly Communication," *Journal of Academic Librarianship* 34, no. 3 (2008): 186–95, https://doi.org/10.1016/j.acalib.2008.03.014; Holly Mercer et al., "Structure, Features, and Faculty Content in ARL Member Repositories," *Journal of Academic Librarianship* 37, no. 4 (2011): 333–42, https://doi.org/10.1016/j.acalib.2011.04.008.
3. Liz Bishoff and Carissa Smith,"Managing Digital Collections Survey Results," *D-Lib Magazine* 21, no. 3–4 (2015): 1–7, accessed June 5, 2015, http://www.dlib.org/dlib/march15/bishoff/03bishoff.print.html; Davis and Connolly, "Institutional Repositories"; Elizabeth Hertenstein,"Student Scholarship in Institutional Repositories," *Journal of Librarianship & Scholarly Communication* 2, no. 3 (2014): eP1135, accessed October 6, 2016, http://jlsc-pub.org/articles/abstract/10.7710/2162-3309.1135/; Karen Markey et al., "Institutional Repositories: The Experience of Master's and Baccalaureate Institutions," *portal: Libraries & the Academy* 8, no. 2 (2008): 157–73, https://doi.org/10.1353/pla.2008.0022; Soo Young Rieh et al., "Census of Institutional Repositories in the U.S.: A Comparison Across Institutions at Different Stages of IR Development," *D-Lib Magazine* 13, no. 11–12 (2007): 1–13, accessed September 2, 2014, http://dlib.org/dlib/november07/rieh/11rieh.html.
4. Ellen Dubinsky, "A Current Snapshot of Institutional Repositories: Growth Rate, Disciplinary Content and Faculty Contributions," *Journal of Librarianship & Scholarly Communication* 2, no. 3 (2014): eP1167, accessed January 16, 2015, http://jlsc-pub.org/articles/abstract/10.7710/2162-3309.1167/;

Jantz and Wilson, "Institutional Repositories Faculty Deposits," 186–95; Cat S. McDowell, "Evaluating Institutional Repository Deployment in American Academe since Early 2005: Repositories by the Numbers, Part 2," *D-Lib Magazine* 13, no. 9–10 (2007): 1–14, https://doi.org/10.1045/september2007-mcdowell; Mercer et al., "Structure, Features, and Faculty Content"; Melissa Nykanen, "Institutional Repositories at Small Institutions in America: Some Current Trends," *Journal of Electronic Resources Librarianship* 23, no. 1 (2011): 1–19, https://doi.org/10.1080/1941126X.2011.551089; Betty Rozum et al., "We Have Only Scratched the Surface: The Role of Student Research in Institutional Repositories," ACRL 2015 Conference. Portland, OR: Association of College and Research Libraries. (2015): 804–12, accessed May 4, 2016, http://www.ala.org/acrl/sites/ala.org.acrl/files/content/conferences/confsandpreconfs/2015/Rozum_Thoms_Bates_Barandiaran.pdf; Jingfeng Xia and David B. Opperman, "Current Trends in Institutional Repositories for Institutions Offering Master's and Baccalaureate Degrees," *Serials Review* 36, no. 1 (2010): 10-18, https://doi.org/10.1080/00987913.2010.10765272; Hong Xu, "The Current Situation of Faculty Participation in Institutional Repositories—A Study of 40 DSpace Implementations Supporting IRs," *Proceedings of the American Society for Information Science and Technology* 44, no. 1 (2007): 1–3, https://doi.org/10.1002/meet.1450440332.

5. Markey et al., "Institutional Repositories: The Experience," 159; Dubinsky, "A Current Snapshot," 17–18.
6. Markey et al., "Institutional Repositories," 167; McDowell, "Evaluating Institutional Repository Deployment," 10–11; Nykanen, "Institutional Repositories at Small Institutions," 13; Yuji Tosaka, Cathy Weng, and Eugenia Beh, "Exercising Creativity to Implement an Institutional Repository with Limited Resources," *Serials Librarian* 64, no. 1 (2013): 254–62, https://doi.org/10.1080/0361526X.2013.761066.
7. Indiana University Center for Postsecondary Research, *The Carnegie Classification of Institutions of Higher Education Interim Site*, accessed March 6, 2015, http://carnegieclassifications.iu.edu/.
8. University of Southampton School of Electronics and Computer Science, *Registry of Open Access Repositories (ROAR)*, accessed May 21, 2016, http://roar.eprints.org/; University of Nottingham Centre for Research Communications, *The Directory of Open Access Repositories—OpenDOAR*, accessed May 21, 2016, www.opendoar.org/.
9. Lynch and Lippincott, "Institutional Repository Deployment," 2–4.
10. McDowell, "Evaluating Institutional Repository Deployment," 4–5.
11. Rieh et al., "Census of Institutional Repositories," 3.
12. Bishoff and Smith, "Managing Digital Collections," 2.
13. Raym Crow, "The Case for Institutional Repositories: A SPARC Position Paper," *ARL Bimonthly Report*, no. 223 (August 2002): 5, accessed June 9, 2015, http://sparcopen.org/wp-content/uploads/2016/01/instrepo.pdf.
14. Mildred Coates, "Electronic Theses and Dissertations. Differences in Behavior for Local and Non-Local Users," *Library Hi Tech* 32, no. 2 (2014): 285–99, https://doi.org/10.1108/LHT-08-2013-0102.
15. Jantz and Wilson, "Institutional Repositories Faculty Deposits," 192–93.
16. Mercer et al., "Structure, Features, and Faculty Content," 335.
17. Beth St. Jean et al., "Unheard Voices: Institutional Repository End-Users," *College & Research Libraries* 72, no. 1 (2011): 29–30, 35, https://doi.org/10.5860/crl-71r1.
18. Lynch and Lippincott, "Institutional Repository Deployment," 4.
19. McDowell, "Evaluating Institutional Repository Deployment," 6.
20. Nykanen, "Institutional Repositories at Small Institutions," 9.
21. Hertenstein, "Student Scholarship," 4; Lynch and Lippincott, "Institutional Repository Deployment," 6; Mercer et al., "Structure, Features, and Faculty Content," 335; Rieh et al., "Census of Institutional Repositories," 9-10.
22. Nykanen, "Institutional Repositories at Small Institutions"; McDowell, "Evaluating Institutional Repository Deployment"; Xu, "The Current Situation."
23. Anne M. Casey, "Does Tenure Matter? Factors Influencing Faculty Contributions to Institutional Repositories," *Journal of Librarianship & Scholarly Communication* 1, no. 1 (2012): eP1032, accessed May 13, 2016, https://doi.org/10.7710/2162-3309.1032; Lynch and Lippincott, "Institutional Repository Deployment," 5-6; Xia and Opperman, "Current Trends in Institutional Repositories," 14.
24. Hertenstein, "Student Scholarship," 11; Nykanen, "Institutional Repositories at Small Institutions" 14, 17; Rozum et al., "We Have Only Scratched the Surface," 811.
25. Hertenstein, "Student Scholarship," 7.
26. Rozum et al., "We Have Only Scratched the Surface," 810.
27. Bishoff and Smith, "Managing Digital Collections"; Dubinsky, "A Current Snapshot."
28. Jonathan Bull, Jonathan, and Bradford Lee Eden, "Successful Scholarly Communication at a Small University: Integration of Education, Services, and an Institutional Repository at Valparaiso University," *College & Undergraduate Libraries* 21, no. 3–4 (2014): 263–78, https://doi.org/10.1080/10691316.2014.93226; Gregory J. Kocken and Stephanie H. Wical, "'I've Never Heard of it Before': Awareness of Open Access at a Small Liberal Arts University," *Behavioral & Social Sciences Librarian* 32, no. 3 (2013): 140–54, https://doi.org/10.1080/01639269.2013.817876.
29. Carnegie Commission, "Interim Site."
30. Clifford A. Lynch,"Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age," *portal: Libraries and the Academy* 3, no. 2 (2003): 328, https://doi.org

/10.1353/pla.2003.0039.

31. Jantz and Wilson, "Institutional Repositories: Faculty Deposits," 190.
32. Alexander J. A. M. van Deursen and Jan A. G. M. van Dijk, "Using the Internet: Skill Related Problems in Users' Online Behavior," *Interacting with Computer* 21, no. 5–6 (2009): 398, https://doi.org/10.1016/j.intcom.2009.06.005.
33. Lynch and Lippincott, "Institutional Repository Deployment," 5–6; McDowell, "Evaluating Institutional Repository Deployment," 9–10.
34. Carnegie Commission of Institutions of Higher Education, *The Carnegie Classification of Institutions of Higher Education Basic Classification Methodology*, accessed May 21, 2016, http://carnegieclassifications.iu.edu/methodology/basic.php; Institute of Education Sciences, National Center for Education Statistics. *IPEDS, 2013-2014 Final Data*, accessed March 30, 2016, https://nces.ed.gov/ipeds/datacenter/Default.aspx.
35. Lynch and Lippincott, "Institutional Repository Deployment," 4.
36. Markey et al., "Institutional Repositories: The Experience," 162.
37. McDowell, "Evaluating Institutional Repository Deployment," 4-5.
38. Nykanen, "Institutional Repositories at Small Institutions," 9.
39. Xia and Opperman, "Current Trends in Institutional Repositories," 14.
40. Mercer et al., "Structure, Features, and Faculty Content," 334.
41. Davis and Connolly, "Institutional Repositories," 13.
42. Hertenstein, "Student Scholarship," 4.
43. Jantz and Wilson, "Institutional Repositories: Faculty Deposits," 192.
44. Xia and Opperman, "Current Trends in Institutional Repositories," 12.
45. McDowell, "Evaluating Institutional Repository Deployment," 10; Nykanen, "Institutional Repositories at Small Institutions," 13; Xia and Opperman, "Current Trends in Institutional Repositories," 14.
46. Rozum et al., "We Have Only Scratched the Surface," 811.
47. McDowell, "Evaluating Institutional Repository Deployment," 10; Nykanen, "Institutional Repositories at Small Institutions," 13; Xia and Opperman, "Current Trends in Institutional Repositories," 12.
48. Hertenstein, "Student Scholarship," 4.
49. Ibid, 5.
50. McDowell, "Evaluating Institutional Repository Deployment," 10.
51. Nykanen, "Institutional Repositories at Small Institutions," 13.
52. Rieh et al., "Census of Institutional Repositories," 3.
53. Bishoff and Smith, "Managing Digital Collections," 2.
54. Xia and Opperman, "Current Trends in Institutional Repositories," 16.
55. Nykanen, "Institutional Repositories at Small Institutions," 13.
56. Jantz and Wilson, "Institutional Repositories: Faculty Deposits," 193.
57. Lynch, "Institutional Repositories: Essential Infrastructure," 335.

Notes on Operations

IGAPS: A Taxonomy and Facet Classification System

John Pell and Meghan Huppuch

This paper describes an assessment of the information management practices at International Planned Parenthood Federation/Western Hemisphere Region and the development and implementation of an information management pilot for that organization. The pilot included the development of a taxonomy to classify the organization's documents, training in basic citation practices, and a decentralized model for building an organized library of documents within the citation management software Mendeley. The authors discuss the pilot's taxonomy within the context of literature on taxonomy development and offers strategic recommendations for improving the information management practices of not-for-profit organizations that lack dedicated information management staff.

International Planned Parenthood Federation/Western Hemisphere Region (IPPF/WFR), the not-for-profit organization that sponsored this project phased out its library in 2011 when the librarian it employed left to accept another position. The organization did not seek a replacement, rationalizing that the information it needed for research and publications was easily accessible online for free. The organization's communications and evaluations officers soon noticed the development of difficulties with the retrieval of research material and problems with complete and accurate citation in organizational publications. Accordingly, they suggested the following two main objectives for this project: 1) develop a taxonomy to facilitate access to journal and grey literature, and 2) train staff to use of the taxonomy and citation management software to increase their ability to systematically document, track, and locate citations of data and literature in speeches, presentations, papers, and internal publications. The organization's communications and evaluations officers tasked a temporary information management specialist to carry out these objectives. An academic librarian provided pro bono consultation and training for the communications and evaluations officers and information management specialist. This paper provides an overview of the literature on taxonomy development that informed the approach to taxonomy development taken at the organization, a description of the methods used to assess the need for information management at the organization, and a discussion of the results of the project.

Literature Review

A search of Library and Information Science Source, Library and Information Science Abstracts, Business Source Premier, and the Directory of Open access journals retrieved 924 papers published between 2004 and 2015 that include the word "taxonomy" in the title; none of these papers described developing a taxonomy for use within a citation manager. After excluding papers that were primarily descriptions of taxonomies in domains unrelated to women's reproductive health or that described automated methods of taxonomy production based on

John Pell (jpell@hunter.cuny.edu) is an Assistant Professor at Hunter College Library in New York, New York. **Meghan Huppuch** (meghan.huppuch@gmail.com) is a communications professional at a philanthropic foundation.

Manuscript submitted August 4, 2016; returned to authors for revision September 20, 2016; revised manuscript submitted November 16, 2016; manuscript accepted for publication March 1, 2017.

inaccessible tools, the search yielded twenty-two papers that dealt explicitly with methods of taxonomy creation or assessment, or description of taxonomies related to the domain of women's health. This review uses those papers to accomplish three functions: define the features of taxonomies, describe methods of producing taxonomies, and analyze the costs and benefits of using and creating taxonomies.

Defining Taxonomy

A taxonomy is commonly understood as a system of knowledge organization that is closely related to the practice of classification. Confusion may ensue in discussions of taxonomy as systems for knowledge organization are described variously in the literature of information science and business as classifications, frameworks, typologies, taxonomies, and ontologies that may vary in the complexity of their structure from flat lists of terms to densely interconnected hierarchies with arrays of branching subcategories.[1]

Some of these differences in description are interpreted as different approaches to analysis of the same underlying relationships. For example, in the National Library of Medicine's (NLM) Medical Subject Heading (MeSH) database, "classification" is listed as a child element, or subcategory, of both "information science" and "documentation." A taxonomist could represent the relationships among these terms in the form of a network in which "classification" has multiple parent elements. Alternatively, a taxonomist could represent "classification" as existing in multiple hierarchical lists. Nickerson, Varshney, and Munterman describe taxonomies as defined sets of objects. These objects have dimensions and the dimensions in turn have defined characteristics. They posit two restrictions on taxonomies requiring that characteristics be mutually exclusive, meaning that each dimension should have exactly one characteristic, and collectively exhaustive, meaning that each object should have a characteristic for every one of its dimensions. They explicitly invoke Miller's "Magical Number Seven, Plus or Minus Two" as a possible objective criteria related to their requirement that a taxonomy be concise.[2] Neelameghan and Rhagavan extend this argument for the significance of Miller's number in classification schemes through a broad overview of knowledge organization drawing examples from Vedanta philosophy, religious mysticism, and modern approaches to systems development; they note that prominent classification systems that persist over time have between five and nine top-level categories.[3]

A taxonomy may be as simple as a list of terms, but that does not imply it is neutral with regard to social values or theoretical perspectives. Classification can raise ethical issues when determinations about the scientific merit of materials are in question, such as when a college library director was asked to classify creationist materials as "science" rather than "religion" in the college's library.[4] A theoretical perspective may drive taxonomy development and help to further the development of a field, as McKinney and Yoos intended their taxonomy of views on information to help advance the field of information science.[5]

While agreement about what is meant by "taxonomy" may further the development or unification of a discipline, differences in perspective or purpose about a taxonomy's entities may divide disciplines. A historic example of this type of difference of perspective is found in the phrenetic and cladistic approaches to taxonomy development in systematics: the phrenetic approach groups organisms on the basis of shared characteristics and the cladistic approach groups organisms on the basis of shared ancestry. The phrenetic approach dominated when physical observation was the primary research method in the field, but as techniques for analyzing genetic distance emerged, cladistic taxonomies were developed to focus on the common ancestry of group of organisms. The schism between scientists using these two taxonomies engendered not only scientific debates, but also impassioned disputes over bias toward one form of taxonomy or the other in the editorship of scientific journals.[6]

Another more recent example is from scientometrics in the form of taxonomic disputes over citation metrics. Bornman follows Kuhn's theory to characterize scientific revolutions as taxonomic changes in a research field and proposes that such a revolution is currently underway in the field of scientometrics.[7]

If the relatively stolid fields of systematics and scientometrics are vulnerable to controversy and upheaval through disputes over taxonomy, a taxonomist would be well advised to tread carefully in a field that carries more politicized controversy, such as reproductive health. In her taxonomy and framing analysis of abortion weblogs, Park distinguishes between advocacy versus objectivist framings of the blog posts.[8] While such distinctions may be made with relative ease academically, considering research that suggests perceptions of objectivity may vary with an individual's political bias, classifying information as advocacy or objectivity could prove controversial in an organization with politically diverse views, or it could seem to have limited usefulness for an organization that views its purpose as closely aligned with advocacy.[9]

Despite the controversies that may ensue, theoretical basis is not necessarily a vulnerability to avoid in taxonomy development; rather, it may be a goal to strive toward. As taxonomic classifications can serve as the basis for organizational decision-making, it is important that those classifications are based in meaningful characteristics. Tezanos, Vazquez, and Sumner criticize the use of per capita income as the criteria for defining developing nations because such classification lacks a sound theoretical basis. They point out

that, even absent a developed and agreed upon theory, a multidimensional taxonomy that takes more entities and characteristics into account will provide greater utility for decision-making and theory development.[10]

Additionally, taxonomies may differ in more pragmatic terms regarding whether they classify their objects by physical characteristics, historical origins, or intended uses. User intent for interacting with a taxonomic entity is often an important consideration in taxonomy development. The RoMEO taxonomy for copyright transfer agreements categorizes agreements based on which format of publication an author may distribute and how and to where the author may distribute it.[11]

Doria's theoretical taxonomy of document uses included eight categories: individual work document, collaborative work document, project monitoring document, trade document, auxiliary resource document, referential document, external document, and record document. While Doria noted that this theoretical taxonomy could be applied to any department in an organization, her empirically developed taxonomy of a document collection from an engineering firm's research department produced fifty-seven categories, including budget, needs analysis, and case scenario; although Doria's theoretical categories promised potential, they required significant modification to meet the needs of actual users within a specific context.[12]

Taxonomy Development Methods

Nickerson, Varshney, and Munterman reported that over forty of the papers that they surveyed in their review of taxonomy development literature did not report a methodology for taxonomy development. Reports that described a method classified it as inductive, deductive, or intuitive, and observed that nearly a third of these reports used an intuitive, or ad hoc, approach to develop the taxonomy.[13]

As an alternative to ad hoc approaches, Nickerson, Varshney, and Munterman present a four-stage model of taxonomy development that incorporates recursive processes. This model begins with the determination of a "meta-characteristic" that serves to dictate the characteristics to be included in the taxonomy. Following the determination of the meta-characteristic, ending conditions for the development of the taxonomy are set. Development then proceeds through recursive stages of identification, modification, and evaluation until the previously defined ending conditions are met. Even the meta-characteristic may not clearly emerge until multiple iterations of the development approach have conducted.

In its study of the Functional Requirements of Bibliographic Records (FRBR), the International Federation of Library Associations and Institutions (IFLA) applied an entity analysis technique. At a basic level, entity analysis consists of isolating the entities of interest to users in a particular domain. These entities are defined in a way that focuses on the entities themselves, rather than the data about them.[14] An example from the domain of reproductive health would be women with cervical cancer, as opposed to statistics about women with cervical cancer. After entities are defined, each entity's characteristics may be enumerated. This method of entity analysis may be extended to include relationships between entities and user tasks.

In a paper outlining the basic phases and best practices of taxonomy development, Cisco identifies the four basic phases of taxonomy development as planning and analysis; design, development, and testing; implementation; and maintenance. Cisco's best practice are: keeping the taxonomy closely related to the organizational strategy, incorporating existing taxonomy and metadata, making categories well-defined and distinct, developing the taxonomy in an iterative process, and providing for adequate resources to maintain the taxonomy.[15]

Since taxonomies are only structured lists of terms, the terms included in the taxonomy can strongly influence the taxonomy's usefulness. It is important to decide what morphological form to use in the taxonomy. Faith's linguistic analysis of taxonomy recommends avoiding terms such as company or brand names, organizational titles, or acronyms that are subject to sudden change or confusion; however, this recommendation is tempered by her acceptance of the principle that scope is the key factor in determining which terms belong in a taxonomy.[16] Scope, in turn, is defined by the context of the taxonomy's intended users.

While user input is important in the taxonomy development process, basing the taxonomy on a limited group of users can restrict the taxonomy's utility in more diverse contexts. Alexander offers an approach to assess the process of decision-making in taxonomy creation projects that uses four criteria to characterize the taxonomy's objectivity and subjectivity. Those criteria are: openness to criticism, responsiveness to criticism, public accessibility of standards, and equality of intellectual authority. She uses the metaphor of "taxonomer as politician" to explain how taxonomy development balances these criteria to achieve objectivity, which Alexander characterizes as "open intersubjectivity."[17] On platforms that support widespread collaboration, collaborative tagging, or user supplied tagging, can be used to generate a corpus of meaningful terms to serve as the basis for a more structured taxonomy; however, lack of guidelines and variation in background knowledge make it difficult to reuse collaboratively assigned tags.[18] Although there are many approaches to taxonomy creation, most of the reports in this review stress the importance of user input in the taxonomy development process. There may not be a single best method of taxonomy creation; a taxonomy's usability and usefulness are ultimately determined by the users.

Costs and Benefits of Taxonomies

Well-designed taxonomies can enable efficient retrieval of relevant information. Two important measures of information retrieval are precision (the percentage of relevant documents in a search result set) and recall (the amount of relevant documents in a result set expressed as a percentage of the total available number of relevant documents). It is not unusual for database searches to miss a great deal of relevant information (low recall) and to return a great deal of irrelevant information (low precision). Haig reports on group searches that ranged from capturing 6.5 to 19.6 percent of the available relevant documents, with most search result sets consisting of a small percentage of relevant papers, 6 to 30 percent.[19] In contrast, Wang found recall scores ranging from 62.5 to 87.1 percent and precision scores ranging from 41.6 to 97.4 percent when evaluating the navigation effectiveness of a taxonomy for a library and information science school.[20]

Precise searches mean less time searching. High recall means more complete use of available information. In a paper highlighting the return on investment for taxonomy development, Ekionea and Swain align this capacity of taxonomy to increase efficient and effective use of resources with successful and sustainable business strategies.[21] Classification can also play an important role in information retrieval systems designed to answer questions; it is the first step in the process of connecting information with a purpose.[22] The utility of a taxonomy is not restricted to information retrieval. It can also perform an important role in knowledge transfer: browsing or studying a taxonomy can provide a user with subject knowledge, especially in the case of highly developed and specialized taxonomies.[23]

Although a fully developed and implemented taxonomy can be a timesaving resource for an organization, constructing taxonomies is very time consuming. It is common to consider using existing resources whenever possible; however, the specific context of an organization's purpose may not be reflected in an existing taxonomy. Haig et al. evaluated nine thesauruses related to medicine, education, and medical education and found them insufficient for describing medical education in the United Kingdom.[24]

Even when a taxonomy is established, its implementation may be very time consuming, particularly if consistency is a concern. In Park's taxonomy of weblogs, it took seventy hours of training to achieve acceptable interrater reliability among the seven coders using the taxonomy.[25]

Method

The taxonomy development team consisted of the organization's communications and assessment officers, the information management specialist, and the librarian consultant. The team adopted Cisco's best practices for taxonomy development as the project's guidelines. The first step was to determine existing organizational strategies for accessing, organizing, and applying information. However, without a directly applicable example in the literature for obtaining the user input to meet this guideline, the information management specialist devised methods to solicit user input through staff interviews and an organization-wide survey, and conducted interviews with staff at organizations with a similar focus on reproductive health.

Staff Interviews

The communications and evaluations officers at IPPF/WHR selected ten staff members to represent a cross-section of roles, practices, and challenges in the areas of access to the literature, citation, and reference management. Staff members received meeting invitations and a short explanation of the needs assessment, and voluntarily participated in one-on-one interviews that lasted between thirty and fifty minutes. Interview questions probed for information about practices for finding, citing and tracking sources; determining trustworthiness; and sharing information with colleagues. The organization's communications officer vetted and edited the interview questions.

Organization-wide Survey

The survey goal was to understand the organization's need for access to subscription-based resources. The information management specialist obtained a list of thirty-five recommended publications through meetings and informal staff interviews. All staff members received a questionnaire requesting information about which of these recommended publications they subscribed and what publications they wanted to use but to which they lacked access. The information management specialist checked which of the desired journals were available either through the public library or as open access publications.

Interviews with Staff at Related Organizations

The communications and evaluation team suggested six related organizations to contact to get a sense of current practices, systems, and software used by organizations in the area of international sexual and reproductive health and rights. Interview questions focused on current citation management systems and institutional access to literature. Follow-up questions compared the effectiveness of approaches and explored recommendations for effective practice. Interviews lasted between twenty and forty minutes.

Results

Access to the Literature

Staff had varying and extremely limited access to the literature. Many respondents reported encountering pay walls when trying to obtain needed papers. Staff members reported using the logins of interns, friends, or partners with a university library affiliation to access peer-reviewed papers. While no one expressed a desire to use peer-reviewed journals as their main source of data, many ranked them among their most trusted sources. Overall, these conditions limited staff knowledge of the literature in the field, as there was no centralized process for obtaining relevant papers.

Citation and Data Management

There was not a systematic procedure for documenting citations referenced in speeches, presentations, and publications. Staff members were often unable to locate source documents if they wanted to cite them later or if questions arose about the quality or accuracy of the source information. Without organization-wide citation practices, work shared with the public could not be consistently supported with references to evidence, which could potentially weaken the organization's credibility.

Data permeated every staff members' work and played crucial roles in both internal communications and how the staff represented the organization to its many stakeholders. Data use varied greatly by position and responsibilities and was reportedly used for reasons as varied as blog posts, speeches, statements to reporters, publications, conferences, papers, proposals, decision-making (strategy, focus, inclusion in proposals), reports, institutional proposals, and strategy papers. Five of the seven interviewees report that they used citations regularly for a variety of reasons: to confirm sources for their audience; to locate sources at a future date; to justify assertions; to provide context; to demonstrate need in the region; to maintain the organization's reputation; to identify strategies, need, evidence-based practices, or replicable models; to examine opportunities and challenges at regional and national levels; to track whether interventions are successful; and to provide increased transparency.

Many staff respondents reported a preference for data from prominent national and international organizations and felt that these organizations had already vetted the data and could serve as clearinghouses of trustworthy findings. Some interviewees suggested that, relative to peer-reviewed papers, these prominent national and international organizations had name recognition that instilled a sense of confidence in the data.

While many staff members reported saving some of their sources, they saved them in folders on their personal drive. Others, who saved in a shared drive in team-wide resource folders, did not frequently utilize those repositories when they needed data. Wherever these sources were saved, they were not regularly updated and often became outdated. Descriptive file names were not common, which made searching and identifying content difficult. A third category of respondents did not save their sources and choose to repeat web searches when they needed to find a source of information again.

When staff members needed to share data or work across teams, they saw information management challenges manifest themselves. There was a lack of transparency between teams; many respondents reported that data regarding each program resided with the program officer; they did not know what information other teams possessed, and that lack of a formal system for sharing information was a problem. Most sharing of information and papers was through email, in conversations during meetings, or in the staff kitchen. Several respondents described the email channel as oversaturated and as a "black hole" for data. These conditions served to silo the different areas of work according to team/program and to limit the amount of potential cross-team and cross-program synergy.

Practices at Related Organizations

The information management specialist contacted three other large not-for-profit organizations that focus on women's health. Each organization had very different practices and capacities for citation management and access to the literature. Two of the three organizations had some form of library resources and used citation management software to draft reports and other publications. The need for citation management software and access to the literature was recognized and prioritized in organizations dedicated to producing academic papers and acting as a clearinghouse for the field. As an organization with a growing role in producing consumer health information, the organization that conducted this case study found validation for its interest in information management in the practices at other not-for-profits of comparable size and focus.

Discussion

Information Management Pilot Development

Following suggestions from the data management consultant, the organization implemented citation management software to provide a web-based shared library that enabled tagging, annotation, full document searches, collaborative PDF reading and mark-up, citation, and bibliography creation. The organization felt most comfortable with the customer support offered by a hosted service. These needs

IGAPS in Mendely	Expanded IGAPS (Multi-level)
• Information: CDC Reports • Geography: United States • Application: Fact Sheet • Population: Latina • Subject: IUDs	• Information – Government Information • CDC Reports • Geography – North America • United States • Application – Consumer Health Information • Fact Sheet

Figure 1. IGAPS Taxonomy Development

and preferences made Mendeley the best fit for the organization's information management pilot.

IGAPS Taxonomy Development and Relationship to Taxonomy Literature

The pilot development team held a series of meetings to determine the organization's priorities for organizing and using sources; it became apparent that important categories of information were related to the information's format, the geography of focus, the application for the information, the population to which the information relates, and the subject of the information. These categories were expressed using the mnemonic IGAPS (information, geography, application, population, and subject). Although there was interest in developing a hierarchical taxonomy of characteristics related to the IGAPS dimensions, the taxonomy was implemented within Mendeley, which would not support a taxonomy with a complex hierarchical structure at the time that the pilot was to be conducted (see figure 1).

Of the two restrictions proposed by Nickerson et al., the IGAPS taxonomy fits the collectively exhaustive restriction by requiring a characteristic for each one if its dimensions, but does not fit the mutually exclusive restriction in that it permits multiple characteristics in its subject dimension. Nickerson's restriction on mutual exclusivity would theoretically have helped efficient retrieval, but the organization's communications and assessment officers felt this would have made it difficult to classify all of a document's content with a single characteristic. The preference of the intended users for multiple labels for subject characteristic drove the departure from Nickerson's theoretical model. This departure from mutual exclusivity makes the IGAPS system a combination of faceted classification and descriptive metadata standard informed by user preference. While it is not a pure model, the pragmatic decision to defer to user preferences fits Cisco's best practice of keeping the context of the intended user in mind during taxonomy development.[26]

IGAPS is not unique as an approach to taxonomy that departs from Nickerson's mutual exclusivity criteria. In permitting the assignment of multiple characteristics to a single dimension, IGAPS was similar to Park's taxonomy used to classify weblogs.[27] MeSH used by NLM also permits multiple subheadings that are not mutually exclusive.[28] IGAPS followed Doria and IFLA in the effort to link the documents it classifies with user intent through its Action facet.[29]

With only five top-level categories, IGAPS conformed to the interpretation of Miller's rule suggested by Nickerson et al. and Neelameghan.[30] This number of categories, and the faceted approach to analyzing documents, shared some similarity with Ranganathan's Personality, Matter, Energy, Space, Time (PMEST) colon classification system used in Indian libraries.[31]

The pilot development team edited a pre-made keyword guide for "resources related to family planning and reproductive health" and categorized it to fit into the IGAPS categories in an iterative process.[32] This decision followed Cisco's best practice of incorporating existing taxonomic resources to save time; however, this taxonomy still required extensive editing to meet the needs of its intended users.[33] Following Cisco, this editing was executed in an iterative process that incorporated input from the organization's stakeholders.

Pilot Implementation

IPPF/WHR lacked the budget to support a dedicated information/data manager who would assume responsibility for maintaining the taxonomy. Taking plans for taxonomy maintenance into account, the taxonomy was implemented using a decentralized model.

Each of the fifteen staff members participating in the pilot assigned taxonomic terms to documents that they deposited in the Mendeley group. The taxonomic terms were logged on a worksheet that was uploaded as an attachment to the record. The information management specialist transferred terms from the worksheet to the tag field in the record. A version of this worksheet is available in the appendix to this paper.

The worksheet listed the IGAPS categories and provided instructions about the descriptions of the scope of each category and instructions for assigning a range of one to five terms for each category. The worksheet included space for additional terms and questions. The pilot manager followed up on these entries with each staff participant and this input was incorporated into the taxonomy development.

Pilot Assessment

The practice of using data mostly from organizational reports and fact sheets, rather than academic papers, presented a

challenge for easy integration of the pilot programs. Documents published by commonly used organizations often lacked accessible metadata that would allow automated creation of a complete Mendeley record. While the software allowed for making manual edits to records to complete metadata, this manual process was a barrier for staff use.

Staff contributions to the shared collection during the pilot period were not sufficient to create a robust organizational library. With only a few dozen items, it was not possible to test the utility of the IGAPS taxonomy for implementing more efficient searches, and the entire collection could be scanned at a glance. Staff members mentioned having limited time to spend with literature and expressed a desire for more time. The very limited growth of the library during the pilot period could be taken as an indicator that staff did not have much time to conduct literature searches and/or that staff literature searches were excessively time consuming. Qualitative interviews with the organization's leaders in communications and evaluation suggested that the pilot experience made valued contributions to their personal information management practices and their thinking about the organizational roles of information management and citation practices.

Overall, results of the pilot are mixed. Staff use of the IGAPS taxonomy during the pilot was inconsistent. During the pilot, the information management specialist was able to correct inconsistent uses of IGAPS; however, without dedicated staff to oversee the application of IGAPS to documents added to the Mendeley repository, it was clear that the information management system envisioned by the taxonomy development team was not sustainable. Despite this shortcoming, the pilot produced some valued outcomes: 1) it established staff use of Mendeley as a citation management system and organizational repository; 2) it delivered the IGAPS taxonomy as an organizational document; and 3) it provided staff with citation training.

Recommendations

User-generated libraries are challenging, even with a dedicated staff to curate them. Organizations without permanent librarians or information management specialists may face challenges when establishing a new information management system. This paper describes the beginning of a process that would require organizational culture shifts and investment of resources to effect sustainable change. Long-term success would depend on progress and development in three interrelated areas: staff commitment, culture shift, and training.

Staff Commitment

Giving staff an incentive to participate is critical to the long-term success of information management projects. Staff members must be informed and reminded that it is important to make the organizational library a part of their work plan, that this is a significant way to grow as an organization, and especially that it will benefit them as employees.

Culture Shift

The origin of changes to organizational information management practices lies in the desire for a shift in organizational culture. Getting staff to place greater value on tracking and citing sources will lead to an increase in the integrity of information—used both internally and externally—and reduce frustration and time spent backtracking statistics. Citation training was included in the pilot described in this report as a first step towards this culture shift.

Training

Continuing software and methods training is necessary, even for those who gained experience during pilot programs. These training sessions might be refreshers, updates on new features, additional ways to use software, opportunities to ask questions or raise technical issues, etc. It is important to share ideas about how to incorporate new software into daily routines, how each individual's engagement impacts the utility of the library, etc. A taxonomy is not useful if staff do not apply it in their information storage and retrieval practices. Staff need repeated training sessions to inculcate the best practices for applying the taxonomy in tagging to searching. Participants should also be updated when the taxonomy changes. As the project progresses, training on points of access and search methods and best citation practices will be necessary. These long-term considerations form a key framework that should continue to be discussed as steps are taken to address reference management and access to literature within a not-for-profit organization.

References

1. Robert C. Nickerson, Upkar Varshney, and Jan Muntermann, "A Method for Taxonomy Development and Its Application in Information Systems," *European Journal of Information Systems* 22, no. 3 (2013): 336–59.
2. Ibid.
3. A. Neelameghan, "Seminal Mnemonics in Knowledge Organization: Ancient Traditions and Modern Practices," *Information Studies* 12, no. 1 (2006): 5–26.
4. Daniel CannCasciato, "Ethical Considerations in Classification Practice: A Case Study Using Creationism and Intelligent Design.," *Cataloging & Classification Quarterly* 49, no. 5 (June 2011): 408–27.
5. Earl H. McKinney Jr., J. Charles, and I. I. Yoos, "Information

About Information: A Taxonomy of Views," *MIS Quarterly* 34, no. 2 (2010): 329–344.
6. David L. Hull, "Bias and Commitment in Science: Phenetics and Cladistics," *Annals of Science* 42, no. 3 (985): 319–38, https://doi.org/10.1080/00033798500200231.
7. Lutz Bornmann, "Is There Currently a Scientific Revolution in Scientometrics?," *Journal of the Association for Information Science & Technology* 65, no. 3 (2014): 647–48, https://doi.org/10.1002/asi.23073.
8. Sung-Yeon Park et al., "Inside the Blogosphere: A Taxonomy and Framing Analysis of Abortion Weblogs," *Social Science Journal* 50, no. 4 (2013): 616–24, https://doi.org/10.1016/j.soscij.2013.04.014.
9. Natalie Jomini Stroud, Ashley Muddiman, and Jae Kook Lee, "Seeing Media as Group Members: An Evaluation of Partisan Bias Perceptions," *Journal of Communication* 64, no. 5 (2014): 874–94, https://doi.org/10.1111/jcom.12110.
10. Sergio Tezanos Vázquez and Andy Sumner, "Revisiting the Meaning of Development: A Multidimensional Taxonomy of Developing Countries," *Journal of Development Studies* 49, no. 12 (2013): 1728–45, https://doi.org/10.1080/00220388.2013.822071.
11. Celia Jenkins, Charles Oppenheim, and Steve Probets, "RoMEO Studies 7: Creation of a Controlled Vocabulary to Analyse Copyright Transfer Agreements," *Journal of Information Science* 34, no. 3 (2008): 290–307.
12. Orélie Desfriches Doria, "The Role of Activities Awareness in Faceted Classification Development," *Knowledge Organization* 39, no. 4 (2012): 283–91.
13. Nickerson, Varshney, and Muntermann, "A Method for Taxonomy Development."
14. IFLA Study Group on the Functional Requirements for Bibliographic Records, "Functional Requirements for Bibliographic Records," Final Report, IFLA Series on Bibliographic Control (Munich, 1998), accessed August 8, 2015, http://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf.
15. Susan Cisco and Wanda Jackson, "Creating Order out of Chaos with Taxonomies," *Information Management Journal* 39, no. 3 (2005): 44–50.
16. Ashleigh Faith, "Linguistic Analysis of Taxonomy Facet Creation and Validation," *Key Words* 21, no. 1 (2013): 11–15.
17. Fran Alexander, "Devising a Framework for Assessing the Subjectivity and Objectivity of Information Taxonomy Projects," *Journal of Documentation* 70, no. 1 (2014): 4–24, https://doi.org/10.1108/JD-09-2012-0117.
18. Eric Charton et al., "Using Collaborative Tagging for Text Classification: From Text Classification to Opinion Mining," *Informatics* 1, no. 1 (2013): 32–51, https://doi.org/10.3390/informatics1010032.
19. Alex Haig et al., "METRO—the Creation of a Taxonomy for Medical Education," *Health Information & Libraries Journal* 21, no. 4 (2004): 211–19.
20. Zhonghong Wang, Christopher S.G. Khoo, and Abdus Sattar Chaudhry, "Evaluation of the Navigation Effectiveness of an Organizational Taxonomy Built on a General Classification Scheme and Domain Thesauri," *Journal of the Association for Information Science & Technology* 65, no. 5 (2014): 948–63, https://doi.org/10.1002/asi.23017.
21. Jean-Pierre Ekionea and Deborah Swain, "Developing and Aligning a Knowledge Management Strategy: Towards a Taxonomy and a Framework," *International Journal of Knowledge Management* 4, no. 1 (2008): 29–45.
22. Maheen Bakhtyar et al., "Creating Multi-Level Class Hierarchy for Question Classification with NP Analysis and WordNet," *Journal of Digital Information Management* 10, no. 6 (2012): 379–88.
23. Cisco and Jackson, "Creating Order out of Chaos with Taxonomies."
24. Haig et al., "METRO—the Creation of a Taxonomy for Medical Education."
25. Park et al., "Inside the Blogosphere."
26. Cisco and Jackson, "Creating Order out of Chaos with Taxonomies".
27. Park et al., "Inside the Blogosphere.".
28. "Introduction: What Is MeSH?," Training Material and Manuals, *US National Library of Medicine*, accessed January 21, 2014, http://www.nlm.nih.gov/bsd/disted/meshtutorial/introduction/index.html.
29. Doria, "The Role of Activities Awareness in Faceted Classification Development."
30. Nickerson, Varshney, and Muntermann, "A Method for Taxonomy Development"; Neelameghan, "Seminal Mnemonics in Knowledge Organization: Ancient Traditions and Modern Practices."
31. S. R. Ranganathan, *Colon Classification: Basic Classification* (Ess Ess Publications, 2007).
32. Knowledge for Health Project, "A User's Guide to POPLINE Keywords" (Johns Hopkins Center for Communication Programs, 2014), accessed August 11, 2015, http://www.popline.org/sites/default/files/KWGuide_10thEd.pdf.
33. Cisco and Jackson, "Creating Order out of Chaos with Taxonomies."

Appendix. IGAPS Worksheet

Refer to the taxonomy while you're reviewing the document—flipping back and forth between the article and the taxonomy—to remind yourself of IGAPS categories while reading.

Select tags for each category of IGAPS while reading. Adhere to the following rules:

1. None of the IGAPS categories should have more than five tags.
2. Be concise by using few tags as possible. Focus on the overarching themes of the article.
3. Each category of IGAPS should have at least one tag.
4. Use the terms closest to the language used in the article.

IGAPS Term selections:

Information:

(What type of information is this document?)

Geography:

(What country and/or region does this document focus on?)

Application:

(How will this information be applied?)

Population:

(What groups does this document focus on?)

Subject:

(What are the main themes that appear in this document?)

Additional Comments:

Notes on Operations

Using Automation and Batch Processing to Remediate Duplicate Series Data in a Shared Bibliographic Catalog

Elaine Dong, Margaret Anne Glerum, and Ethan Fenichel

The application of divergent local practices in a shared bibliographic database can result in unexpected display issues that adversely affect user experience. This is especially problematic when merging databases from multiple institutions accustomed to adopting local practices for their own constituents. The authors describe their experience with the application of automation tools, such as MarcEdit, Excel, and Python, during a large-scale remediation project. They used these tools to analyze, compare, and batch process bibliographic records to remediate obsolete and redundant series data in their shared bibliographic database.

Along with accuracy and comprehensiveness, consistency in cataloging practice improves discovery and identification of resources. Conversely, varying cataloging practice, whether due to local needs or changes to national standards, can result in inconsistent data within a shared bibliographic catalog. The consolidation of bibliographic databases in library consortia may exacerbate these inconsistencies. To maintain metadata quality and update older data to newer standards, catalogers can build on their traditional knowledge and also use data analysis, scripting, and batch manipulation when performing large-scale remediation.

The authors are catalogers at institutions comprising the State University Libraries (SUL) of Florida. As members of the Bibliographic Control and Discovery Subcommittee of the Council of State University Libraries, they formed the Multiple-Series Cleanup Task Force. The Task Force members were chosen due to their complementary skill sets. Two of the members have extensive experience and training in cataloging practice while the third had substantial experience with databases and systems technology before a career in librarianship. One of the members had experience developing Python scripts as a content systems analyst at a financial information provider. Another member has experience with developing XSLT and JavaScript programs. Although these tools were not used for this remediation project, experience with programing language provided a conceptual understanding that assisted with interpreting the Python scripts. All the members had varying experience with data analysis, batch processing, and batch loading as part of their assignments. To aid in these efforts, they independently learned to use MarcEdit through trial and error, webinars, and from peers. Similarly, they also learned how to take advantage of Excel's powerful data analysis tools.

The Task Force was charged with creating a plan to remediate duplicate series data that were causing issues in the catalog's discovery tool. To fulfill its charge, the Task Force identified records in SUL's shared bibliographic database that included obsolete and duplicate series fields that caused display problems.

Elaine Dong (edong@fiu.edu) is the Database & Metadata Management Librarian at Florida International University. **Margaret Anne Glerum** (aglerum@fsu.edu) is the Associate University Librarian and Head of Complex Cataloging at Florida State University Libraries in Tallahassee. **Ethan Fenichel** (fenichele@fau.edu) is the E-Resources Access Management Librarian at Florida Atlantic University.

Manuscript submitted July 11, 2016; returned to authors for revision October 7, 2016; revised manuscript submitted December 5, 2016; accepted for publication March 29, 2017.

The Task Force first analyzed the records using MarcEdit and Excel, and then developed a Python script to compare a subset of the records in the shared bibliographic database of the SULs—known as the Shared Bib—with their corresponding OCLC master records. They ultimately updated the problematic Shared Bib records using a locally developed batch-loading tool. The application of these automation tools saved a significant amount of time rather than manually updating each record. The workflows and processes used for this project serve as an example for how catalogers can approach future remediation projects in an efficient and effective manner.

Literature Review

How best to incorporate quality bibliographic description into a library's catalog has been a topic of discussion in literature for decades.[1] In 2008, *Cataloging and Classification Quarterly* devoted an entire special issue to the topic.[2] High-quality bibliographic description is generally defined as accurate, usable, complete, and consistent.[3] These components are needed for a positive impact on the user experience. Petrucciani writes about the need for consistency and accuracy as prerequisites for establishing trust among the users that the catalog will provide "clear and effective navigation functions among controlled bibliographic entities."[4] Dunsire states, "The efficiency and effectiveness of any information retrieval service requires coherency and consistency in metadata."[5] Harmon acknowledges the direct relationship between the presence of information in the bibliographic record and the library users' retrieval of that record in the discovery interface, and asserts that it is the cataloger's responsibility to support the organization's public service mission in providing access to research materials.[6] Among the key findings of the 2009 OCLC Report, *Online Catalogs: What Users and Librarians Want*, is that "appropriate, accurate and reliable data elements . . . are critical" in retrieving bibliographic descriptions and that "search results must be relevant and the relevance must be obvious."[7] It is that last statement that directly relates to the issues outlined in this paper—multiple series statements and access points are coded to display only in particular local discovery tools, leaving users wondering why the record was retrieved when it does not display the search terms entered.

To maintain the desired quality in their bibliographic database, libraries can outsource their database maintenance, provide it in-house, or use a combination of both. Guajardo and Carlstone describe a Resource Description and Access (RDA) conversion project at the University of Houston Libraries using Marcive, a bibliographic services company, plus in-house staff, to update their catalog records to the new standard.[8] Williams describes an authority remediation project provided by Marcive, followed by subsequent review by the London School of Economics Library staff.[9] Similarly, Finn described an authority control workflow at Virginia Tech that began with an updated authority file provided by Library Technologies Incorporated (LTI), followed by staff using MarcEdit, a free database maintenance program developed by Terry Reese, to edit the authority fields of vendor records before batch loading them.[10] Park and Panchyshyn discuss how they contracted with Backstage Library Works to enrich their MARC records with RDA elements while staff used MarcEdit in-house to create AACR2-RDA hybrid records during Kent State University Libraries' database enrichment project.[11]

Outsourcing database remediation was not an option for the Multiple Series Cleanup Project, so it was performed solely by the Task Force, drawing on earlier projects. Draper and Lederer at Colorado State University Libraries discuss a project using MarcEdit to generate particular field and subfield counts in a set of MARC records in preparation for batch loading. At the University of Minnesota Libraries, Traill and Genereux explained how they transformed Microsoft Office Excel spreadsheets into MARC records using MarcEdit.[12] Sanchez et al. at the Alkek Library at Texas State University-San Marcos described methods using MarcEdit and both Excel and Microsoft Office Word to provide quality control for vendor-supplied records.[13] Myntti and Neatrour demonstrated how MarcEdit and OpenRefine, a free, open-source program, were used to scrub and transform data to update the controlled vocabulary of existing data and to further enrich the metadata with Uniform Resource Identifier (URI) values in preparation for linked data capabilities at the University of Utah.[14]

When there is sufficient in-house expertise, computer programs can be developed for bibliographic database analysis and processing. Myntti and Cothran developed a process to achieve automated authority control for metadata in the University of Utah's digital collection. This process adapted existing services provided by Backstage Library Works to utilize algorithms for reconciling uncontrolled names and subject terms in XML data and replace them with authorized constructions.[15] Frank outlined a method of batch-processing MARC records using MarcEdit and Python, an open source programming language, plus PyMARC, a Python library for parsing MARC record data.[16] To automate the importing of metadata and content during a data migration into the DSpace archive directory format, Walsh at Ohio State University Libraries used Excel, Python, and Perl, another open-source programming language.[17] Mitchell and McCallum explored computational techniques for migrating metadata using OpenRefine and Python.[18] Mitchell later studied data analysis techniques for comparing different library holdings using Python, PyMARC, MySQL, and command line scripts.[19] For the Dewey Decimal Classification

Number "clean-up" at the Library of the Pontifical University Santa Croce, Bargioni et al. shared that seven different Perl programs were developed for queries via the API for their open source ILS, Koha.[20]

SUL Shared Bibliographic Database Overview

SUL members use Ex Libris' Aleph as their integrated library system. In June 2012, the eleven SUL members, in collaboration with the Florida Virtual Campus (FLVC), merged their twenty-three million bibliographic records from separate databases into the Shared Bib of about eleven million records.

SUL members have used OCLC records and vendor records for more than forty years, during which cataloging rules and practices have changed. Part of the need for the Multiple Series Cleanup Project stemmed from the 2008 change when the MARC 440 field (Series Statement/Added Entry-Title) was made obsolete.[21] Another key development was in June 2006, when the Library of Congress (LC) stopped creating authorized series access points (formerly referred to as headings) in conjunction with the transcribed series statement, a practice known as tracing, on its newly created bibliographic records.[22] An untraced series is indicated in the MARC 490 field with a first indicator "0" (490 0_). However, Program for Cooperative Cataloging (PCC) participants and other libraries continued to trace series. In MARC, traced series are encoded as MARC 490 field with a first indicator "1" (490 1_), which indicates that there is a corresponding authorized series access point in a MARC 80X-83X 8xx field.[23]

Before the Shared Bib merge, some SUL members imported different versions of the same OCLC or vendor supplied record, which contained variants in common fields. SUL members also added fields for local data specific to the items at their institution. During the merge, multiple copies of a bibliographic record were combined into one. Due to the difficulty of identifying the particular local data, it was agreed to that all the varying forms of fields would be retained. The subfield $5 was established to label fields with potentially local data. As a result, repeated fields with variations were added to Shared Bib records, including series fields that repeated due to the slight variations of the transcription, incorrect subfield coding, or varying tracing practices. The authors requested a report from FLVC that identified 209,671 records with multiple series MARC fields (440s and 490s).

SUL members share a statewide union discovery layer named Mango, which was developed by FLVC's predecessor the Florida Center for Library Automation (FCLA). Several institutions use a local instance of Mango in addition to the union version for statewide access.[24] To control the display of institution-specific data, FLVC configured Mango to use the SUL members' OCLC MARC Organization Code in MARC subfield $5.

Subsequent to the merge process, the subfield $5 protected fields from being overwritten during the updating of a Shared Bib record with an OCLC master record. Since many fields marked with subfield $5 are not necessarily local data, FLVC later changed the function of subfield $5 to only control display and not to protect the field. The SUL members identified thirty-four fields to protect, irrespective of a subfield $5, since those fields would be likely to contain local data.[25]

Display Issues in the Discovery Layer

The multiple functionalities and extensive use of the subfield $5 resulted in several problems in the Mango discovery layer. The Task Force focused on these issues affecting series data:

1. If a MARC 440 or 490 field includes a subfield $5, that field's series data will display only in the local Mangos corresponding to the MARC organization codes. Figures 1 and 2 show that the University of West Florida (UWF) and the University of North Florida (UNF) Mangos display only the series statements that have MARC 490 fields with the MARC Organization Code for its library, FPeU and FJUNF respectively.
2. If a MARC 440 or 490 field includes a subfield $5, that field will not display in the Union Mango nor in any other local Mango that lacks a corresponding subfield $5 code. Figure 3 shows that the Union Mango does not display any series statements because every MARC 490 has a subfield $5, yet series access points found in the MARC 830 fields do display because they lack subfield $5.
3. Due to the legacy functionality of MARC, Mango treats the MARC 440 as a series access point. If both MARC 440 and 830 are present, both fields display in the local Mango, even if those fields have the same text string. Figure 4 shows that since the 490 fields do not include subfield $5 with the MARC Organization Code for its library (FTS), the USF catalog does not display any series statements. However, since the 440 fields do have subfield $5 FTS, the University of South Florida (USF) catalog displays series access points found in both MARC 440 and 830 fields.

The following screenshots are various displays of the same Shared Bib MARC record containing these series statement and access point fields.

=440 0_ $a Essays in history, economics, & social science, $v 8 $5 FTS
=440 0_ $a Burt Franklin research & source works series, $v 163 $5 FTS
=490 0_ $a Burt Franklin research & source works series #163 $5 FPeU
=490 1_ $a Essays in history, economics, & social science #8. $5 FJUNF $5 FPeU
=830 _0 $a Burt Franklin research & source works series $v no. 163
=830 _0 $a Selected essays in history, economics, & social science, $v 8.

Series note:	Burt Franklin research & source works series #163 Essays in history, economics, & social science #8.
Series:	Burt Franklin research & source works series no. 163 Selected essays in history, economics, & social science, 8.

Figure 1. UWF Mango Catalog

Series note:	Essays in history, economics, & social science #8.
Series:	Burt Franklin research & source works series no. 163 Selected essays in history, economics, & social science, 8.

Figure 2. UNF Catalog

Series:	Burt Franklin research & source works series no. 163 Selected essays in history, economics, & social science, 8.

Figure 3. Union Mango Catalog

Series:	Burt Franklin research & source works series no. 163 Burt Franklin research & source works series, 163 Essays in history, economics, & social science, 8 Selected essays in history, economics, & social science, 8.

Figure 4. USF Mango Catalog

Shared Bibliographic Record Issues

In establishing best practices, SUL members understood that a Shared Bib record should represent a single manifestation, therefore the series statements should not differ among SUL members. However, consortial guidelines allowed for different tracing practices. Please see example 1 below for a case where member institutions chose different tracing practices.

When updating a Shared Bib record, obsolete MARC 440 fields should be replaced with a MARC 490 and its corresponding MARC 830 authorized access point. As discussed, this is a challenge when the MARC 440 fields are indicated as being specific to one of the SUL members (see example 2 below). As discussed in the previous section, the ambiguity around which fields truly are specific to one of the SUL members largely stems from the Shared Bib merge. Example 3 illustrates how this can make cataloging practice more difficult.

Example 1: Multiple MARC 490 fields for different tracing practice on a Shared Bib Record

=001 020001295
=035 __$a(OCoLC)49356140
=440 _0$a**Explorations** in sociology$vv.62$5FTS
=**490 0**_$a**Explorations** in sociology $vv.62$5FBoU$5FU
=**490 1**_$a**Explorations** in sociology ;$v62$5FTaFA$5FMFIU$5FTaSU
=830 _0$a**Explorations** in sociology ;$vv. 62.

The corresponding OCLC record

=001 ocm49356140\
=003 OCoLC
=**490 1**_$a**Explorations** in sociology ;$vv. 62
=**830** _0$a**Explorations** in sociology ;$vv. 62.

Example 2: Obsolete MARC 440 fields on a Shared Bib Record

=001 020000022
=035 __$a(OCoLC)00000069
=**440** _0$aReprints of economic classics **$5FMFIU$5FU**
=**440** _4$aThe Adam Smith library**$5FMFIU**
=490 0_$aReprints of economic classics$5FJUNF$5FTaFA$5FPeU
=490 0_$aThe Adam Smith library $5FJUNF$5FTaFA$5FPeU$5FU

Example 3: Multiple MARC 490 and 830 fields with same tracing practice on a Shared Bib Record

=001 020000093
=035 __$a(OCoLC)00000311
=**490** 1_$aBollingen series, 35:10. The A. W. Mellon lectures in the fine arts**$5FTaSU**
=**490** 1_$aBollingen series, 35. The A. W. Mellon lectures in the fine arts, 10 **$5FSsNC$5FMFIU$5FJUNF$5FPeU$5FBoU$5FTaFA$5FTS$5FU**
=**490** 1_$aBollingen series, 35:10**$5FOFT**
=**490** 1_$aBollingen series, 35. The A. W. Mellon lectures in the fine arts,$v10**$5FFmFGC**
=**830** _0$aBollingen series,$v35.
=**830** _0$aA.W. Mellon lectures in the fine arts ;$v10.

=**830** _0$aA.W. Mellon lectures in the fine arts.
=**830** _4$aThe A. W. Mellon lectures in the fine arts ;$v1961
=**830** _4$aThe A. W. Mellon lectures in the fine arts,$v10

The corresponding OCLC record

=001 ocm00000311\
=003 OCoLC
=**490** 1_$aBollingen series, 35. The A.W. Mellon lectures in the fine arts, 10
=**830** _0$aBollingen series ;$v35.
=**830** _0$aA.W. Mellon lectures in the fine arts ;$v10.

Project Goal

The project's goal was to resolve the issues affecting the display of series data in both the local and the Union Mango while preserving any data specific to each institution. The Task Force devised an automated resolution due to the sheer number of records with problematic attributes. After examining some of these problematic Shared Bib records, the Task Force found that most of the records originated from OCLC. SUL members have relied on OCLC, the largest bibliographic database in the world, as a main source for importing and updating local bibliographic records, even predating the creation of the Shared Bib. Accordingly, the Task Force discovered that most of these problematic local bibliographic records were imported from OCLC a long time ago and have not been updated since.

The example records displayed in the preceding section illustrate problematic Shared Bib records that were no longer compliant with current standards. Most corresponding OCLC master records had since been updated and contained only accurate series pairs. In contrast, the local records contained various forms of series fields that had been contributed over time in each library's individual catalog. These various forms of series fields were then merged into a single Shared Bib record. In addition to correct series data, the OCLC records contained enhancements contributed by OCLC members plus the automatic maintenance performed by OCLC over the years, such as RDA updates and FAST headings. For an example of a full record in Shared Bib compared to its corresponding OCLC record, see appendix A.

The Task Force determined that the best way to update these problematic Shared Bib records would be to overlay them with their latest OCLC master records. This would correct the specific problems with the series data with the added benefits of updating other fields in the local records, including RDA enhancements and additional access points. The Task Force also needed to identify which records were acceptable for overlay and to protect local data. The following section describes the analytical method and the tools used to achieve this goal.

Analysis of Shared Bib and OCLC Records

To identify which Shared Bib records were candidates for overlay, the Task Force performed the following analysis:

1. Shared Bib Records: MARC 035 Field Analysis

The MARC 035 field contains the system control number for the Shared Bib records. The purpose of the MARC 035 field analysis was to identify the locally held records that originated from OCLC and represent the same manifestation compared to those that were provided by other vendors or derived from OCLC records for different manifestations. Examples of the last case were the Shared Bib records in formats different from the corresponding OCLC records.

The authors created a random sample of 1,000 Shared Bib records from the report of 209,671 problematic records.[26] After extracting the MARC records from Shared Bib, the Task Force used MarcEdit to extract just the MARC 035 fields and to copy and paste the results into Excel. The values were sorted and the data were separated into the following four groups:

1. Records with OCLC numbers only (674 records, 67 percent)
2. Records containing more than one MARC 035 field where one of the MARC 035 field values is an OCLC number and another is a vendor identifier (63 records, 6 percent). The majority of these records were identified as vendor records. A separate remediation project is currently underway to address this type of record.
3. ProQuest CIS microfiche records in the Shared Bib with both a MARC 035 field containing an OCLC number and a MARC 035 field containing a proprietary ProQuest number.[27] Some OCLC numbers end with an "x" on the end (36 records, 3 percent).These Shared Bib records are used for microforms and were created by ProQuest from print format OCLC records. These records should not be overlaid by their corresponding OCLC records.
4. Vendor records lacking an OCLC number in MARC 035 fields (285 records, 28 percent). These records could not be updated by the overlaying method since they did not have OCLC records.

After discussion, the Task Force agreed that records in Group 2-4 were not suitable for overlay.

2. Shared Bib Records: Format Analysis

The Task Force developed a Python script to return the necessary MARC data to examine the records in Group 1.[28] In particular, the Task Force focused on the record format as determined by the fixed fields, mainly the MARC 008 field. Using the script, they identified 7,535 records matching Group 1 parameters from 10,000 records that were drawn from the original problem set. The distribution of the formats is shown in table 1.

Table 1 shows that the majority of Group 1 records (89 percent) are print format. There are also small percentages of electronic (5 percent), microform (6 percent), and unknown format (0.3 percent). The Task Force spot checked records for each format and determined that each format needed to be treated differently.

A portion of Shared Bib records coded as microform contained MARC 035 or MARC 019 fields matching OCLC records encoded as print format. In light of this finding, the Task Force added a comparison of the formats of the OCLC records and Shared Bib records as part of the automated analysis. They also determined that records with mismatched formats were not suitable candidates for overlay.

The Task Force determined that records coded as electronic format were not candidates for overlay. The provider-neutral cataloging policy that the PCC implemented in 2009 led to provider-specific records for electronic resources being merged into single provider-neutral records in OCLC.[29] This policy raised concerns about the consistency and comprehensiveness of description in the OCLC master records relative to local records. Before automating the overlay of records for electronic resources, the Task Force wanted to apply additional rigor to the analysis. To complete an iteration of enhancement without resolving this problem, the Task Force simply decided to exclude this category of records.

After reviewing the Shared Bib records with programmatically undetermined formats, the Task Force discovered that they were mostly map or GIS format records. They agreed that these and the Shared Bib records coded as print format were candidates for overlay.

Table 1. Format of Group 1 Records

No. of records with a MARC 035 field beginning with (OCoLC) prefix only	7,535	Percentage
Format: print	6,697	89.0
Format: electronic	391	5.0
Format: microform	422	6.0
Format: unknown	25	0.3

3. OCLC Master Records: MARC 490 and 830 Field Analysis

At this stage of analysis, the Task Force wanted to ensure that any potentially local series data in the Shared Bib would not be lost during the overlay process. To accommodate local practices, they wanted to avoid reversing the tracing of the series in the Shared Bib if the series was not traced on its corresponding OCLC record.

Among the 7,535 OCLC master records corresponding to the Group 1 records that were still candidates for overlay, the Task Force identified eighty-three OCLC bibliographic records (1 percent) that lacked any MARC 490 or 830 fields. Since their Shared Bib records contained MARC 440, 490, or 830 fields, which might be local series, the Task Force agreed that these records were not candidates for overlay. Instead, they created a set of records to be reviewed for authority control by a separate team.

The Task Force also identified 1,222 OCLC bibliographic records (16 percent) that contained MARC 490 0_. They discovered that some of the corresponding series authority records included a MARC 645 subfield $a with a value of "n" (untraced), subfield $5 DLC, and were created before 1989, hence the series were correctly untraced in the OCLC bibliographic records according to LC and PCC standards in Section Z1 of the Descriptive Cataloging Manual.[30] However, some series statements should be changed to traced (MARC 490 1_ and 830 _0 combination) since their series access points were established and should be traced according to their MARC 645 subfield $a with a value of "t" (traced). After discussion, the Task Force decided that these 1,222 records (16 percent) should be parsed for authority review and were not pursued as candidates for overlay.

Unprotected Local Series Data and Access Points in the Shared Bib

Focusing on the preservation of unprotected local series data and access points, the authors collaborated with SUL representatives and colleagues to collect information about data created by each library. This information helped in developing the Python script and determining the best method to identify and protect local data from overlay. The Task Force identified the following three types of local data created by SUL members that was not in the thirty-four protected fields:

1. Access Points for Local Collections

Some SUL members create access points that are only related to materials in their libraries, such as names of specific collections. Since these access points are relevant to a single institution, they do not need to be established

in any authority files. The purpose of these locally created access points is for easy retrieval of associated bibliographic records when these phrases do not appear on the materials. After the Shared Bib merge, SUL members have adopted the use of MARC 79X and 89X fields for these locally created access points.[31] For example, Florida International University (FIU) created and added a MARC 899 field with the phrase "George Wise Collection" to all the bibliographic records for materials donated by George Wise.[32] Before the Shared Bib merge, SUL members used other MARC fields for these locally made access points, including MARC 710, 490, and 830. Below is an example of an access point for a local collection in a MARC 710 field. In Shared Bib, MARC 79X and 89X are among the thirty-four fields protected from OCLC overlay. However, MARC 710, 490, and 830 are not. To preserve data in these fields, the Task Force agreed that the 710 fields should be protected during the overlay process.

=001 020173100
=035 __$a(OCoLC)09370337
=490 0_$aBulletin / Department of Agriculture [new series] ;$vno. 188$5FTaSU
=490 1_$aBulletin / Department of Agriculture, State of Florida ;$v[new ser.], no. 188$5FU$5FTS
=**710 2_$aFloridiana Collection.$5FTS**
=830 _0$aBulletin (Florida. Department of Agriculture) ;$vno.188.
=830 _0$aBulletin (Florida. Department of Agriculture) ;$vnew ser., no. 188.

2. Local Tracing Practices for Series-like Phrases

Some SUL members preferred to trace series-like phrases so that they are indexed and searchable as series and title in Aleph, whereas their SARs in the national authority file instruct catalogers to use the series-like phrases as quoted notes only. For example, authority record number (ARN) 5234175 "Black circle book," a SAR established in the national authority file, instructs catalogers to use the title as a quoted note only. Table 2 below shows the difference between the local and national SAR:

Some SUL members have added the series-like phrases to MARC 490 and 899 fields in Shared Bib bibliographic records as shown in the following example:

=001 032057831
=035 __$a(OCoLC)00289583
=**490** 1_$aA black circle book$5FTaSU
=**899** _0$aBlack circle book$5FTaSU

The local practice of adding the series-like phrase in the indexed MARC 490 and 899 fields will not be found in the corresponding OCLC record. As shown below, the OCLC record uses the series-like phrase as a MARC 500 quoted note only.

Table 2. Series Authority Record for "Black circle book" in Local and National Authority File

SAR in Local Authority Database	SAR in LC/NACO Authority File
=040 \\$aFNP$cFNP	=010 no 00040240
=130 \0$**aBlack circle book**	=040 \\$aNcU$beng$cNcU
=643 \\$aNew York$bGrove Press	=130 \0$aBlack circle book
=644 \\$af$5FJUNF	=643 \\$aNew York$bGrove Press
=645 \\$at$5FJUNF	=667 \\$**aGive phrase as quoted note**.
=646 \\$as$5FJUNF	

=001 289583
=003 OCoLC
=**500** __$a"a black circle book."

To retain data from these local tracing practices, MARC 490, 830 fields in Shared Bib records would need to be compared to the corresponding fields in OCLC master bibliographic records prior to the remediation process to determine if those fields contain local data.

3. Locally Created Series Authority Records

Prior to the SUL members' participation in the Library of Congress (LC)/Name Authority Cooperative Program (NACO), locally created authority records, including those for series headings, existed in each SUL member's local databases. In Florida, the University of Florida (UF), Florida International University (FIU), Florida State University (FSU), and University of North Florida (UNF) libraries are the earliest NACO contributors. NACO participants contribute authority records for names, uniform titles, and series headings to the LC/NACO Name Authority File (NAF). In October 2008, seven libraries, including five university libraries (UF, UNF, FIU, FSU, and USF), two college libraries, and one public library in Florida joined the Florida NACO Funnel. A UF librarian served as the funnel coordinator. This joint endeavor consolidated members' efforts to make a larger contribution to the national authority file and has improved the quality of authority records originating in Florida.[33]

After the Shared Bib merge, all of the locally created authority records were migrated to a combined local authority file in Shared Bib. The Task Force examined a sample of locally created SARs and found that many of them were established in the LC/NACO NAF. The comparison between SARs created locally and those in the national file showed that most of them have the same form of authorized access point (MARC 130 field), while some provided

a different treatment (e.g. Analyzed versus Not analyzed, Traced versus Not traced, Classified as a collection versus Classified separately).[34] These locally created series were added to MARC 440/490/830 fields on Shared Bib records; here is an example:

Locally Created Series in MARC 440 and 490 field on a Shared Bib record

=001 020001980
=**440** _0$aAddison-Wesley series in metallurgy and materials$5FMFIU$5FTS$5FTaSU
=**490** 0_$aAddison-Wesley series in metallurgy and materials$5FJUNF$5FBoU$5FU

The locally created series with unestablished SARs in the national authority file would need to be identified and retained during the overlay process.

Project Workflow and Implementation

Based on the findings from record analysis and information collected about local series practice, the Task Force developed an initial remediation plan. After testing the first 10,000 problem records, analyzing the test results, and adjusting the program logic, the Task Force finalized the workflow (see figure 5). For an account of the project's timeline, please see appendix B.

The Task Force took the following steps to remediate the problematic records:

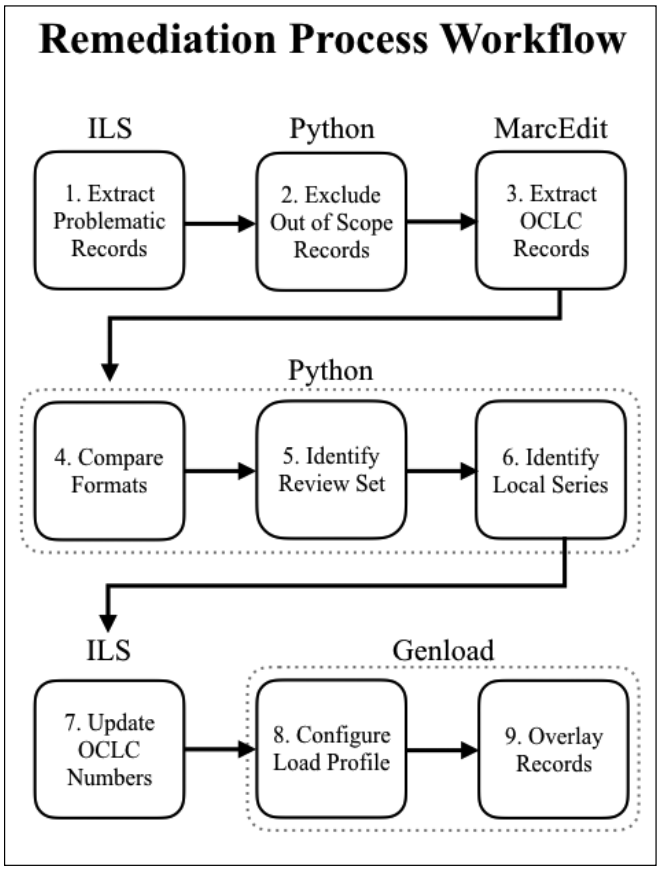

Figure 5. Overall Workflow

Step 1. Use Aleph Services to Extract Problematic Shared Bib Records

The Task Force first extracted the 222,404 problematic MARC records from the Shared Bib in twenty-three batches using a function for record retrieval native to the consortium's cataloging system, Aleph.

Step 2. Use Python Script to Remove Records beyond Scope of Analysis

In the section *Analysis of Shared Bib and OCLC Records*, it is established that when updating Shared Bib records, the Task Force wanted to overlay only non-electronic resource records that could be firmly established as OCLC records. To do this, they collaboratively created a Python script to identify records that originated from OCLC defined by having only "OCoLC" in the MARC 035 prefix. The script also classified the record formats to filter out electronic resource records. After completion of this step, there were 130,692 Shared Bib records remaining.

Step 3. Use MarcEdit to Extract OCLC Master Records

Using the MarcEdit Z39.50 Client's batch processing function, the Task Force retrieved the corresponding OCLC master records.

Step 4. Use Python Script to Compare Formats of Shared Bib and OCLC Records

Following the download, the Task Force developed a second Python script to compare the format of the Shared Bib records with their corresponding OCLC master records (see appendix C). Record pairs with mismatched formats were identified and excluded.

Step 5. Use Python Script to Identify OCLC Records for Authority Review

The Task Force used the same script for format comparisons to build the Authority Review Set. This set would contain the Shared Bib records whose corresponding OCLC records that either lacked a MARC 490 field or contained a MARC

490 0_. These records would not be considered for overlay but instead were referred to a separate team to analyze for compliance with consortial authority policies. After this process was applied to all twenty-three batches, the Task Force identified 25,951 OCLC records and corresponding Shared Bib records as the Authority Review Set.

Step 6. Use Python Script to Identify Local Series

After excluding records with mismatched formats and records from the Authority Review Set, the script performed a text string comparison between the series data on the Shared Bib records and those on the corresponding OCLC master records (see appendix D). The goal of the comparison was to identify local series on the Shared Bib records and to flag them for the "Do Not Overlay" set. The records with matching series data were placed in the "Suggest Overlay" set.

To eliminate non-critical mismatches between text strings, the Task Force added additional rules to normalize the data prior to the comparison process. The Task Force chose to remove "his," "her," or "him" from the beginning of the series text string because the words were used inconsistently, especially in Shared Bib records. The differences were not critical enough to classify as a mismatch. The Task Force chose to remove numbers from subfields $a and $p since the series numbering had been incorrectly entered in these subfields. Diacritics were normalized so that differences in character encodings did not result in a mismatch.[35] All of the text normalization rules applied in the script are listed below.

- Strip out the following data in subfield $a and subfield $p for MARC 440, 490, and 830 fields before comparison:
  - Initial articles in English, French, and Spanish: the, a, an, el, los, la, las, un, unos, una, unas, le, la, l', les, un, une, des
  - His, Her, Him
  - Punctuation marks including ‘ ’ “ ” ... ! : ; , . [ ] < > ( ) { } - | / \
  - Numbers
  - Volume and number abbreviations ("NO" "V" "VOL")
- Additional text manipulation
  - Convert all text to uppercase
  - Normalize text encoding of diacritical marks to use UTF-8

Comparison Logic

After the script normalized the series data in the form of text strings, it performed a series of comparisons. The order of the comparisons was significant and in each comparison, either the text strings were considered as equal or the Shared Bib record would not be considered a candidate for overlay and was flagged for the "Do Not Overlay" set. In each comparison, only the subfields $a and $p were used from the MARC fields 440, 490 and 830.

First, the script compared all of the MARC 440 fields from a Shared Bib record with the MARC 490 and 830 fields of its corresponding OCLC master record. If the script determined that the series data did not match, the Shared Bib record was placed in the "Do Not Overlay" set. If the Shared Bib MARC 440 matched the OCLC master record's series data, the script proceeded to the next step.

Second, the script compared all of the MARC 490 fields from the Shared Bib record with its corresponding OCLC master record's series data. If the script determined the series data did not match, the Shared Bib record was placed in the "Do Not Overlay" set. If the Shared Bib MARC 490 matched the OCLC master record's series data, the script proceeded to the next step.

In the third and final comparison, the script compared all of the MARC 830 fields from the Shared Bib record with its corresponding OCLC master record's MARC 830 data. If the script determined the series data did not match, the Shared Bib record was placed in the "Do Not Overlay" set. If the Shared Bib MARC 830 matched the OCLC master record series data, the script added the Shared Bib record to the "Suggest Overlay" set. It would then repeat the comparisons for the next Shared Bib record in the batch. In total, by using the script, the Task Force placed 53,802 records in the "Suggest Overlay" set. For diagrams of steps 5 and 6, see figure 6.

Step 7. Use Aleph Services to Update OCLC Numbers of Shared Bib Records

While performing the comparisons in step 3, the script identified 243 cases where the Shared Bib record's OCLC number in the MARC 035 field did not match any OCLC Master record due to a merge of OCLC records. To accurately update the Shared Bib records, the Task Force first updated the MARC 035 field value to match the current OCLC number. The Task Force completed this using an automated service native to Aleph. If the current OCLC record was also in the Shared Bib, the Task Force deleted the duplicate record.

Step 8. Use GenLoad Profile to Protect Local Fields from Overlay

GenLoad is a record loading utility created by FLVC for SUL members to load MARC data into the Shared Bib.[36] GenLoad performs each load based on the profile configuration.

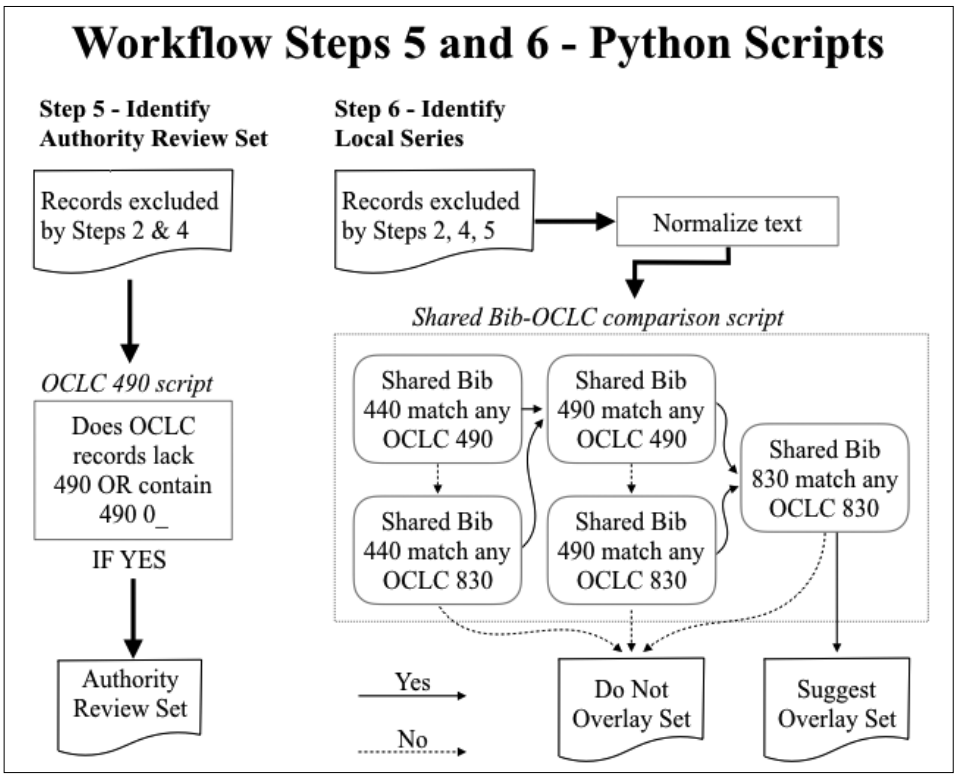

Figure 6. Expanded Workflow, Steps 5–6

The profile controls which data to insert and which data to protect or replace. The Task Force created a custom configuration to protect local data during the overlay process.

The Task Force added all of the fields to be protected to the GenLoad Profile. These include the thirty-four fields established by FLVC. Among these thirty-four fields, there are two non-standard MARC fields. The first is LKR, which is used to link bibliographic records in the Shared Bib. The second is TKR, which is a pre-merge holdover used to create an indexed string on the bibliographic records. Beyond the thirty-four fields, the Task Force included another non-standard MARC field to indicate record status, abbreviated as "STA." For example, "STA $aProvisional" on a Shared Bib record indicates it's a provisional record. The Task Force also included MARC 520, 599, and 710 fields, because they are likely to contain local data.

The profile protects local data in the following fields:

1. Established MARC fields to protect: LKR, TKR, 351, 500, 501, 506, 520, 533, 540, 541, 542, 545, 561, 562, 563, 583, 584, 590, 690, 691, 699, 790, 791, 797, 845, 896, 897, 898, 899, 909, 951, 970, 655 _7 with the following subfield $2: rbprov, rbbin, rbgenr, rbpap, rbpri, rbpub, rbtyp,
2. Added MARC fields: STA, 520, 599, 710, 655 _7 $2 local

Step 9. Used GenLoad to Batch Overlay Shared Bib Records with OCLC Records

Following a review period in which other SUL members provided feedback, the Task Force proceeded to the final step. They downloaded the OCLC master records that corresponded to the Shared Bib records in the "Suggest Overlay" set using MarcEdit. Following initial testing, the Task Force used GenLoad to overlay 51,818 Shared Bib records within two weeks.

Results

The Task Force's analysis and the resulting procedure that they developed culminated in the identification of 53,802 records as candidates for overlay, including approximately 2,000 duplicates from the originally identified 222,404 records with multiple series issues. Following the Task Force's work, a total of 51,818 Shared Bib records were overlaid. See appendix E for an example of a Shared Bib record before and after the overlay process.

This duplicate series data remediation project has made significant improvements in the quality of the Shared Bib database. The updates to series fields improved presentation, retrieval, and access for users of the consortial discovery systems. The project has also impacted the Shared Bib environment for internal maintenance. Concurrently, SUL members were preparing to merge their database with the State College libraries, as part of a migration to a new Next-Generation Integrated Library System. The improvements have reduced the overall amount of work required to complete that migration.

Future Projects Possibilities

The steps taken to remediate series data in our shared bibliographic database utilizing OCLC master records demonstrates a process that is repeatable and expandable. In our project, the use of PyMarc allowed us to create a customized process for analyzing and manipulating a large amount of MARC data. The writing of scripts by members of a cataloging team opens the possibilities for new procedures. Cataloging units could replicate the process to remediate

MARC fields containing local data by distinguishing locally created data in local records from non-local data in OCLC master records. As part of the remediation, the units could move local data to appropriate locally defined fields, such as MARC 89x fields.

A future project expanding on this work would allow bibliographic records to receive automated quality checks. Scripts could identify problems in local records, errors in OCLC records prior to loading into a local database, or perform comparisons between local and OCLC records. Resolutions for identified problems would follow either through further scripts or human intervention. By building the scripting into workflows such as database maintenance, copy cataloging, and batch loading, bibliographic records are reviewed automatically for predictable problems.

In the future, a shift to a linked data bibliographic environment will reduce the need for this process. The procedure relies on the model of a bibliographic record in the database as a document. Shifting bibliographic description to records as data graphs, or serialized data, will remove the need to analyze the full bibliographic document since data will be updated at a more granular level.[37] This question remains to be addressed as the structures and models of linked library data are developed.[38] The expansion of the scripting abilities in cataloging units is likely to be an essential component in the transition to the new models and workflows.

Conclusion

When large-scale changes to library bibliographic data are required, cataloging departments may lack the resources to suspend other projects and will spend hours manually updating records. By exploiting new technologies and skills, they can quickly adapt their data to the latest systems, cataloging standards, and changes in practice. The ability to utilize automated tools to analyze and batch process data is now an essential skill for librarians responsible for bibliographic data.

SUL faced large-scale changes that began with a system migration and were exacerbated by revisions to the practices of recording series data. When it became apparent that the existing practices were adversely affecting users, the Task Force identified how to bulk update series data. By using generally available tools—Microsoft Excel, MarcEdit, Python, and a locally developed data loader, GenLoad—the Task Force eased the analysis and largely automated the record update process. Three tech savvy catalogers completed this work without the involvement of formal software developers or systems experts. The Task Force made significant updates to the Shared Bib environment for all SUL members, with minimal help. In doing so, they demonstrated the value of leveraging automation in consortial collaboration.

While updating the Shared Bib records with OCLC master records, the Task Force made improvements beyond the series data that were initially the target for enhancement. In many cases, bibliographic records in the Shared Bib had not been updated in a long time. The latest versions of the OCLC master records contained improved description that would not have been captured through normal workflow processes. For example, the updated OCLC records contained the results of OCLC automated enhancements and authority control such as RDA updates and FAST subject headings.

The Task Force's analysis helped highlight the benefits of establishing best practices between SUL members. Accordingly, the Task Force made recommendations for SUL members on how to transcribe series in general and to add local series. One issue that the Task Force encountered was different tracing practices among individual libraries in a shared database. The Shared Bib Guidelines that all SUL members follow state that individual libraries may apply varying practices for analysis, tracing, and classification practice found in the LC Authority File. The Task Force recommended that the best practice is to use the OCLC master record's treatment of the series fields rather than alter the Shared Bib record. If the OCLC bibliographic or authority record needs to be revised to the current standard, that should also be done. If the library initiating the change is not authorized to edit the OCLC record, they can contact an SUL member who is authorized to do so.

Another issue that the Task Force observed is that individual libraries have used different fields for local series before Shared Bib. After the records have been merged into a single database, it takes a significant effort to identify and protect local data from being overlaid and causes serious challenges for data remediation. The authors feel that in a shared database, it is better to put local series and other local data into actual locally defined fields such as the MARC 590, 69X (local subject access fields), 79X (local added entry fields), 89X (local series added entries), and 9XX (local data elements) and minimize the use of other fields for local data. If a future library system allows it, it would be ideal to record local data in a separate section (e.g., holding records), not in bibliographic records, which would make the management and maintenance of the shared database much easier and efficient.

References and Notes

1. Janet Swan Hill, "Is It Worth It? Management Issues Related to Database Quality," *Cataloging & Classification Quarterly* 46, no. 1 (2008): 5–26, https://doi.org/10.1080/01639370802182885; Barbara Schultz-Jones et al., "Historical and Current Implications of Cataloguing Quality for Next-Generation Catalogues," *Library Trends* 61, no. 1 (2012): 49–82.

2. "Special Issue: Bibliographic Database Quality," *Cataloging & Classification Quarterly* 46, no. 1 (2008).
3. Hill, "Is It Worth It?"; Peter S. Graham, "Quality in Cataloging: Making Distinctions," *Journal of Academic Librarianship* 16, no. 4 (1990): 213–18; Heather Moulaison Sandy and Felicity Dykas, "High-Quality Metadata and Repository Staffing: Perceptions of United States-Based OpenDOAR Participants," *Cataloging & Classification Quarterly* 54, no. 2 (2016): 101–16, https://doi.org/10.1080/01639374.201.1116480; Philip Hider and Kah-Ching Tan, "Constructing Record Quality Measures Based on Catalog Use," *Cataloging & Classification Quarterly* 46, no. 4 (2008): 338–61.
4. Alberto Petrucciani, "Quality of Library Catalogs and Value of (Good) Catalogs," *Cataloging & Classification Quarterly* 53, no. 3–4 (2015): 303–13.
5. Gordon Dunsire, "Collecting Metadata from Institutional Repositories," *OCLC Systems & Services* 24, no. 1 (2008): 51–58.
6. Joseph C. Harmon, "The Death of Quality Cataloging: Does It Make a Difference for Library Users?," *Journal of Academic Librarianship* 22, no. 4 (1996): 306–7.
7. Karen S. Calhoun et al., "Online Catalogs: What Users and Librarians Want: An OCLC Report" (Dublin: OCLC, 2009), accessed December 4, 2016, http://www.oclc.org/content/dam/oclc/reports/onlinecatalogs/fullreport.pdf.
8. "Marcive," accessed December 4, 2016, http://home.marcive.com; Richard Guajardo and Jamie Carlstone, "Converting Your E-Resource Records to RDA," *Serials Librarian* 68, no. 1–4 (2015): 197–204.
9. Helen K. R. Williams, "Cleaning up the Catalogue," *Library & Information Update* (2010): 46–48.
10. Library Technologies, Inc. "We are the Authority Control Specialists," accessed December 4, 2016, https://www.authoritycontrol.com; Mary Finn, "Batch-Load Authority Control Cleanup Using MarcEdit and LTI," *Technical Services Quarterly* 26, no. 1 (2009): 44–50; "About MarcEdit," accessed December 4, 2016, http://marcedit.reeset.net/about-marcedit.
11. Amey L. Park and Roman S. Panchyshyn, "The Path to an RDA Hybridized Catalog: Lessons from the Kent State University Libraries' RDA Enrichment Project," *Cataloging & Classification Quarterly* 54, no. 1 (2016): 39–59, http://www.tandfonline.com/doi/abs/10.1080/01639374.2015.1105897.
12. Daniel Draper and Naomi Lederer, "Analysis of Readex's Serial Set MARC Records: Improving the Data for the Library Catalog," *Government Information Quarterly* 30, no. 1 (2013): 87–98, https://doi.org/10.1016/j.giq.2012.06.010; Stacie A. Traill and Cecilia Genereux, "Strategies for Catalog Management of Electronic Monographs in Series," *Serials Librarian* 65, no. 2 (2013): 167–80.
13. Elaine Sanchez et al., "Cleanup of NetLibrary Cataloging Records: A Methodical Front-End Process," *Technical Services Quarterly* 23, no. 4 (2006): 51–71, https://doi.org/10.1300/J124v23n04_04.
14. Jeremy Myntti and Anna Neatrour, "Use Existing Data First: Reconcile Metadata before Creating New Controlled Vocabularies," *Journal of Library Metadata* 15, no. 3–4 (2015): 191–207, https://doi.org/10.1080/19386389.2015.1099989; "OpenRefine," accessed December 4, 2016, http://openrefine.org.
15. Jeremy Myntti and Nate Cothran, "Authority Control in a Digital Repository: Preparing for Linked Data," *Journal of Library Metadata* 13, no. 2-3 (2013): 95–113; "Backstage Library Works," accessed December 4, 2016, http://www.bslw.com/about/.
16. Heidi Frank, "Augmenting the Cataloger's Bag of Tricks: Using MarcEdit, Python, and PyMARC for Batch-Processing MARC Records Generated From the Archivists' Toolkit," *Code4lib Journal* 20 (2013), accessed December 4, 2016, http://journal.code4lib.org/articles/8336; "About Python™: Python.org," accessed December 4, 2016, http://www.python.org/about/; "PyMARC," accessed December 4, 2016, http://pymarc.sourceforge.net.
17. Maureen P. Walsh, "Batch Loading Collections into DSpace: Using Perl Scripts for Automation and Quality Control," *Information Technology & Libraries* 29, no. 3 (2010): 117–27; "About Perl: www.perl.org," accessed December 4, 2016, http://www.perl.org/about.html; "DSpace: DSpace is a turnkey institutional repository application," accessed December 4, 2016, http://dspace.org.
18. Erik Mitchell and Carolyn McCallum, "Old Data, New Scheme: An Exploration of Metadata Migration using Expert-Guided Computational Techniques," *Proceedings of the American Society for Information Science and Technology* 49, no. 1 (2012) 1–10, https://doi.org/10.1002/meet.14504901091.
19. Erik T. Mitchell, "Reconciling Holdings Across Multiple Libraries: A Study in Data Analysis Techniques," *Technical Services Quarterly* 33, no. 2 (2016): 154-169, https://doi.org/10.1080/07317131.2016.1135000.
20. Stefano Bargioni et al., "Obtaining the Dewey Decimal Classification Number from Other Databases: a Catalog Clean-up Project," *Italian Journal of Library & Information Science* 4, no. 2 (2013): 176, https://doi.org/10.4403/jlis.it-8766.
21. "MARC Proposal No. 2008-07," accessed December 4, 2016, http://www.loc.gov/marc/marbi/2008/2008-07.html.
22. "Series at the Library of Congress: June 1, 2006" (Washington, DC: Library of Congress, 2006), accessed December 4, 2016, http://www.loc.gov/catdir/cpso/series.html.
23. "MARC 21 Bibliographic: 80X-83X - Series Added Entry Fields" (Washington, DC: Library of Congress, 2008), accessed December 4, 2016, http://www.loc.gov/marc/bibliographic/bd80x83x.html.
24. "Florida Academic Library Services Cooperative (FALSC): Discovery Tools" (Tallahassee, Florida: FALSC, 2015), accessed December 4, 2016, https://libraries.flvc.org/discovery-tools.

25. Bibliographic Control and Discovery Subcommittee of the Council of State University Libraries, "Shared Bib Guidelines. Appendix IIIA: Fields to Protect on Overlay from OCLC Gateway Import," (Gainesville, FL), accessed December 4, 2016, https://sharedbib.pubwiki.fcla.edu/wiki/index.php/Shared_Bib_Guidelines_Online#APPENDIX_IIIA:_Fields_to_Protect_on_Overlay_from_OCLC_Gateway_Import.
26. A VBA macro was created using MOD(ROW(),209)=1 to select every 209th row.
27. "Legislative Publications: CIS Congressional Bills, Resolutions & Laws on Microfiche (1933–2008)" (Ann Arbor, Michigan: ProQuest, 2014), accessed December 4, 2016, http://cisupa.proquest.com/ws_display.asp?filter=cis_leaf&item_id={1D481C6F-CA7A-4929-B4B0-BC90D20FAC71}.
28. Ethan Fenichel, "FLVC_490_Duplicates," GitHub, accessed December 4, 2016, https://goo.gl/ttjBtp.
29. "Program for Cooperative Cataloging (PCC) Provider-Neutral E-Resource MARC Record Guidelines," accessed December 4, 2016, http://www.loc.gov/aba/pcc/scs/documents/PCC-PN-guidelines.html.
30. "Descriptive Cataloging Manual Section Z1 and LC Guidelines Supplement to MARC 21 Format for Authority Data" (Washington, DC: Library of Congress, 2016), accessed December 4, 2016, http://www.loc.gov/catdir/cpso/z1andlcguidelines.html.
31. Bibliographic Control and Discovery Subcommittee of the Council of State University Libraries, "Shared Bib Guidelines. Section 3.4.8: Local Series" (Gainesville, FL), accessed December 4, 2016, https://sharedbib.pubwiki.fcla.edu/wiki/index.php/Shared_Bib_Guidelines_Online#3.4.8_Local_Series.
32. "899 Local Series Added Entry-Uniform Title" (Dublin, Ohio: OCLC, 2016), accessed December 4, 2016, https://www.oclc.org/bibformats/en/8xx/899.html.
33. Priscilla William, "TSPC Authorities Subcommittee Report on the Florida NACO Funnel" (Gainesville, FL, 2010), accessed December 4, 2016, http://csul.net/sites/csul.fcla.edu/uploads/authorities-NACOrpt-11-03-10.pdf.
34. "MARC 21 Format for Authority Data—64X Series Treatment General Information" (Washington, D.C.: Library of Congress, 2008), accessed December 4, 2016, www.loc.gov/marc/authority/ad64x.html.
35. Geoffrey Spear, "More Unicode Issues - Diacritics This Time" [Online forum comment]. Aug.10, 2015, Message posted to https://groups.google.com/forum/#!msg/pymarc/w9iy9dTb5xQ/RcbEg4VaHQAJ.
36. "GenLoad" (Gainesville, FL: Florida Virtual Campus, 2012), accessed December 4, 2016, https://support.flvc.org/knowledge-base/kbdw/KBA-01484-R4V7.
37. "RDF 1.1 Concepts and Abstract Syntax," accessed December 4, 2016, http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/.
38. "BIBFRAME Training at the Library of Congress" (Washington, DC: Library of Congress, 2016), accessed December 4, 2016, http://www.loc.gov/catworkshop/bibframe/.

Appendix A. A Full Record in Shared Bib and its Corresponding OCLC Record

Shared Bib Record:

=LDR 04517cam a22008534a 4500
=001 020001295
=005 20111213154942.0
=008 020313s2002\\\\enka\\\\\b\\\\001\0\eng\\
=010 \\$a2002024878
=015 \\$aGBA2-54901
=019 \\$a50433750$a51052019$a51681752
=020 \\$a0333984994 (alk. paper)
=020 \\$a0333984994
=035 \\$a(OCoLC)49356140
=040 \\$aDLC$beng$cDLC$dUKM$dTJC$dMUQ$dNLGGC$dBAKER$dBTCTA
$dYDXCP$dOCLCG$dIG#$dKAAUA$dGEBAY$dOCLCQ$dFUG
=650 \7$aTi=042 \\$apcc
=050 00$aHM656$b.S63 2002
=082 00$a304.2/3$221
=084 \\$a71.02$2bcl
=245 00$aSocial conceptions of time :$bstructure and process in work and everyday life /$cedited by Graham Crow and Sue Heath.
=260 \\$aHoundmills, Basingstoke, Hampshire ;$aNew York :$bPalgrave MacMillan,$c2002.

=300 \\$axvii, 266 p. :$bill. ;$c23 cm.
=440 \0$aExplorations in sociology$vv.62$5FTS
=490 0\$aExplorations in sociology$vv.62$5FBoU$5FU
=490 1\$aExplorations in sociology ;$v62$5FTaFA$5FMFIU$5FTaSU
=504 \\$aIncludes bibliographical references (p. 247-263) and index.
=650 \0$aTime$xPsychological aspects.
=650 \0$aTime$xSocial aspects.
=650 \6$aTemps$xAspect psychologique.
=650 \6$aTemps$xAspect social.
=650 07$aZeit.$2swd
=650 07$aAlltag.$2swd
=650 07$aZeitwahrnehmung.$2swd
=650 07$aAufsatzsammlung.$2swd
=650 17$aSociologische aspecten.$2gtt
=650 17$aPsychologische aspecten.$2gtt
=650 17$aTijd.$2gtt
=700 1\$aHeath, Sue.
=700 1\$aCrow, Graham.
=700 1\$aHeath, Sue,$d1964-
=830 \0$aExplorations in sociology ;$vv. 62.

Its Corresponding OCLC Record:

=LDR 03318cam a22007574a 4500
=001 ocm49356140\
=003 OCoLC
=005 20150914140806.0
=008 020313s2002\\\\enka\\\\\b\\\\001\0\eng\\
=010 \\$a 2002024878
=040 \\$aDLC$beng$cDLC$dUKM$dTJC$dMUQ$dNLGGC$dBAKER$dBTCTA
$dYDXCP$dOCLCG$dIG#$dKAAUA$dGEBAY$dOCLCQ$dOCLCF$dOCLCO$dOCLCQ
=015 \\$aGBA254901$2bnb
=019 \\$a50433750$a51052019$a51681752
=020 \\$a0333984994$q(alk. paper)
=020 \\$a9780333984994$q(alk. paper)
=035 \\$a(OCoLC)49356140$z(OCoLC)50433750$z(OCoLC)51052019$z(OCoLC)51681752
=042 \\$apcc
=050 00$aHM656$b.S63 2002
=082 00$a304.2/3$221
=084 \\$a71.02$2bcl
=245 00$aSocial conceptions of time :$bstructure and process in work and everyday life /$cedited by Graham Crow and Sue Heath.
=260 \\$aHoundmills, Basingstoke, Hampshire ;$aNew York :$bPalgrave MacMillan,$c2002.
=300 \\$axvii, 266 pages :$billustrations ;$c23 cm.
=336 \\$atext$btxt$2rdacontent
=337 \\$aunmediatedbn2rdamedia
=338 \\$avolumebnc2rdacarrier
=490 1\$aExplorations in sociology ;$vv. 62
=504 \\$aIncludes bibliographical references (pages 247-263) and index.
=650 \0$aTime$xSocial aspects.
=650 \0$aTime$xPsychological aspects.
=650 \6$aTemps$xAspect social.
=650 \6$aTemps$xAspect psychologique.

```
=650  \7$aTime$xPsychological aspects.$2fast$0(OCoLC)fst01151056
=650  \7$aTime$xSocial aspects.$2fast$0(OCoLC)fst01151066
=650  17$aTijd.$2gtt
=650  17$aPsychologische aspecten.$2gtt
=650  17$aSociologische aspecten.$2gtt
=650  07$aZeit.$2swd
=650  07$aAlltag.$2swd
=650  07$aZeitwahrnehmung.$2swd
=650  07$aAufsatzsammlung.$2swd
=700  1\$aCrow, Graham.
=700  1\$aHeath, Sue,$d1964-
=830  \0$aExplorations in sociology ;$vv. 62.
```

Appendix B. Project Implementation Timeline

Jan. 2015	A report of 209,671 Shared Bib records with multiple series (MARC 440/490/830) fields was generated by FLVC
Mid-April	Multiple-Series Cleanup Task Force was formed to analyze potential solutions for the issues resulting from the multiple series in these records
May-Aug.	Task Force analyzed sample records and began fact-finding
June	Task Force developed strategy: use Python program to flag records that contain local data, and use GenLoad to batch overlay records with obsolete and duplicate series using their corresponding OCLC master records
June-Aug.	Task Force developed, tested and finalized the Python scripts
Last week of Aug.	Task Force configured and tested GenLoad profile for loading OCLC master records. Following the successful test loads, FLVC approved the GenLoad profile.
Sep. 3	Task Force requested and received an updated report from FLVC that included 222,404 Shared Bib records with multiple series.
Sep.	Task Force executed the Python script against the new report resulting in the identification of the following: • 53,802 records in the Suggest Overlay Set • 106 duplicate records from Suggest Overlay Set which were sent for deduplication • 243 Shared Bib records whose OCLC number needed to be updated due to merge of OCLC master records
Oct.	Task Force presented the project at the Council of State University Libraries (CSUL) Cataloging, Authorities and Metadata Committee (CAM) and the FLVC Members Council on Library Services Technical Services Standing Committee (TSSC) meeting. This also began the review period where the Task Force solicited feedback prior to any additional updates.
First two weeks of Nov.	Task Force batch loaded the OCLC master records from Suggest Overlay Set to update problematic records in Shared Bib

Appendix C. Python Script for Format Determination

```
# called from main script, to get format information
# Return Formats
# lFormat = returnFormat(lDict[key])
# lDict[key] is the dictionary of fields for a given MARC record
# mFormat = returnFormat(mDict[aDict[key]])

def returnFormat(dict):
    # extract the code from the 008 23 values
    formatCode = 'None'
    for tag in dict['fields']:
        for k in tag:
            if k == '008':
```

```
            formatCode = tag[k][23:24]
    format = ''
    if formatCode in ['s', 'o', 'q']:
        format = 'electronic'
    elif formatCode in [' ', 'r', 'd']:
        format = 'print'
    elif formatCode in ['a', 'b', 'c']:
        format = 'microform'
    else:
        format = 'unknown'

    return format
```

Appendix D. Python Script for Identification of Local Series Values

```
# part of the main script - calls to the function that does the comparisons
# l440 is the cleaned series strings from the SharedBib MARC 440
# l490 is the cleaned series strings from the SharedBib MARC 490
# l830 is the cleaned series strings from the SharedBib MARC 830
# m490 is the cleaned series strings from the OCLC MARC 490
# m830 is the cleaned series strings from the OCLC MARC 830

# placeholder list, wasteList allows the procedure to send a value to the function as a placeholder
wasteList = ['-1']

#Compare Local440
compResult = betterComparison(l440, m490, m830, wasteList, wasteList)
# this part follows each comparision (is excluded from below cases)
if len(compResult) > 0:
        sendForLocalCheckResults = [lSysNumber, oclcNumberL, '440', local440]
        writeLocalCheckResults(sendForLocalCheckResults, lSysNumber)
        logString = logString+'\n\tComparison Strings Not Found (440):'+'\n\t\t'+compResultString
        writeBibsForOverlay(lSysNumber, oclcNumberL, '0')
        logResult(str(keyCounter), logString)
        keyCounter += 1
        continue

#Compare Local490
compResult = betterComparison(l490, m490, m830, wasteList, wasteList)

#Compare Local830
compResult = betterComparison(l830, m830, wasteList, wasteList, wasteList)

# the betterComparison function called that actually does the comparisions.
def betterComparison(lista, listb, listc, listd, liste):
    unfoundSeriesStringList = []
    badEndingValues = ['V']
    beginningWords = ['he ', 'her', 'his', 'she']

    listaa = []
    listbb = []
    listcc = []
```

```
    listdd = []
    listee = []

    for a in lista:
        if len(a) > 0 and a.upper()[-1] in badEndingValues:
            a = a[0:len(a)-1].strip()
        listaa.append(a.upper())
    for a in listb:
        if len(a) > 0 and a.upper()[-1] in badEndingValues:
            a = a[0:len(a)-1].strip()
        listbb.append(a.upper())
    for a in listc:
        if len(a) > 0 and a.upper()[-1] in badEndingValues:
            a = a[0:len(a)-1].strip()
        listcc.append(a.upper())
    for a in listd:
        if len(a) > 0 and a.upper()[-1] in badEndingValues:
            a = a[0:len(a)-1].strip()
        listdd.append(a.upper())
    for a in liste:
        if len(a) > 0 and a.upper()[-1] in badEndingValues:
            a = a[0:len(a)-1].strip()
        listee.append(a.upper())

    for series in listaa:
        if series in listbb:
            continue
        if series in listcc:
            continue
        if series in listdd:
            continue
        if series in listee:
            continue

        unfoundSeriesStringList.append(series)

    return unfoundSeriesStringList
```

Appendix E. Example of a Shared Bib Record before and after the Overlay Process

Before

=001 020014504
=035 __$a(OCoLC)00001935
=040 __$aDLC$cDLC$dm.c.$dFNP
=050 0_$aBL1405$b.D4
=050 0_$aBQ1138$b.D4
=090 __$aBL1405$b.D4
=092 __$a294.3$bD286b
=100 1_$aDe Bary, William Theodore,$d1919-$ecomp.
=245 14$aThe Buddhist tradition in India, China & Japan.$cEdited by Wm. Theodore De Bary. With the collaboration of Yoshito Hakeda and Philip Yampolsky and with contributions by A. L. Basham, Leon Hurvitz, and Ryusaku Tsunoda.

=260 __$aNew York,$bModern Library$c[1969]
=300 __$axxii, 417 p.$c20 cm.
=**440** _0$aReadings in Oriental thought**$5FTS$5FTaSU**
=**440** _4$aThe Modern library of the world's best books$v<205>**$5FTS**
=**490** **0**_$aThe Modern library of the world's best books [205]**$5FTaSU$5FPeU$5FU$5FMFIU**
=**490** **0**_$aReadings in Oriental thought**$5FPeU$5FJUNF$5FBoU$5FU$5FMFIU**
=**490** **0**_$aThe Modern library of the world's best books**$5FBoU**
=**490** **1**_$aThe Modern library of the world's best books$v[205]**$5FJUNF**
=504 __$aBibliography: p. [399]-401. Bibliographical footnotes.
=650 _0$aBuddhism$xCollections.
=650 _0$aBuddhism$xSacred books.
=**830** _0$aModern library of the world's best books;$v[205]
=**830** _0$aReadings in Oriental thought.
=**899** _0$aWedig collection.$5FMFIU
=951 __102aFAU01:000255331;$5FBoU
=951 __104aFIU01:001349334;$5FMFIU
=951 __105aFSU01:000547336;$5FTaSU
=951 __109aNFU01:000231416;$5FJUNF
=951 __110aSFU01:000000770;$5FTS
=951 __108aUFU01:000812032;$5FU
=951 __111aWFU01:000222854;$5FPeU

After

As a result of the overlay process, MARC 440, 490, 830, and other fields were updated to reflect the OCLC master record. As an added benefit, the Shared Bib record received the more complete data in the master record including the MARC 33x fields, FAST headings, extra subject access points, MARC 505, and 710 fields were added. Local fields on the Shared Bib record, including MARC 899 and 951 fields, were protected.

=001 020014504
=019 __$a1261666$a462181729$a911553216
=035 __$a(OCoLC)00001935
=040 __$aDLC$beng$cDLC$dOCL$dBTCTA$dITC$dCBC$dHIL$dDEBBG$dOCLCF $dP4I$dOCLCQ$dOCLCO$dTWS$dTAMSA
=050 00$aBQ1138$b.D4
=050 14$aBL1405$b.D4
=100 1_$aDe Bary, William Theodore,$d1919-$ecompiler.
=245 14$aThe Buddhist tradition in India, China & Japan.$cEdited by Wm. Theodore De Bary. With the collaboration of Yoshito Hakeda and Philip Yampolsky and with contributions by A. L. Basham, Leon Hurvitz, and Ryusaku Tsunoda.
=260 __$aNew York,$bModern Library$c[1969]
=300 __$axxii, 417 pages$c20 cm.
=336 __$atext$btxt$2rdacontent
=337 __$aunmediated$bn$2rdamedia
=338 __$avolume$bnc$2rdacarrier
=**490** **1**_$aReadings in Oriental thought
=**490** **1**_$aThe Modern library of the world's best books [205]
=504 __$aIncludes bibliographical references (pages 399-401. Bibliographical footnotes).
=505 0_$aEarly Buddhism -- The life of Buddha as a way of salvation -- "The greater vehicle" of Mahayana Buddhism -- Tantricism and the decline of Buddhism in India -- The coming of Buddhism to China -- The schools of Chinese Buddhism -- The introduction of Buddhism to Japan -- Saicho and the lotus teaching -- Kukai and esoteric Buddhism -- Amida and the pure land -- Nichiren's faith in the lotus -- Zen.
=650 _0$aBuddhism$vSacred books.
=650 _7$aBuddhism.$2fast$0(OCoLC)fst00840028

=650 07$aBuddhismus.$2swd
=651 _7$aChina.$2swd
=651 _7$aIndien.$2swd
=651 _7$aJapan.$2swd
=650 07$aBuddhismus.$0(DE-588)4008690-2$2gnd
=651 _7$aChina.$0(DE-588)4009937-4$2gnd
=651 _7$aIndien.$0(DE-588)4026722-2$2gnd
=651 _7$aJapan.$0(DE-588)4028495-5$2gnd
=655 _7$aCollections.$2fast$0(OCoLC)fst01424032
=710 2_$aRogers D. Spotswood Collection.$5TxSaTAM
=776 08$iOnline version:$aDe Bary, William Theodore, 1919-$tBuddhist tradition in India, China & Japan.$dNew York, Modern Library [1969]$w(OCoLC)610373932
=**830** _0$aModern library of the world›s best books ;$v205.
=**830** _0$aReadings in Oriental thought.
=**899** _0$aWedig collection.$5FMFIU
=951 __102aFAU01:000255331;$5FBoU
=951 __104aFIU01:001349334;$5FMFIU
=951 __105aFSU01:000547336;$5FTaSU
=951 __109aNFU01:000231416;$5FJUNF
=951 __110aSFU01:000000770;$5FTS
=951 __108aUFU01:000812032;$5FU
=951 __111aWFU01:000222854;$5FPeU

Notes on Operations

GMD or No GMD

RDA Implementation for a Consortial Catalog

James Kalwara, Melody Dale, and Marty Coleman

This paper explores the benefits of establishing item-specific terms for General Material Designations (GMDs) for library consortia implementing Resource Description and Access (RDA). While RDA includes a new approach towards the description and categorization of an item's physical medium through the assignment of content, media, and carrier types (CMCs), thus replacing the GMD, libraries may still benefit from GMD retention in their online catalogs to help support user tasks and help contextualize CMC information. This paper presents the challenges that Mississippi State University Libraries experienced in leading RDA enrichment for the Mississippi Library Partnership (MLP) consortium. Additionally, it discusses parameters for libraries to consider when working with a vendor for RDA enrichment in a consortial environment.

The Library of Congress's implementation of RDA in March 2013 prompted many libraries to reassess their local cataloging practices. A major change with RDA was the change from the General Material Designation (GMD) to the content, media, and carrier types (CMCs) provided in MARC 336, 337, and 338 (commonly referred to as 33X) fields.[1] In 2010, a librarian voiced concern at the loss of GMDs when addressing a columnist: "Dear Elsie, Is it true that the GMD will disappear with RDA? If so, how will we alert our patrons, and ourselves, to the fact that a title is a CD, a DVD, and so on? Designated, and would like to stay that way, in Decatur."[2] At that time, Mississippi State University's (MSU) catalogers shared the same concern and began taking steps to develop training and implementation plans for this new cataloging standard while considering the future of GMDs in the consortial catalog.

During RDA training, MSU's catalogers discussed display and indexing decisions for RDA elements and whether the GMD would remain useful with the standard's new rules. Catalogers agreed that GMDs contextualize 33X terms and clearly differentiate materials that share the same title. After discussing this concern with various library departments and consortial partners, MSU's catalogers determined that retaining GMDs remained essential to supporting resource discoverability. However, catalogers would need to update legacy GMD terms by selecting more item-specific terms to better support user tasks. Members of several departments in MSU Libraries suggested using "common terms" in place of GMDs for their local bibliographic records to support patron search behaviors in the consortial catalog. For instance, the common term "DVD" would in some cases replace the GMD "videorecording," and similarly, the common term "MP3" would replace the GMD "electronic resource" in some cases.

The decision to implement RDA and to retain GMDs affected the MLP, which is comprised of fifty-four libraries with distinct user needs. As several consortial libraries already used non-standard common terms locally in place of GMDs, a collective decision was reached to continue this practice to establish consistent metadata across the catalog while proceeding with RDA enrichment.

James Kalwara (james.kalwara@colorado.edu) is a Monographic Cataloger at the University of Colorado Boulder. **Melody Dale** (mdale@library.msstate.edu) is an Education Librarian at Mississippi State University. **Marty Coleman** (mcoleman@library.msstate.edu) is an Acquisitions Librarian at Mississippi State University.

Manuscript submitted May 25, 2016; returned to authors for revision September 2, 2016; revised manuscript submitted October 30, 2016; manuscript returned to authors for minor revision January 4, 2017; revised manuscript submitted February 3, 2017; accepted for publication March 31, 2017.

This paper is based on a presentation that was delivered at the Southeastern SirsiDynix Regional Users Group Conference on Tuesday, August 4, 2015 at Mississippi State University and also at the Association of Library Collections and Technical Services Catalog Management Interest Group Meeting at the American Library Association Midwinter Meeting on Saturday, January 9, 2016 at the Boston Convention & Exhibition Center.

The authors would like to thank Anita Winger at Mississippi State University for supplying vital data and timelines for this paper.

MLP libraries agreed to use the established common terms and to incorporate RDA elements and practices into non-RDA records, including spelling out abbreviations found in the MARC 300 and 504 fields, adding the 33X fields, and converting the 260 field to appropriate 264 fields, to maintain consistency. Additionally, MSU Libraries agreed to provide training and documentation to MLP members as needed.

Literature Review

This literature review explores both the historical and the current climate of GMD usage in library catalogs and how libraries have respectively handled GMD replacement and CMC inclusion following RDA implementation. GMDs originated from the necessity to distinguish between different material types with the same title. For instance, the GMD "videorecording" distinguishes a title's medium from other possible manifestations of the same title, including a sound recording or an electronic resource.

Initially, media and print materials were housed in separate catalogs; however, in some libraries, media items were uncataloged and simply stored in particular workrooms.[3] In the 1960s, libraries recognized the advantages of providing bibliographic records for all material types within a unified catalog.[4] Libraries began using media codes, which were later renamed media designators, to identify non-print materials.[5] The Anglo-American Cataloging Rules First Edition (AACR) standardized a small vocabulary of media designators; however, they were not applied to all types of materials, nor were all media types included in the code.[6] Media designation was renamed to "general material designation" with the second edition of AACR (AACR2), which strategically placed the GMD directly after the title proper to notify the user of an item's physical medium.[7]

While GMDs remain beneficial for users, there are also limitations to their usefulness. Caudle and Schmitz discovered that patrons sought more detailed information regarding a resource's format type than what was presented by the GMD.[8] GMDs also do not consistently communicate an item's mode of issuance or carrier information.[9] For example, the GMD "filmstrip" represents only one physical format; whereas the GMD "sound recording" can represent multiple carrier types, including audio cassettes, compact discs, or audiotape reels. Schmitz argued that providing the mode of issuance and carrier type information for an electronic resource enables users to clearly distinguish between a newspaper and an electronic book, or to clarify whether an audio disc is a music CD or a vinyl record.[10] Oliver indicated that GMDs inconsistently describe an item's physical medium since they represent the attributes of an item on a work, expression, and manifestation level but inadequately provide description on an item level.[11] Ou and Saxon reiterated the shortcomings of GMDs' capacity for item-level description and categorization by illustrating that while the GMD "electronic resource" describes a resource's carrier type, the same resource could also be assigned the GMD "cartographic material," which describes the resource's content type.[12] Furthermore, a motion picture may be assigned the GMD "videorecording," yet when the same title is issued as a streaming video, the GMD "electronic resource" is assigned since that is considered as the primary medium.[13] Additionally, Seikel and Steele suggested that GMDs have become irrelevant with user search patterns, due to updates to terms such as "sound cassette" and "videodisc," which have been superseded by the more commonly used terms "audio tape" and "DVD."[14]

RDA seeks to address and remedy the GMD's limitations and issues by replacing them with CMCs, which are provided in the MARC 336 Content Type, 337 Media Type, and 338 Carrier Type fields.[15] The content type is the form of communication through which a work is expressed.[16] The media type reflects the general type of intermediation device required to view, play, run, or access the content of a resource.[17] The carrier type reflects the format of the storage medium and housing of a carrier in combination with media type.[18] Bernstein suggests that implementing CMCs allows for a more hierarchical structure for categorizing resources that addresses the complexities found in categorizing non-print materials.[19]

While RDA takes a more granular approach to resource description, online public access catalogs (OPACs) and discovery systems are still developing functions to fully support RDA's practical applications. RDA's theoretical foundation is based on the Functional Requirements for Bibliographic Records (FRBR), which focuses on representing entities, attributes, and relationships.[20] However, in 2011, the MARC format had incorporated relatively few developments that could take full advantage of the FRBR model.[21] Since then, the MARC environment and integrated library systems (ILS) remain under development. Cronin illustrated that libraries had varying degrees of control over the indexing and record display options of their ILS.[22] Historically, many institutions have had success in displaying CMC information through open-source software and cultivating support from their systems departments. Currently, major ILS systems offer CMC display options enabling libraries to choose which CMC information to display based on patron needs. By using an open-source online catalog, catalogers and library systems associates at Auburn University customized display functions necessary to display CMC information in their online catalog.[23] Panchyshyn additionally proposed an innovative OPAC solution to commercial online catalogs by combining the item type icon with the RDA carrier type data from a bibliographic record.[24]

RDA conversion and enrichment have recently been prominent in the library literature as more libraries have implemented RDA. Panchyshyn and Park concluded that RDA enrichment is a necessary step to enhance legacy bibliographic metadata, which ultimately improves patron experience in the online catalog.[25] Guajardo and Carlstone described their RDA enrichment procedure, including the addition of material type codes, which served to replace GMDs.[26] While no uniform resolution can correct OPAC display issues, libraries strive to support user tasks by developing their own solutions. Until OPAC and bibliographic systems can fully support all theoretical aspects of RDA, many libraries will continue working towards creating a positive user experience by modifying aspects of national cataloging practices to best support their local needs.

Case Study

MSU Libraries leads cataloging efforts for fifty-four libraries within its statewide consortium, the MLP. Many libraries within the consortium lack adequate staffing and resources to undergo a catalog enrichment project. MSU Libraries' cataloging department includes staff members who provide original and complex cataloging for serials and monographs. Since it is fully staffed, MSU Libraries was well suited to lead RDA enrichment for the consortial catalog and to establish RDA cataloging procedures for the MLP.

MSU Libraries' Trajectory to RDA

Formal discussions regarding MSU Libraries' RDA implementation began in spring 2010 with catalogers tracking the trend via discussion lists. Following various RDA discussion list threads gave MSU's catalogers the opportunity to learn how similar institutions were planning RDA implementation. MSU's catalogers also followed LC's efforts, which began transitioning to RDA in June 2011, with full implementation in March 2013. MSU catalogers realized that in a consortial environment, RDA implementation was not limited to their own cataloging workflows and would also impact MLP's original cataloging practices. RDA was a major discussion topic at the 2013 and 2014 Southeastern SirsiDynix Regional Users Group Conferences (SERUG) where MSU cataloging and systems librarians and the MLP staff members discussed concerns about omitting GMDs when bibliographic records were enriched to incorporate RDA elements and practices.

As members of the Name Authority Cooperative Program (NACO) since 2002, MSU catalogers understood that they would have to incorporate RDA practices into the authority records that they contributed to the LC/NACO Authority File. From July 2011 to November 2012, MSU catalogers received training from an LC representative in creating personal name, corporate body, and series authority records using RDA. After completing RDA NACO training in November 2012, catalogers were ready to implement RDA authority control practices at MSU Libraries. The next step in MSU Libraries' RDA planning included training in creating bibliographic records using RDA.

In September 2013, MSU Libraries applied for OCLC "Enhance" status, which involved a training and review period with an LC representative. This period allowed them the opportunity to create original bibliographic records that were sent to a reviewer who provided feedback prior to contributing master records to WorldCat. Following the review period, in late 2013, MSU Libraries received "Enhance" status, which then presented the opportunity to apply to the LC Monographic Bibliographic Record Cooperative Program (BIBCO). After completing four webinars and training with an LC representative to learn how to create original RDA BIBCO bibliographic records and how to enhance non-RDA bibliographic records to BIBCO status, MSU Libraries were granted BIBCO authorization in April 2014.

By gaining independence to contribute RDA BIBCO bibliographic and NACO authority records, MSU catalogers demonstrated that they were prepared to implement RDA policy standards both at MSU Libraries and to their MLP partners. In addition to establishing RDA standards policy documentation for MLP's original and copy cataloging procedures, an important component of the proposed RDA implementation was enriching bibliographic records in the online catalog to incorporate RDA elements, thus hybridizing its catalog. This enrichment entailed adding RDA elements including the 33X fields, spelling out abbreviations in the 300 and 504 fields, and converting the publication statement from the 260 to 264 field in all bibliographic records in the consortial catalog. By enriching its records with the aforementioned RDA elements, the consortial catalog would provide clean and consistent metadata for its users. Planning for RDA enrichment and implementation additionally prompted MSU Libraries and the MLP to reassess its vocabulary of GMDs and discuss their retention in the catalog.

Retaining GMDs

In addition to MSU's cataloging unit, several committees were involved in the discussion of GMD retention, including MSU's Library Technologies Committee, the OPAC subcommittee, the Library Administrative Council, and an ad-hoc committee consisting of MLP and MSU Libraries' personnel. These committees collaborated in the decision-making process, concluding that retaining GMDs throughout RDA enrichment in the consortial catalog would best support user tasks and establish consistent metadata for

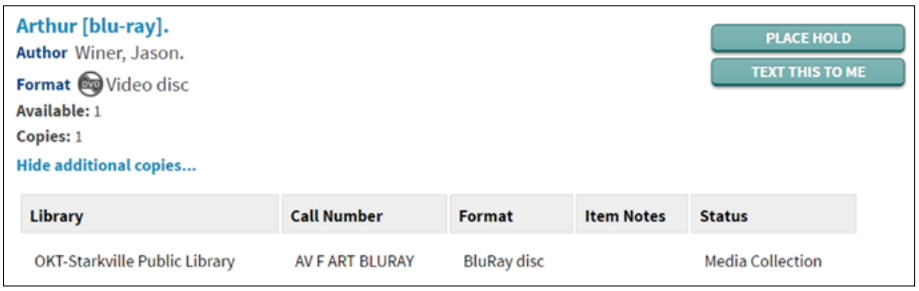

Figure 1. Example of GMD that Clarifies Format Icon Information

user convenience. MSU's cataloging and computer systems departments moved forward with RDA enrichment and implementation.

RDA Implementation

In April 2014, Backstage Library Works (Backstage) approached MSU Libraries to serve as a testbed institution for vendor-supplied RDA enrichment. Backstage provides the benefit of establishing a customized RDA Profile, allowing an institution to define MARC data element parameters for RDA enrichment. To establish an RDA Profile for MSU Libraries, coding options for legacy bibliographic and authority records in the online catalog were explored to reflect specific RDA elements. By creating a customized RDA Profile for bibliographic record validation and authority record cleanup, MSU Libraries established preferences used in creating an algorithm to generate newly revised GMDs, replace the 260 with 264 fields, add 33X fields, and spell out abbreviations in the 300 and 504 fields of bibliographic records for subsequent quarterly batch-record loading.

Feedback on user information seeking behaviors was compiled from discussions with public services librarians. The cataloging department met informally with staff members from the MSU Libraries' Research Services Department from May to June 2014 to gain broader perspectives on the usage of GMDs. Research Services librarians indicated that patrons, research librarians, and support staff relied heavily on GMDs to support searching, identifying, and in some cases selecting resources of interest. They concluded that GMDs were most useful in instances when the GMD differed from the icon used to display an item's format type.

Figure 1 provides an example of how a GMD may be useful in identifying an item's format despite ILS display limitations. In this case, the only icon displayed in the ILS is for a DVD. With the GMD's presence, patrons can more easily interpret and confirm information describing the item's format as a Blu-ray disc. Catalogers also met with MSU's electronic resources personnel to discuss expanding terms that might improve description of various types of electronic resources in the online catalog.

MSU catalogers felt strongly about retaining GMDs to support their workflows, specifically when performing routine database maintenance and copy cataloging procedures, such as differentiating between media and non-media titles published in multiple formats. Establishing consistent and clean metadata for the catalog was another major concern in deciding to retain GMDs. After receiving feedback from various departments and cataloger recommendations, MSU Libraries concluded that retaining GMDs would best establish clean metadata for the consortial catalog and best support user tasks of searching, identifying, and selecting when conducting basic or advanced search queries in the online catalog. However, after reviewing previously used GMDs, it became apparent to MSU and MLP catalogers that revising GMDs would maximize their usefulness and enhance the user experience.

GMD Expansion and Updating the Catalog's Existing GMDs

The first step in revising GMDs was to review a list of "common terms" from Backstage, presented in figure 2, that could appropriately replace GMDs.[27]

While some of the listed "common terms" were already being used by a handful of consortial libraries, many, including MSU Libraries, used AACR2's GMDs for cataloging practices. After reviewing the list of Backstage "common terms" and AACR2 GMDs, MSU catalogers identified outdated GMDs that could be revised or omitted.[28] To do this, catalogers discussed each common term from the Backstage list designated for a particular format type. For example, figure 3 displays four possible options, found in green boxes, which represent a DVD video, including "DVD," "videodisc (DVD)," "videodisc," and "videorecording (DVD)." After discussing the clarity of each option, catalogers concluded "DVD" was the clearest term that would best support user needs.

To ensure that GMDs appropriately applied to their corresponding item type in an item record, catalogers mapped them to all known item types from the catalog, and then determined how they could be separated and enhanced for clarity (see hdl.handle.net/11668/13644). For instance, every item type that corresponded to the GMD "videorecording" was recorded in an Excel spreadsheet. These included "BLURAY," "DVD," "DVD-SET," "DVD-ROM," "VIDEO," "NF-AV," and "VHS." Catalogers devised new "common term" GMDs to include "Blu-ray," "DVD," "electronic

activity card	globe	record
art original	government document	serial
art reproduction	graphic	slide set
atlas	kit	sound recording
audiocassette	large print	sound recording (cassette)
braille	laser disc	sound recording (compact disc)
cartographic chart	LP	sound recording (CD)
cartographic material	manuscript	sound recording (LP)
cartographic material (tactile)	map	study print
CD recording	map (tactile)	technical drawing
CD-ROM	microfiche	text
CDV	microfilm	text (large print)
chart	microform	toy
chart (large print)	microopaque	transparency
compact disc	microprint	US document
diorama	microscope slide	VHS
DVD	model	video CD
DVD-ROM	motion picture	video single disc
electronic resource	music	videocassette
electronic resource (CD-ROM)	music (braille)	videodisc
federal document	newspaper	videodisc (DVD)
filmstrip	periodical	videorecording
flash card	photograph	videorecording (DVD)
floppy	picture	videorecording (VHS)
game	realia	VSD

Figure 2. Backstage "Common Terms" List

video," "nonfiction audiovisual," and "VHS" that would be used to replace each previous instance of the GMD "videorecording." Referencing the Backstage table helped catalogers to create granular and item-specific GMDs. For example, when prompted to derive a new GMD for "DVD-ROM," catalogers concluded that "DVD-ROM (computer only)" would provide clear information about both the item's carrier and media type to contextualize the RDA terms used in the MARC 337 and 338 fields.

Catalogers also revised the GMD "electronic resource" for the consortial catalog. They agreed that this GMD should provide more granular information regarding its description and categorization since "electronic resource" was previously used to describe and categorize multiple format types of electronic origin. By deriving common terms such as "CD-ROM," "computer game," "video game," "electronic book," "electronic video," and "software," MSU catalogers created item-specific GMDs to allow users to easily disambiguate item types of various library materials that were similarly categorized as "electronic resource."

After catalogers revised legacy GMDs, the newly revised GMDs were mapped to the corresponding RDA 33X fields, and then to the item types to create a Word document table that MSU and MLP catalogers could use as a reference tool (see lib.msstate.edu/_assets/docs/mlp/MLP-RDA.pdf). This table featured a comprehensive list of mappings from the item type to 33X fields to the revised GMD. After revising GMDs for the MLP in December 2014, MSU Libraries distributed the Word document table mapping GMDs to item types and to the newly revised "common term" GMDs to each MLP cataloging unit.

atlas	audiocassette	cartographic chart	cd recording
cd-rom	cdv	compact disc	dvd
dvd-rom	electronic resource (cd-rom)	equipment	federal document
floppy	globe	government document	graphic
large print	laser disc	lp	map
map (tactile)	microfiche	microfilm	microopaque
microprint	miscellaneous	newspaper	periodical
photograph	record	serial	slide set
sound recording (cd)	sound recording (lp)	sound recording (cassette)	sound recording (compact disc)
study print	us document	vhs	video single disc
video-cd	videocassette	videodisc	videodisc (dvd)
videorecording (dvd)	videorecording (vhs)	vsd	

Figure 3. Choosing Common Terms Options for a DVD

Testing the Process

Prior to the conversion, Backstage developed an algorithm that extracted data values from the Leader, 007, 008, 245, 300, 500, and 538 fields in a bibliographic record to generate the RDA 33X fields and the newly revised GMDs. To test this algorithm, MSU Libraries and Backstage began implementing the conversion with a sample of bibliographic records that featured various item types including audio recordings, books, electronic resources, graphic novels, government documents, media, microforms, and photographs that were randomly selected from different libraries to ensure that there was appropriate representation from the consortium.

This sample was tested three times, and four catalogers spent an estimated 100 hours reviewing the tests. The first test file contained 793 bibliographic records and was sent to Backstage for processing in late October 2014, and was returned to MSU Libraries on October 30, 2014. This file tested the basic RDA enrichment algorithm provided by Backstage without changes to GMDs. In addition to basic RDA enrichment, the second test file also assigned the revised "common term" GMDs, and was received from Backstage on November 12, 2014. Catalogers reviewed these bibliographic records to ensure the modifications to Backstage's algorithm were accurately generating the newly revised GMDs.

For the third test file, the Computer Systems Assistant added fourteen bibliographic records to the initial 793 records, including three titles from the Early English books collection on microfilm, which features various "bound-with titles" (mlp.ent.sirsi.net/client/en_US/msstate/search/detailnonmodal/ent:$002f$002fSD_ILS$002f0$002fSD_ILS:2861289/ada?rt=CKEY|||CKEY|||false) and added eleven titles with mixed media item types such as books with

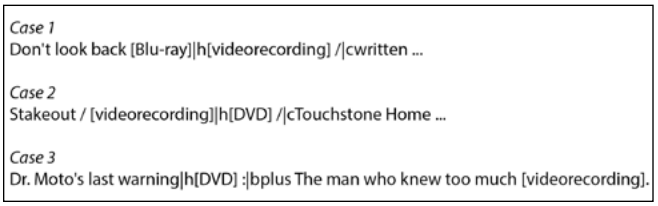

Figure 4. Examples of Two GMDs within Single Title Statement

accompanying CDs. By adding these bibliographic records, catalogers confirmed how Backstage's algorithm modified records for these complex formats. The third test file was returned to MSU Libraries on December 12, 2014. After reviewing results from the third test file, MSU Libraries decided to proceed with full RDA enrichment for the entire consortial catalog.

MSU's Computer Systems Assistant sent the full set of 1,800,186 bibliographic records from the catalog to Backstage on December 15, 2014 to be processed. This set was returned to MSU Libraries on December 19, 2014 to be loaded into the online catalog. Due to the holiday recess, the loading process on the production server did not begin until January 6, 2015. The entire process took one week, and was not without issues. During this process, the catalog was unavailable for editing bibliographic records until all data was loaded on the production server. However, patrons could still access and use the catalog during this time.

Findings

Several months after the conversion, the authors systematically reviewed RDA-enriched records to check for accuracy, as they sought to identify, correct, and prevent future problems. Some of the inconsistencies found included incorrect 33X fields, multiple GMDs, and incorrect GMDs. Various keyword searches were conducted in the online catalog to identify GMD issues that may have occurred following RDA enrichment. Since Backstage processes and converts newly loaded records quarterly, the authors limited search results to retrieve titles loaded before September 30, 2015 to eliminate any identifiable post-enrichment errors that may have been corrected during subsequent Backstage processing.

The authors discovered many inconsistencies that resulted from cataloging errors, including the presence of two GMDs, only one of which was in MARC subfield $h. One particular example included a title statement with two GMDs, "Blu-ray" and "videorecording." Figure 4 illustrates three cases with Blu-ray or DVD titles in which two existing GMDs were provided within a single title statement.

MSU catalogers have yet to determine a consistent reason for two simultaneously occurring GMDs, but will correct these errors in the future. Catalogers identified this problem when conducting search queries using the catalog's "Advanced Search" page for previously used GMDs such as "videorecording" and "sound recording" and revised GMDs such as "DVD" and "music CD." Table 1 presents search results yielding two simultaneously existing GMDs, one obsolete and one revised term, within a single title statement.

Table 1. Search Results for Two Existing GMDs in Online Catalog

Search Query	Titles Retrieved
"blu-ray" and "videorecording"	41
"DVD" and "videorecording"	734
"VHS" and "videorecording"	24
"CD" and "sound recording"	219
"audiocassette" and "sound recording"	8
"music CD" and "sound recording"	11

Catalogers also discovered incorrect 33X fields, which they concluded were generated from Backstage's algorithm. Since Backstage extracted data from the 007 fields to generate the 33X fields, a record with either an incorrect 007 field or multiple 007 fields consequently produced incorrect 33X fields. For instance, if the item type was designated as a DVD, the first data element in the 007 field should have been coded as "v" while second data element should have been coded as "d." However, there were twenty-four records in which the second 007 data element was coded as "f," indicating the item is designated as a videocassette, which consequently generated the GMD "VHS."

Many of the records with this issue were created following earlier local policies, in which print and electronic versions of works were represented in the same bibliographic record. When the records were later separated to represent their specific format types, the 007 field was retained on the record that represented the print version of the work due to lack of appropriate bibliographic and item record maintenance. Problems with the 007 field data accuracy caused issues generating 33X fields and also with the OPAC's format display. When performing routine tasks in the ILS, serials and monographic catalogers discovered that GMDs and the format icons did not consistently match for every record, prompting them to review the OPAC for further display issues and inconsistencies.

After catalogers recognized that incorrect 007 field data could generate inaccurate GMDs and 33X fields in the bibliographic record and icons in the OPAC display, they initiated a policy change in the treatment of 007 fields. First, MSU catalogers must review the coded data element in the first 007 field of a bibliographic record and confirm whether it accurately categorizes and describes the primary format type of the work in hand. Second, should multiple 007 fields

Figure 5. Example of Title with Display Icon for Globe and GMD "vinyl record"

exist in one bibliographic record, catalogers must identify the 007 field that corresponds to the primary item type in hand and revise the order of the 007 fields to list the primary 007 field first. By reorganizing the 007 fields to list the primary item type as the first 007 field, the OPAC display generates the appropriate icon for an item type.

Another frequently occurring error was the presence of the 007 field for globes in bibliographic records for vinyl records, which was identified in 305 records. After discovering this error, catalogers searched the online catalog for all titles with this format and discovered 357 results, of which only one correctly represented globe as a resource. While the origin of this error remains unclear, catalogers will correct this issue as part of the ongoing cleanup. Figure 5 illustrates instances where the GMD "vinyl record" was present in a record with corresponding item type designated as "globe."

After catalogers discovered inconsistencies between 007 fields and GMDs, they coordinated with the acquisitions and systems departments to develop reports comparing information in the 007 field with the GMD. Each report was produced as an Excel spreadsheet and provided a listing of titles with 007 information and its corresponding GMDs. For instance, from 5,519 individual titles with the first two 007 subfield elements coded as "a" and "j," the assigned GMDs included eleven with "electronic book," sixty-four with "CD-ROM," thirty-four with "cartographic material," three with "graphic," 3,124 with "electronic resource," and 2,218 with "map."

Inconsistencies in the designation of certain item types also generated incorrect GMDs. While the majority of libraries in the consortium used the item types "audio-cd" for audiobooks and "music-cd" for music CDs, several libraries assigned inaccurate designations and used the item type "audio-cd" to categorize music CDs. Since GMDs were mapped to item types, all music CDs with a corresponding item type of "audio-cd" had a GMD of "audiobook" following the conversion. Similarly, one library system within the consortium used an item type designated as "NF-AV" for nonfiction audiovisual materials. This item type was established to enable catalogers to designate an extended circulation period for nonfiction audiovisual materials than was previously available and to distinguish circulation periods for fiction audiovisual materials. However, this new item type, which was designed to include several different item format types including DVDs, VHS tapes, and CDs, was problematic when when MSU catalogers tried to revise GMDs for the consortia. Catalogers devised the GMD "nonfiction audiovisual" as a temporary solution and will reassess its value in the future. Although the online catalog currently includes an estimated 4,100 records with the GMD "nonfiction audiovisual," the location designation and icon provided help to clarify the corresponding item format. Figure 6 represents an online catalog record display with usage of the GMD "nonfiction audiovisual."

Following the conversion, the Cataloging Department worked closely with the Systems Department to generate reports reflecting conversion errors, including reports containing 007 fields and item types. Catalogers reviewed those reports to identify the most frequently occurring errors, to plan for subsequent online catalog maintenance, and to prevent errors in quarterly batch loading services from Backstage. Although catalogers initially ran bibliographic record reports and conducted searches to identify errors related to the conversion, unrelated errors were also discovered. For example, the authors discovered a significant batch of previously unreceived order records in the ILS, which were identified for future maintenance. As a result, the authors are currently coordinating with others at MSU Libraries to develop hands-on training sessions and enhance cataloging and acquisitions documentation for internal and MLP practices.

Discussion

In addition to developing training sessions, MSU Libraries catalogers continue to update local cataloging policies to establish consistent practices across the MLP. For instance, to prevent previously noted display errors originating from multiple 007 fields, all monographic catalogers met in October 2015 to establish a new policy, which involved analyzing all 007 fields in a bibliographic record prior to linking an item record. According to the new policy, only the 007 field correlating to a resource's primary format type are retained in the bibliographic record; all others are deleted or revised so that the primary 007 field information is listed first in the bibliographic record.

A more recent issue highlights inconsistent cataloging practices within the MLP that caused discrepancies between

the 33X fields and GMDs following RDA enrichment. The authors discovered that several consortial libraries were not creating item records in the catalog for DVDs and other media types, which occurred when a large batch of records were loaded for an online streaming video service. Additionally, acquisition records lack item information until they are received, and bibliographic records for cancelled orders are not always deleted.

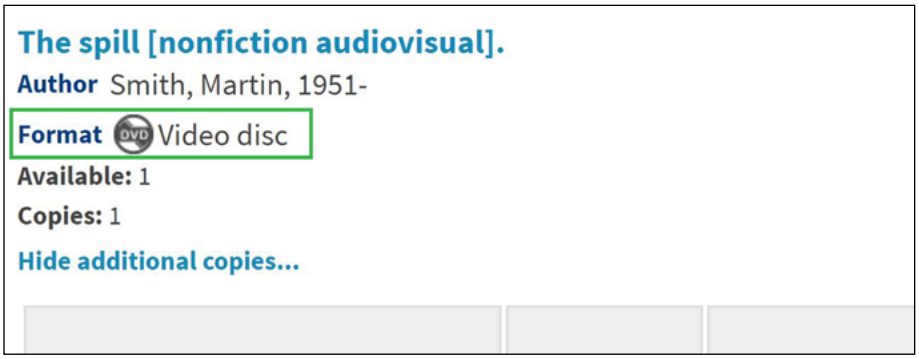

Figure 6. Example of a Title with GMD "nonfiction audiovisual"

After identifying and analyzing issues that occurred post-enrichment, the authors concluded it was necessary to revise MLP cataloging procedural guidelines to ensure that each member library follows consistent practices for original and copy cataloging. A recent discovery involved creation of original bibliographic records for monographs that did not follow any established cataloging standard. After investigating this issue, it was revealed that two consortial libraries had created brief bibliographic records strictly for local usage and were not applying national cataloging standards. Such practices are problematic since there were pertinent RDA elements missing from the records, plus outdated GMDs. To address these issues and provide updated procedural guidelines, MSU Libraries hosted a training session on June 24, 2016 for MLP catalogers. The authors will also propose regular consortial catalog maintenance sessions for media resources to clean up metadata in the catalog.

While a cataloger can edit individual records manually, retrospective editing of large amounts of data is not ideal or cost-effective. If MSU Libraries pays a vendor for RDA enrichment services, it is logical that they want to limit maintenance post-enrichment. Therefore, it is important that all necessary bibliographic record elements be consistent for vendor-supplied RDA enrichment to be fully effective. Furthermore, without clear policies and procedural guidelines for catalogers to reference, inaccurate cataloging practices will likely continue despite vendor-supplied enrichment services.

Conclusion

As display and functionality of CMC information are still evolving, libraries have the opportunity to supplement this information locally with GMDs to best support their user needs. Although MSU Libraries displays the 33X fields in its OPAC, the clarity of this information is not easily interpreted by patrons. To remedy this issue, creating a vocabulary featuring more granular GMDs derived from "common terms" or item-specific information in the consortial catalog enabled MSU Libraries' catalogers to provide consistent and clean metadata necessary to support the user tasks of selection and identification, which was challenging with the legacy vocabulary of GMDs. While catalogers continue to assess the usefulness of the revised GMDs, patrons have expressed satisfaction with retaining GMDs in the online catalog. Moreover, supplying revised GMDs helps users to clarify the vocabularies used to represent RDA's CMC information displayed in the OPAC. For example, by providing the GMD "music CD," a user has more specific information regarding an item's format and the devices needed to access the content on the item, as compared to the ambiguous information found in the 33X fields such as "audio disc." From this RDA implementation project, MSU Libraries have identified a temporary solution to the limitation of its ILS display functions.

The authors believe that collaborating with the MLP and Backstage to standardize GMD and CMC processing before RDA conversion was the most effective action in providing clean and consistent metadata in the consortial catalog. Mapping item type designations was also useful in revising legacy GMDs. However, this RDA implementation project illustrates the importance of uniform cataloging practices in a consortial environment when considering working with a vendor to enrich or "hybridize" all the consortium's bibliographic records. Without uniform cataloging practices, a vendor-supplied conversion will not yield consistent results.

Furthermore, careful consideration should be taken in selecting a test sample of bibliographic records in the beginning stages of an enrichment process. Without a fully representative and accurate sample of bibliographic records, it is difficult to identify algorithmic problems that appear following batch conversions since there are a multitude of variables involved in the vendor's algorithm for GMD and CMC processing. From this, the authors concluded that bibliographic records must include consistent metadata prior to conversion in order to yield optimal results. By better understanding the vendor algorithm prior to conversion, catalogers

and computer systems librarians can reduce errors that will allow for smooth RDA implementation.

References

1. Sevim McCutcheon, "RDA and the Reference Librarian: What to Expect from the New Cataloging Standard," *Reference Librarian* 53, no. 2 (2012): 125, https://doi.org/10.1080/02763877.2011.607409.
2. "Dear Elsie," *ILA Reporter* 28, no. 6 (2010): 28, accessed October 30, 2016, http://www.ila.org/content/documents/Reporter_1210.pdf.
3. Jean Weihs and Lynne C. Howarth, "Designating Materials: From 'Germane Terms' to Element Types," *Cataloging & Classification Quarterly* 45, no. 4 (2008): 3–24, https://doi.org/10.1300/J104v45n04_02.
4. Ibid.
5. Philip Hider, "A Comparison Between the RDA Taxonomies and End-User Categorizations of Content and Carrier," *Cataloging & Classification Quarterly* 47, no. 6 (2009): 544–60, https://doi.org/10.1080/01639370902929755.
6. Weihs and Howarth, "Designating Materials," 5-6.
7. Ibid.
8. Dana M. Caudle and Cecilia Schmitz, "Keep It Simple: Using RDA's Content, Media, and Carrier Type Fields to Simplify Format Display Issues," *Journal of Library Metadata* 14, no. 3–4 (2014): 222–38, https://doi.org/10.1080/19386389.2014.984572.
9. Chris Oliver, *Introducing RDA: A Guide to the Basics* (Chicago: American Library Association, 2010), 50–51.
10. Caudle and Schmitz, "Keep It Simple," 229.
11. Oliver, *Introducing RDA*, 50–52.
12. Carol Ou and Sean Saxon, "Displaying Content, Media, and Carrier Types in the OPAC: Questions and Considerations," *Journal of Library Metadata* 14, no. 3–4 (2014): 239–54, https://doi.org/10.1080/19386389.2014.990846.
13. Michele Seikel and Thomas Steele, "How MARC Has Changed: The History of the Format and Its Forthcoming Relationship to RDA," *Technical Services Quarterly* 28, no. 3 (2011): 322–34, https://doi.org/10.1080/07317131.2011.574519.
14. Ibid.
15. McCutcheon, "RDA and the Reference Librarian," 125.
16. OCLC, "336 Content Type," *OCLC Bibliographic Formats and Standards*, last modified August 22, 2016, http://www.oclc.org/bibformats/en/3xx/336.html.
17. OCLC, "337 Media Type," *OCLC Bibliographic Formats and Standards*, last modified August 25, 2016, http://www.oclc.org/bibformats/en/3xx/337.html.
18. OCLC, "338 Carrier Type," *OCLC Bibliographic Formats and Standards*, last modified August 22, 2016, http://www.oclc.org/bibformats/en/3xx/338.html.
19. Steven Bernstein, "Beyond Content, Media, and Carrier: RDA Carrier Characteristics," *Cataloging & Classification Quarterly* 52, no. 5 (2014): 463–86, https://doi.org/10.1080/01639374.2014.900839.
20. Oliver, *Introducing RDA*, 17–23.
21. Terry Willan, "RDA in the Library System: Implementation and Beyond," *Catalogue & Index* 163 (2011): 14–17.
22. Christopher Cronin, "From Testing to Implementation: Managing Full-Scale RDA Adoption at the University of Chicago," *Cataloging & Classification Quarterly* 49, no. 7–8 (2011): 626–46, https://doi.org/10.1080/01639374.2011.616263.
23. Caudle and Schmitz, "Keep It Simple," 222–38.
24. Roman S. Panchyshyn, "RDA Display and the General Material Designation: An Innovative Solution," *Cataloging & Classification Quarterly* 52, no. 5 (2014): 487–505, https://doi.org/10.1080/01639374.2014.902893.
25. Roman S. Panchyshyn and Amey L. Park, "Resource Description and Access (RDA) Database Enrichment: The Path to a Hybridized Catalog," *Cataloging & Classification Quarterly* 53, no. 2 (2015): 214–33, https://doi.org/10.1080/01639374.2014.946574.
26. Richard Guajardo and Jamie Carlstone, "Converting Your E-Resource Records to RDA," *Serials Librarian* 68, no. 1–4 (2015): 197–204, https://doi.org/10.1080/0361526X.2015.1025654.
27. Backstage Library Works, "Common Terms: Level 2," accessed September 30, 2016, http://ac.bslw.com/mars/guide/RDA_Planning_Guide.pdf.
28. Backstage Library Works, "AACR2 GMD Terms Checked," accessed September 30, 2016, http://ac.bslw.com/mars/guide/RDA_Planning_Guide.pdf.

Book Reviews

Elyssa M. Gould

Digital Library Programs for Libraries and Archives: Developing, Managing, and Sustaining Unique Digital Collections. By Aaron D. Purcell. Chicago: ALA Neal-Shuman, 2016. 256 p. $85.00 softcover (ISBN 978-0-8389-1450-2).

Digital Library Programs for Libraries and Archives: Developing, Managing, and Sustaining Unique Digital Collections is a well-organized text that helps readers better understand the historical context and development of digital collections in libraries into the present, and provides a useful step-by-step process for the management and sustainment of digital programs with the goal to move the concept of a digital program into reality. This text serves as a workbook for leaders and managers in libraries and archives and is highly relevant to all levels of staff including students that are involved or interested in the process of creating a digital program. The use of this text can extend to practitioners working with digital collections in government agencies and corporations in the public or private sector. Creating a digital program is still a relatively new endeavor for many institutions with limited resources and is often misunderstood by those with limited knowledge of the process. This text can help these professionals understand the different facets and requirements of creating and sustaining a digital program while maintaining a big picture view.

As a current professor and special collections director at Virginia Tech, and a former archivist at the University of Tennessee, Purcell writes from the collective perspective and experience of someone who has worked with digital collections and created digital programs in an academic environment. Most will find his strategies and exercises applicable for institutions and projects both large and small. The text does not contain technical jargon or explain how to digitize, how to apply digital forensics, recommend specific technology, or standards needed for the preservation of digital objects. Rather, the author provides highly useful strategies and exercises that serve to guide readers like librarians, archivists, students, and anyone involved or interested in creating a digital program. Purcell does not refer to any specific case studies in his text but indicates that the literature is full of stories of other professionals' digital projects and program experiences that make excellent resources for the novice. Purcell points out the existence of limited resources that serve a role to guide practitioners through the systematic process for developing a digital library program and provides this text as a way to fill that gap (xviii).

This book is systematically divided into three parts with many of the chapters ending with "Key Points" that provide a summary of the concepts and ideas covered in the chapter. Part one consists of three chapters that outline the historical context of digital libraries and how various professions influenced their development. In addition, it covers advances in technology, patron expectations and the effect these developments have on library services and the changes in the library's role and environment. Also covered are reasons why digital collections are created, highlights of important aspects of the development, and long-term needs of a digital library program.

The chapters in part two outline the detailed process needed for the creation, planning, and management of a digital library program and, in particular, the importance of creating a vision for the future of the program. The author justifies the preparatory aspects of planning for a digital program while providing reasons why many programs fail. Purcell stresses that developing a vision is crucial to create a "sense of purpose" and is a "powerful motivational tool" that provides opportunities to strengthen and improve the program while creating an image of the future for the program (65).

Challenges and roadblocks are to be expected in different aspects of the process. To overcome many of these issues, Purcell explains the importance of utilizing resources and partnerships wisely, the process of evaluating and selecting materials for digital collections and their consequences, followed by a discussion about preservation needs like standards and metadata, and outreach and methods of sustainability. Acting like a workbook, the set of open-ended questions after each chapter are designed to help the reader thoughtfully organize and develop a plan, prepare for challenges, and define technical needs to efficiently manage the project.

Many librarians are stymied by the technical know-how needed to curate and preserve digital objects. Purcell takes the stress off the practitioner by saying that it is not necessary to be knowledgeable of every aspect about technology in order to be successful (114). Purcell recommends that one or more of the team members working on the project are up to date on the technology, standards, and best practices resulting in a team of professionals whose varied knowledge and experience come together to strengthen the project and its success (114). Working with a team and utilizing partnerships will help ensure the success of the program.

Purcell reviews the technical standards and their importance without making specific recommendations. Particularly helpful is his "technical elements of digitization" on page 118 that entail the four elements: creation, inputs, repository, and output along with figure 7.3. This reviewer found his explanation an easy to understand, simplified version of the OAIS Reference Model.[1] However, Purcell does not specifically name the model as a resource for the reader. The OAIS reference model, immensely influential in digital preservation, outlines a framework of concepts and functions needed for the preservation of digital objects. The reference model is a recommended component for anyone involved in the management of digital programs and should be mentioned as an important resource for digital preservation in this chapter.

Part three consists of eight exercises pertaining to digital library planning and relates to topics covered in the prior chapters. These exercises encourage the reader to engage thoughtfully to build a vision and create a variety of plans and preparational lists that support the creation and development of a digital program. The exercises are followed by a bibliography and list of relevant websites for those interested in learning about best practices and other institutions involved in the world of digital collections.

Scanning a collection of images is not all it takes to create a digital program. The author fulfills his goal to outline the varied and faceted aspects necessary to run and maintain a digital program or project with attention to the varieties of necessary metadata, and standards, without making specific recommendations. Other necessary aspects include a preservation plan, consideration of technological needs or limitations, followed by buy-in, long-term support, outreach, and integrating the day-to-day processes into the daily workflow of the department. Overall, Purcell has provided a detailed, thorough, and thoughtful step-by-step process for beginning practitioners who are interested in creating, managing, and sustaining digital archives programs. This reviewer highly recommends this text for practitioners who need guidance and those who can use a refresher. Both are bound to pick up new ideas to enhance their digital program management skills.—*Meghan Bailey (meghan.bailey@umb.edu), University of Massachusetts Boston, Boston, Massachusetts*

References

1. Consultative Committee for Space Data Systems - CCSDS. *Reference Model for an Open Archival Information System (OAIS): Recommended practice, Issue 2. CCSDS 650.0-M-2. Magenta Book* (Washington, DC: CCSDS, June 2012), http://urlib.net/sid.inpe.br/mtc-m18/2012/07.12.18.08.

Managing Metadata in Web-scale Discovery Systems. Ed. Louise F. Spiteri. London: Facet Publishing, 2016. 197 p. $85.00 paperback (ISBN 978-1-78330-069-3); hardback (ISBN 978-1-78330-116-4); e-book (ISBN 978-1-78330-154-6).

Managing metadata in libraries today presents challenges to information professionals concerned with quality control, providing relevant search results, and taming the volume of items available for access in a web-scale discovery system. No longer are libraries limited to the collections they "own." Catalogers and metadata professionals now assume the responsibility of providing access to millions of resources, often with limitations on who can access that resource. Relationships with vendors provide opportunities to help manage the gargantuan scale of information. Of course those opportunities come with their own problems as relationships among vendors can be contentious, leaving metadata managers to figure out quality control on a grand scale. In addition to this politicized information landscape, new ways of managing and creating metadata are emerging, leaving information professionals with the task of managing multiple schema in different formats. The essays in *Managing Metadata in Web-scale Discovery Systems* seek to address issues in managing the large scale of information overwhelming catalogers today, with potential solutions for taming the beast of exponentially increasing data.

The book begins with an essay on sharing metadata by Marshall Breeding, Angela Kroeger, and Heather Moulaison Sandy. The authors provide an overview of how discovery works in libraries compared to the historical aspects of cataloging. The current landscape of discovery services offered by the top vendors in our profession, such as ProQuest and EBSCO, are discussed in length. When comparing these new discovery tools with traditional library catalogs, some of the features of discovery are problematic to quality control. The size and scope of a centralized index means librarians must work closely and diligently with vendors to provide the best data with many disparate metadata schema, which can sometimes be inoperable if not properly encoded or mapped. Other problems librarians encounter have more to do with volatile vendor relationships, resulting in having to choose a system that works best to provide access to local subscriptions. Understanding the system in which a librarian works is also crucial to providing the best access in these new systems. Breeding, et. al. leaves us with the task of focusing efforts "on improving shared metadata, rather than on making local enhancements that benefit only a single catalogue" (42). The end goal of improving interoperability becomes increasingly important as more and more data from outside the library becomes available.

In "Managing linked open data across discovery systems," Ali Shiri and Danoosh Davoodi address the

responsibility of libraries to open their resources as linked data. They discuss the benefits, as these expand opportunities for libraries to enhance the findability of their resources. The authors address opportunities for development of linked data through the advancement of projects such as BIBFRAME. Though they do not address how librarians will educate themselves and implement linked data in their own libraries, there are examples provided in the library world to follow as developments in linked data unfold.

A common theme in many of the chapters touches on quality control in library discovery systems, or lack thereof. Christine DeZelar-Tiedman discusses the changes in the management of resources and what those mean in discovery systems, addressing issues such as granularity of description for search and access. She acknowledges the daunting task of managing licensed resources as a balancing act between our use of time and our role as stewards of information resources. Aaron Tay addresses the sheer volume of content in our discovery systems, asking whether providing access to everything risks quality of the returned results. Trying to fill indexes with as much content as possible and relying on relevancy ranking is problematic for libraries trying to maintain the content and the end-user experience. He provides a thoughtful approach as to how libraries will maintain or give up control of resources in the future, and the effect that has on searching. Tay argues that librarians should be thoughtful about the search experience in an index as large as a discovery system. Consider whether users will benefit from the vast amounts of owned and unowned collections a library offers, especially when relying on search results that favor high results over quality ones. In "Managing outsourced metadata in discovery systems," Laurel Tarulli grapples with a healthy conversation about the lack of transparency in discovery systems metadata. The ultimate loser in the fight for transparency with outsourced metadata is the end user. Librarians will have to continue to fight harder for standardized metadata and work closely with vendors to find a balance that benefits their users.

The final chapter, written by editor Louise F. Spiteri, argues for the importance of user-generated metadata. She discusses the social features of discovery systems and the benefits to enhancement of bibliographic information with user-generated content. Her particular focus is on enhancing subject access with social tagging, highlighting the benefits to such library services as readers' advisory.

While the book aims to address issues of quality access of metadata within web-scale discovery systems for all types of librarians, it is most appropriate for academic professionals already managing, or considering management of, data within these systems. There are redundant histories of library data management sprinkled throughout each chapter, which Spiteri addresses in the introduction as intentional. The chapters can therefore be read individually or as a whole; however, there lacks an overall cohesiveness when taken in full. The book has a nice balance of the practical, describing challenges of managing metadata in web-scale discovery systems, and the theoretical, encouraging libraries to explore those "what if" moments in discovery systems. Important conversations about the quality of data being offered in discovery systems take place. As user experience and the search process becomes more and more relevant, the topics in *Managing Metadata in Web-scale Discovery Systems* become critical to librarians who manage large volumes of data in discovery systems.—*Brianne Hagen (brianne.hagen@humboldt.edu), Humboldt State University, Arcata, California*

Linked Data for Cultural Heritage (An ALCTS Monograph). Eds. Ed Jones and Michele Seikel. Chicago: ALA Editions, 2016. 134p. $75.00 softcover (ISBN 978-0-8389-1439-7).

While linked data has been on the horizon for librarians, archivists, and other curators of cultural memory nearly since it was first expounded fifteen years ago, for many it has remained an abstraction.[1] Jones and Seikel present six contributions by those engaged in implementing linked data projects across the cultural heritage landscape, seeking to bridge the gap between the idea of linked data and concrete applications that can be adopted at a local level. The focus is not on the technology of linked data, though each of the chapters discuss some technical issues relevant to the projects, but rather on how the technology can overcome the limits of earlier cultural metadata encoding systems (e.g., MARC) and what new challenges and opportunities it presents. By presenting studies of real-world implementations of linked data, this volume effectively communicates the progress made and a sense of what the technology could do for a local collection.

Again, the collection is not a primer on linked data, or a technical manual or a guide to implementation, but each contribution does discuss some technical aspects. The introduction provides a brief overview of the basic structure of linked data, and individual chapters develop particular issues relevant to the projects described; these descriptions of the structure and syntax of linked data are sufficient to follow how the projects used them, but readers without previous familiarity with the topic may wish to review an introduction to linked data, such as Weese and Segal.[2] Again, while the synopses of the individual projects discuss challenges met, the goal of the work is not to provide a roadmap to exposing your data as linked data, such as is provided by Hyvönen or Hooland and Verborgh.[3] Rather, the intent is to highlight the potentials and challenges of linked data for cultural memory

institutions in their current historical moment, updating and expanding the brief Mitchell (2013), and complementing the even briefer Mitchell (2016).[4]

The challenges of converting existing data into linked data emerged as a common theme among the various projects. The volume as a whole presents a picture that there are a number of tools emerging that can help convert datasets, but that, at present, human intervention continues to be needed, particularly where data in the originating record are ambiguous or the structure of the target linked dataset requires higher granularity. For example, Thorsen and Pattuelli in describing their Linked Jazz program note the development of a transcript analyzer that was used to process interview transcripts, find personal names, and generate triples with the predicate rel:knowsOf. The software could not assign more specific relationships, so the data was crowdsourced to refine those predicates to the likes of rel:collaboratedWith or rel:influencedBy. Godby, in describing the OCLC's testing of conversion of MARC bibliographic records to linked data notes that while published monographs could be converted with minimal intervention, more complex works (her example was a video of a live performance of Tchaikovsky's ballet *The Nutcracker* based on the tale by E. T. A. Hoffman) required substantial intervention, e.g., disambiguating the relation of a personal name in a 700 field as being related to the video, the performance, the ballet, or the tale.

The need for controlled vocabularies appears as another key theme among the different projects. Contrary to earlier expectations that a kind of invisible hand would guide the selection of usable vocabularies in a free-web environment, the contributors share a position that carefully created and maintained vocabularies are necessary to connect local metadata with the larger linked data environment, which is one of the main reasons cultural memory institutions would convert their data to linked data in the first place (33–34). Authority control is the focus of O'Dell's chapter, where she takes the perspective that, since authority control is a mature practice within librarianship, the creation, use, and maintenance of controlled vocabularies is an area where libraries are in a position to make a substantive contribution to the linked data community. Huerga and Lauruhn approach the need for authority control from the perspective of science, technology, and medicine (STM), particularly in view of a changing landscape where research data is increasingly openly available and pressure for STM research to be reproducible. In particular, since several STM vocabularies are already available for linked data, and more are likely to be available soon, they point to the need for metadata specialists to select and apply appropriate vocabularies for local data, and for the need to map equivalencies and near-equivalencies of terms between different vocabularies.

The final two chapters share a concern for, among other things, how linked data representations of bibliographic entities can accommodate the Functional Requirements for Bibliographic Records (FRBR) work/expression/manifestation/item model. Godby, reporting on OCLC's linked data conversion project, describes a working model for distinguishing works from manifestations by clustering records with (near-) identical 1xx and 245 fields, where the cluster represents the work, and is assigned appropriate relationships from the individual records, such as schema:about or schema:genre; members of the cluster are assigned the relationship schema:exampleOfWork, which suffices to identify them as manifestations; a relationship of schema:translationOfWork, derived from 41 and 240 fields is sufficient to identify an expression, and so forth. McCallum, reporting on the development of the Bibliographic Framework Initiative (BIBFRAME) at the Library of Congress, compares the BIBFRAME model of work/instance/item with the FRBR model and notes the resulting issues, for example, that every BIBFRAME instance must have a relationship with a BIBFRAME work, but in the data created in the MARC environment, work entities (i.e., authority files) were created in certain conditions.

Altogether, the volume makes an important contribution to the literature on linked data applications for cultural memory institutions. Anyone considering a project to convert their local metadata to linked data will find current perspectives on such questions as what linked data can or cannot (yet) do, what kinds of tools exist to assist the conversion, what level of human intervention will be needed, why are controlled vocabularies needed, and how can they be found and selected. Not all the answers lie within its pages, but the readers will be better able to understand the scopes of their anticipated projects and predict challenges that are likely to arise.—*Paul Ojennus (pojennus@whitworth.edu), Whitworth University, Spokane, Washington*

References

1. Tim Berners-Lee, James Hendler, and Ora Lassila, "The Semantic Web," *Scientific American* 284, no. 5 (2001): 34-43.
2. Keith P. Weese and Dan Segal, *Libraries and the Semantic Web* (San Rafael, CA: Morgan & Claypool, 2015).
3. Eero Hyvönen, *Publishing and Using Cultural Heritage Linked Data on the Semantic Web* (San Rafael, CA: Morgan & Claypool, 2012); Seth van Hooland and Ruben Verborgh, *Linked Data for Libraries, Archives and Museums: How to Clean, Link and Publish your Metadata* (Chicago: Neal Schuman, 2014).
4. Erik Mitchell, *Library Linked Data: Research and Adoption* (Chicago: ALA TechSource, 2013); Erik Mitchell, *Library Linked Data: Early Activity and Development* (Chicago: ALA TechSource, 2016).

Managing Digital Cultural Objects: Analysis, Discovery, and Retrieval. Eds. Allan Foster and Pauline Rafferty. Chicago: ALA Neal-Schuman, 2016. 227 p. $88.00 softcover (ISBN 978-0-8389-1343-7).

Over the past twenty years, libraries, archives, museums, and other institutions have made hundreds of thousands of digitized and born digital cultural heritage objects available online. This momentum is not likely to slow anytime soon. Digitization programs continue to convert analog media, and efforts are ramping up to procure and preserve born digital material. While discussion of technical specifications and skills to support these processes are critical, there is a growing body of research beyond these topics. Some scholars and practitioners have turned their attention towards theory, assessment, and innovative analysis and *Managing Digital Cultural Objects: Analysis, Discovery, and Retrieval* adds to this conversation.

The book is arranged into three parts, each with three chapters. The first part, "Analysis and retrieval of digital cultural objects," aims to give basic introductory and contextual information; the second part, "Digitization projects in libraries, archives and museums: case studies," introduces examples of work; and the third, "Social networking and digital cultural objects," includes chapters on image, music, and film discoverability. However, these divisions do not organize the content particularly well. As the title of the book indicates, each chapter relates to an aspect of analysis, discovery, and retrieval of digital cultural objects and these concepts would have served as better thematic divisions.

Professionals in the library and cultural heritage sector understand the importance of digital cultural objects and the challenges related to making this content accessible and discoverable. One such challenge is creating access points. Descriptive practices are constrained by a number of factors. Besides the fundamental challenges introduced by operating from a particular worldview, there is the desire to adhere to established vocabularies and metadata structures to ensure system functionality and data interoperability. Authors Rafferty, Jörgensen, and Le Barre and Cordeiro, each in their separate chapters, point out the limitations created as a result of particular worldviews, assumptions, and by conforming to standards. Standards remain important, but these authors, along with Higgins, emphasize consideration of user goals, needs, and potential contributions in digital library design and content description.

Digital preservation is key to the retrieval of cultural digital objects and in recent years has burgeoned in study and practice. It is discussed throughout a few chapters in different ways. Weller points out challenges in preserving social media and web content, which is crucial in advocating these media as new historical sources. Pennock and Day introduce both overarching organizational strategies and initiatives surrounding digital preservation plus digital preservation workflows at the British Library. Prentice addresses particularities of digitization and preservation of audiovisual content, which is not always considered in major discussions of digital cultural objects. Ultimately, institutions must make an organizational commitment to digital preservation, integrating it into system architecture and mainstream workflows in order to ensure the retrieval and use of digital content in years to come.

The wealth of digital content that libraries, archives, museums, and others have made available has opened up new opportunities for collaboration. Interdisciplinary teams undertake new and creative analyses of these resources. Two chapters illustrate examples of computer scientists collaborating with librarians and digital humanities researchers. Dee et al. write about their work to automate metadata creation for artworks based on image characteristics. Orio describes a project to identify similarities in music for the purposes of removing duplicates from a collection and providing interesting points of analysis for music scholars. Both chapters include a technical breakdown of computational processes which some readers may find difficult to understand. However, this did not take away from the general aim of each chapter. Both are fascinating and provide inspiration for other collaborations.

The editors state in their introduction that their objective in creating this book was to "inspire prospective students to develop creative and innovative research projects at Masters' and PhD levels" (xvii), and it accomplishes that goal. Those entering the field today have different concerns and considerations than students in years past. This book provides some foundational information, but mainly presents new ideas and issues such as digital preservation, linked data, user-centered design, and digital humanities.

Chapters are written at varying levels of detail. Many chapters serve to introduce ideas and whet the appetite with the expectation that interested individuals will pursue further study. In fact, this is mentioned in the book's introduction. Authors were asked to provide a "broad-ranging bibliography" (xviii) for their chapter to encourage additional research. Each chapter has an extensive list of references providing the reader with plenty of resources to explore these ideas.

This book also offers a variety of perspectives. It was published in the United Kingdom with simultaneous publication in the United States. A majority of the authors are affiliated with European institutions and so drew upon different digital library and research examples than those generally appearing in American literature. This international perspective along with the theoretical, applied, academic, and administrative points of view represented throughout make this an insightful collection of works.

This book serves a good introduction to current areas of research in the sphere of digital cultural heritage.

Both students and professionals alike will benefit from these works on important issues that face this domain. Although the chapter arrangement makes it somewhat difficult to detect, the themes of analysis, discovery, and retrieval bring this collection together overall. This unique volume containing new analyses and case studies is a valuable contribution to the field's body of literature.—*Anne Washington (awashington@uh.edu), University of Houston, Houston, Texas*

ALCTS
Association for Library Collections & Technical Services
a division of the American Library Association
50 East Huron Street, Chicago, IL 60611 • 312-280-5038
Fax: 312-280-5033 • Toll Free: 1-800-545-2433 • www.ala.org/alcts