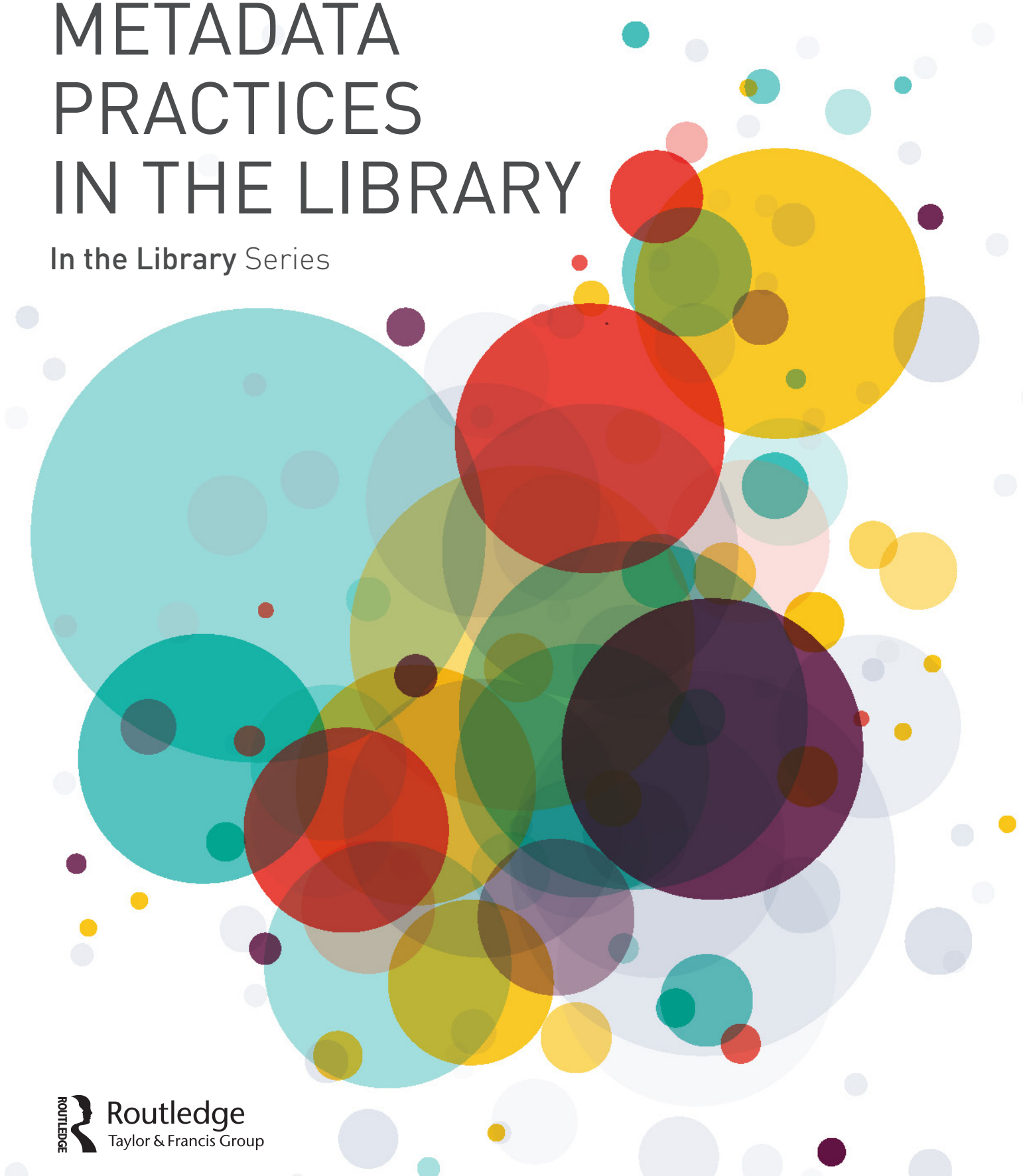
METADATA PRACTICES IN THE LIBRARY

In the Library Series



Routledge
Taylor & Francis Group

TABLE OF CONTENTS

INTRODUCTION

The current technical services operations within libraries needs to be redirected and refocused in terms of format priorities. This FreeBook thus provides library practitioners and students of Library and Information Science (LIS) with a consideration around shifting scholarly publishing, open access, social networking, loss of market share, and declining library funding on collection budgets and digital accessibility to print resources – all of which is in light of Metadata Practice in the Library.

This FreeBook features contributions from experts in their field, including:

Bradford Lee Eden is Associate University Librarian for Technical Services and Scholarly Communication at the University of California, Santa Barbara. He is editor of OCLC Systems & Services: Digital Library Perspectives International and The Bottom Line: Managing Library Finances, has masters and PhD degrees in musicology, and publishes on metadata, technical services, medieval music and liturgy, and J.R.R. Tolkein studies.

Jay Jordan became the fourth president in OCLC's 38-year history in May 1998. He came to OCLC after a 24-year career with Information Handling Services, an international publisher of databases, where he held a series of key positions in top management, including President of IHS Engineering. He is active in professional organizations, including the American Library Association and the Special Libraries Association. He is a Fellow of the Standards Engineering Society.

Jung-ran Park is currently an assistant professor at the College of Information Science and Technology at Drexel University. Her research areas are knowledge organization and representation and computer-mediated communication/online discourse. Dr. Park is currently Editor-in-Chief of Journal of Library Metadata published by Taylor & Francis Group.

Note to readers: As you read through this FreeBook, you will notice that some excerpts reference other chapters in the book – please note that these are references to the original text and not the FreeBook. Footnotes and other references are not included. For a fully referenced version of each text, please see the published title.

READ THE LATEST ON

METADATA PRACTICES

WITH THESE KEY TITLES

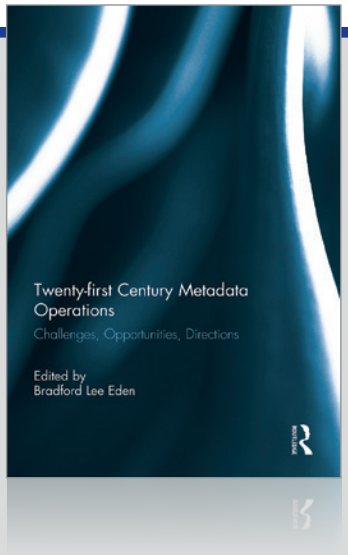
FOR MORE RELEVANT TITLES,
PLEASE VISIT OUR **IN THE LIBRARY** PAGE.

CHAPTER

1

TWENTY-FIRST CENTURY METADATA OPERATIONS

CHALLENGES, OPPORTUNITIES, DIRECTIONS



This chapter is excerpted from
Twenty-First Century Metadata Operations
Edited by Bradford Lee Eden.

LEARN MORE ›

TWENTY-FIRST CENTURY METADATA OPERATIONS

CHALLENGES, OPPORTUNITIES, DIRECTIONS

By Bradford Lee Eden

Excerpted from *Twenty-First Century Metadata Operations*

It has long been apparent to academic library administrators that the current technical services operations within libraries needs to be redirected and refocused, in terms of both format priorities and human resources. A number of developments and directions have made this reorganization imperative:

- While purchased print resources will continue into the future, there will be less of them due to the availability and popularity of online and electronic resources that contain either exact or similar content.
- Every library purchases the same "stuff." It is our special collections, our unique materials that no one else owns and for which there is little if any access either physically or bibliographically, that holds the key to survival for libraries into the future.
- Our current human resources in technical services have focused for too long on purchased print resources as the priority content; libraries need to redirect their scarce resources towards the organization and description of the unique information that each library holds in their special collections and archives, information that is not held anywhere else in the world.
- New directions in libraries, in the areas of metadata, digitization, and digital projects, hold the key to broader collaboration and cooperation in academia with faculty and students, as they struggle with challenges regarding access, curricula, information organization and description, and digital preservation of their created content.
- In the current economic and budget crises, libraries can no longer hire the needed expertise and talent to move forward into these new initiatives, at least not as broadly as they could have five years ago. They must retool and retrain current staff to assist in these initiatives, and make strategic decisions regarding what processes and workflows will no longer be maintained or supported. Technical services staff are uniquely qualified, with their attention to detail and work in metadata standards, to assist libraries as scanning and metadata technicians to digitize and describe objects in the digital environment.
- Our legacy and proprietary integrated library systems (ILSs) cost too much and don't do what we want them to do; open source and Web 2.0 technologies are now advanced enough that, working in consortial and cooperative models, libraries can use combined human resources (especially in the network and programmer areas) to move, manipulate, inventory, purchase, archive/preserve, and provide access to their metadata and digital content in a much more consistent and efficient manner for their patrons, using different cost models and throughputs that are more efficient and cost-effective in the long run, while providing much more user-friendly and interactive search and discovery interfaces.

TWENTY-FIRST CENTURY METADATA OPERATIONS

CHALLENGES, OPPORTUNITIES, DIRECTIONS

By Bradford Lee Eden

Excerpted from *Twenty-First Century Metadata Operations*

- Finally, it is through the retooling, retraining, and re-engineering of technical services staff and their skills from the analogue/print world into the digital world (digitization, digital projects, metadata, etc.) that libraries have a chance to become players in the growing commercialization of accessibility in the information marketplace.

All of this does not take into account the shifting and ever-changing environments surrounding scholarly publishing, open access, social networking, our loss of market share in the information universe, declining state funding of higher education, the effect that the Google book digitization database will have on collection budgets and digital accessibility to print resources, how the Federal Research Public Access Act (FRPAA) will affect libraries' roles in the research and preservation/access process of government grants, etc., etc. Or the fact that libraries need to move into the roles of marketing and outreach.

Overall, there are a number of reports that every librarian should read and digest. The first is No Brief Candle: Reconceiving Research Libraries for the 21st Century (http://www.clir.org/pubs/reports/pub142/pub142.pdf); Anne Kenney's Approaching an Entity Crisis: Reconceiving Research Libraries in a Multi-Institutional Context, which is a response to the previous report (http://www.oclc.org/research/dss/ppt/dss_kenney.pdf); Diane Harley et al., Assessing the Future Landscape of Scholarly Communication: An Exploration of Faculty Values and Needs in Seven Disciplines (http://escholarship.org/uc/cshe_fsc); and the University of Minnesota's Multidimensional Framework for Academic Support (http://www1.lib.umn.edu/about/mellon/docs.phtml). Two recent articles are also worthy of reading and discussion: "Toward a new Alexandria: imagining the future of libraries" The New Republic March 12, 2010 (http://www.tnr.com/article/books-and-arts/toward-new-alexandria) and "Gutenberg 2.0: Harvard's libraries deal with disruptive change" Harvard Magazine May/June 2010 (http://bit.ly/c4m1cy).

One might also want to peruse my contributions to the literature concerning this topic, including "Ending the status quo." American Libraries March 2008 (39:3), p. 38; and "The new user environment: the end of technical services?" Information Technology and Libraries June 2010 (29:2), p. 94-101. I have recently completed chairing the Enterprise-Level Collection Management Services task force as part of the University of California (UC) Libraries' Next Generation Technical Services (NGTS) initiative, charged to develop an operational infrastructure and technical services that can function at an enterprise level (i.e. system-wide) in support of efficient, non-redundant, and collaborative collection services. The charge was:

TWENTY-FIRST CENTURY METADATA OPERATIONS

CHALLENGES, OPPORTUNITIES, DIRECTIONS

By Bradford Lee Eden

Excerpted from *Twenty-First Century Metadata Operations*

...to develop scenarios for enterprise-level collection management services that would support collaborative life-cycle management services for the collective information resources of the UC Libraries. The focus is on acquisition of information resources in all forms and the associated organization of meta-information that enables access by the end user. However, be sure to maintain a broad and holistic perspective that recognizes the role of these services is support of overall collection services including selection, management, archiving, and preservation.

Propose new approaches to technical services processes:

- that support total life-cycle curation for all materials in all UC library collections including special collections and digital materials
- that build upon existing successful system-wide collaborations and that use those successes as models for new collaborations
- that increase access to more materials and that eliminate backlogs and hidden collections
- that provide timely and effective access for the end user that cost less than existing processes

Compare multiple strategies such as:

- decentralized—essentially what we have now but with changes to significantly reduce costs and increase outputs
- centralized—all processing done in a single location
- regionalized—processing done at two locations, one in the north and one in the south
- hybrid—some tasks at a single location, e.g., additional operations similar to the Shared Cataloguing Program

Compare the costs and outputs of each strategy with those for the existing UC technical services operations, including:

- benefits
- obstacles (technical, legal, financial, logistical, service, and HR) cost analysis including savings, transition costs
- impact on end user

TWENTY-FIRST CENTURY METADATA OPERATIONS

CHALLENGES, OPPORTUNITIES, DIRECTIONS

By Bradford Lee Eden

Excerpted from *Twenty-First Century Metadata Operations*

Recommend which strategy or multiple strategies should be implemented and for what reason.

A daunting task, to be sure! More information on the recommendations of this task force, as well as current endeavours and initiatives related to NGTS within the UC Libraries, can be found at http://libraries.universityofcalifornia.edu/about/uls/ngts/docs/ngts_phase2.html, and future budget challenges for the UC Libraries can be found at http://libraries.universityofcalifornia.edu/planning/taskforce/inter-im_report_package_2011-05-00.pdf.

Which brings us to the topic of this book. All of the chapters detail some aspect of technical services reorganization due to downsizing and/or reallocation of human resources, retooling professional and support staff in higher level duties and/or non- MARC metadata, "value-added" metadata opportunities, outsourcing redundant activities, and shifting resources from analogue to digital object organization and description. One chapter specifically discusses the concept of broader cooperative/collaborative sharing of technical services expertise and personnel locally and regionally, while another details a "one person does it all" librarian arrangement that has developed and blossomed at one institution. The first chapter by Mitchell et. al. examines evolving cataloguing roles from a manager's perspective at the University of Houston Libraries. Concepts such as open access, patron-driven acquisitions, batch cataloguing, and locally-curated digital content are discussed, as well as ending the segregation between "cataloguing" and "metadata." The next chapter by El-Sherbini presents a number of models for sharing cataloguing expertise, including the idea of centres of excellence, and the new initiative among OhioLINK libraries called CollaboraTeS. Valentino then details how the University of Oklahoma Libraries integrated digital library metadata creation into the workflow of the Cataloguing Department. John Riemer discusses his philosophy of expanding cataloguing department personnel into the digital arena through his experiences at the University of Georgia and the University of California at Los Angeles (UCLA). A re-visioning process for technical services work- flows at the University of Northern Colorado is detailed by Le er and Newberg in their contribution, followed by an interesting application of the balanced scorecard (BSC) technique for re-engineering the cataloguing department at Hanyang University Library in Seoul, South Korea. Taber and Conger focus on "value-added" cataloguing outside of normal library operations, by developing consultation services and assisting the University of North Carolina at Greensboro with their institutional repository. Cross-training of staff in various services and projects throughout the library at Northern Arizona University is

TWENTY-FIRST CENTURY METADATA OPERATIONS

CHALLENGES, OPPORTUNITIES, DIRECTIONS

By Bradford Lee Eden

described by Pat Headlee et. al. Providing extensive training for library technical services support staff in Enhance and NACO work at Kent State University is described by Lisius et. al., with perspectives from management, expert cataloguer-trainers, and a graduate student. Finally, the merging of technical and public services roles into one librarian position, namely the Cello Music Cataloguer at the University of North Carolina at Greensboro, are detailed by the cur- rent librarian in that position, and how his strengths, talents, and connections assist him in bringing monies and resources into his library.

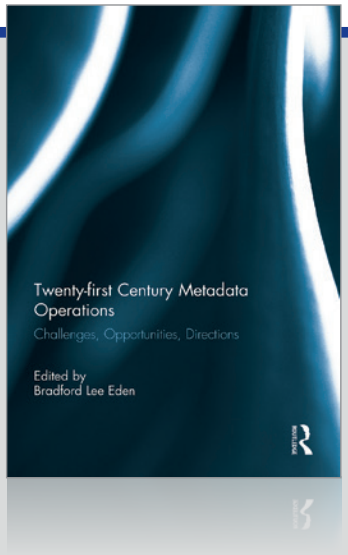
The editor hopes that these contributions to the literature will assist both cataloguers and library administrators with concrete examples of moving technical services operations and personnel from the analogue to the digital environment.

CHAPTER

2

RE-VISIONING TECHNICAL SERVICES

A UNIQUE OPPORTUNITY TO EXAMINE THE PAST, ACCESS THE PRESENT, AND CREATE A BETTER FUTURE



This chapter is excerpted from

Twenty-First Century Metadata Operations

Edited by Bradford Lee Eden.

LEARN MORE ›

RE-VISIONING TECHNICAL SERVICES

A UNIQUE OPPORTUNITY TO EXAMINE THE PAST, ACCESS THE PRESENT, AND CREATE A BETTER FUTURE

By Jennifer J. Leffler and Pamela Newberg

Excerpted from *Twenty-First Century Metadata Operations*

The unexpected resignation of a Technical Services co-manager allowed a unique opportunity for examining and re-visioning all workflows and staffing in a centralized Technical Services department serving the two libraries of the University of Northern Colorado. The past, present, and future of Technical Services, including the growing importance of electronic resources, was researched both within the institution and its peer institutions by a task force. Polling staff and faculty and thinking out of the box helped lead to an organizational model based on timelines rather than materials formats.

INTRODUCTION

The University of Northern Colorado is a doctoral granting university with approximately 13,000 students located in northern Colorado. The students and faculty are served by two libraries: the James A. Michener library and the Skinner Music Library. Technical Services for both libraries are provided by a central department housed in the James A. Michener Library. In April of 2006, a reorganization of the Technical Services Department was realized with the hiring of the Technical Services Co-Manager responsible for Cataloguing. This individual joined the Technical Services Co-Manager of Acquisitions/ Serials to complete the management team for Technical Services. Additional Technical Services personnel included an E-Resources Librarian and twelve classified staff. Previous to this point there were distinct Acquisitions/Serials and Cataloguing departments.

In November of 2007, the Technical Services Co-Manager for Acquisitions/Serials unexpectedly resigned. To ensure interim completion of the tasks performed by this position, responsibility was split between the Technical Services Co-Manager for Cataloguing and the E-Resources Librarian, who was part of the Acquisitions/Serials unit. Management of classified staff was also split between these two librarians. As searches for Acquisitions Librarians had been historically difficult and as both interim managers felt that they could handle the additional duties for an extended period of time, Library Administration decided (in conjunction with both managers) to form a task force to determine the direction that the Technical Services department would take.

LITERATURE REVIEW

The reorganization of Technical Services departments is a not a new topic in library literature. Therefore, literature dedicated to the topic is extensive. An entire chapter of Innovative Redesign and Reorganization of Library Technical Services is dedicated

RE-VISIONING TECHNICAL SERVICES

A UNIQUE OPPORTUNITY TO EXAMINE THE PAST, ACCESS THE PRESENT, AND CREATE A BETTER FUTURE

By Jennifer J. Leffler and Pamela Newberg

Excerpted from *Twenty-First Century Metadata Operations*

to a literature review of the topic.[1] The structures and procedures undertaken by Technical Services departments vary widely. Cataloguing units either as stand-alone entities or a part of a larger Technical Services department are also continuing to undergo massive change. The scope of this change for cataloguers is mentioned in the "Summary of Key Findings" in Emerging Issues in Academic Library Cataloguing & Technical Services.[2] Some articles see the change from cataloguers to "metadata specialists" as all but inevitable.[3] Outsourcing and changes from outside entities, such as the Library of Congress and RDA, have a great impact on the future work of cataloguing departments and greater Technical Services departments and how those departments should be best arranged.[4] Even though reorganization has been a topic of discussion in library literature for decades, the amount of new work before us ensures that the topic will continue to be of interest.

FORMATION OF THE TASK FORCE

The Technical Services Task Force (TSTF), formed by administration after the resignation of the Co-Manager for Acquisitions/Serials, had a unique opportunity to look at Technical Services in depth. In addition, the TSTF could determine a workflow re-arrangement that would better accommodate the ordering, receipt, cataloguing and maintenance of electronic materials in addition to the more traditional tangible items. The Dean of the University Libraries called for task force members both within and outside of Technical Services including faculty and classified staff. In the end, the TSTF was made up of two library administrators, the two interim Technical Services Managers, one cataloguing classified staff member, one acquisitions classified staff member and two serials classified staff members. The first meeting of the TSTF was set for December 21, 2007.

> The University Libraries Administration's charge to the TSTF was: A task force is being established to recommend an organizational structure for Technical Services. The task force will research current trends and roles and how other academic libraries are structured to handle them. The task force will recommend a structure, or alternate structures, designed to meet current needs and be flexible in responding to future needs of both the library and university communities.

At the first meeting of the TSTF, the first task was to brainstorm. The first list created was the guiding principles of the TSTF. These included "Everyone participates," "No

RE-VISIONING TECHNICAL SERVICES

A UNIQUE OPPORTUNITY TO EXAMINE THE PAST, ACCESS THE PRESENT, AND CREATE A BETTER FUTURE

By Jennifer J. Leffler and Pamela Newberg

Excerpted from *Twenty-First Century Metadata Operations*

idea is a bad idea," and "Be civil/tactful/respectful" among others. The second list created was "Adjectives to describe Technical Services in its new form." This list included "efficient," "essential," and "cohesive," among others. In a subsequent meeting the TSTF brainstormed about Technical Services "connections"; who and what Technical Services touches within the library, across campus and off-campus vendors and services. All information generated was placed on a wiki. The use of the wiki facilitated sharing of research done by individuals between meetings as well as real-time recording of meeting notes.

RESEARCH CONDUCTED

The next phase of the TSTF's work was to conduct research. The history of Technical Services within the University Libraries was examined through a review of Charting the Future, a university study published in 2004.[5] Current articles about Technical Services staffing and reorganization were discovered through a literature search and posted on the TSTF wiki. Links to articles on the future of Technical Services and those that focused on the "big picture" were also included. Finally links to cataloguing issues such as On the Record: Report of The Library of Congress Working Group on the Future of Bibliographic Control, Resource Description and Access (RDA), Functional Requirements for Bibliographic Records (FRBR), and next generation catalogues were added to the research page of the TSTF wiki.[6]

The TSTF then looked at the University's peer institutions as determined by NCHEMS. Each committee member chose two peer institutions and gleaned what information they could on the library's Technical Services structure from each library's Web site. In addition, job postings from the previous six months that appeared on AutoCAT, Acq-NET, and ACRL list-serves were gathered. This information gave the TSTF information on how job titles and job duties within Technical Services were currently evolving.

All of the research gathered by the TSTF was summarized in a "Research Results/Findings" document that was forwarded to the Dean of the University Libraries. This document included broad environmental factors impacting academic institutions, issues impacting libraries overall, university- wide issues, University Libraries considerations, and Technical Services considerations.

The TSTF then undertook a survey of the library's Technical Services staff members. The survey consisted of six broad questions:

1. What do we do well?

RE-VISIONING TECHNICAL SERVICES

A UNIQUE OPPORTUNITY TO EXAMINE THE PAST, ACCESS THE PRESENT, AND CREATE A BETTER FUTURE

By Jennifer J. Leffler and Pamela Newberg

Excerpted from *Twenty-First Century Metadata Operations*

2. What could we do better?
3. What are the stumbling blocks you find in your job?
4. What else do we need to do?
5. What can we stop doing?
6. Other comments/concerns?

The answers to the survey questions were gathered electronically by the library administrative assistant and the raw data was formatted into a single document. The same process was then repeated for non–Technical Services faculty and staff with the questions slightly altered:

1. What does Technical Services do well?
2. What could Technical Services do better?
3. What else does Technical Services need to do?
4. What can Technical Services stop doing?
5. Other comments/concerns?

The opinions from both surveys were used in the next steps, which were identifying possible organizational structures and tasks within those structures.

CREATION OF POSSIBLE ORGANIZATIONAL STRUCTURES

To offer as broad of an opportunity for creativity as possible, TSTF members were asked to create at least one organizational chart reflective of the re- search collected and the responses to the surveys. The organizational charts were anonymously submitted to the library administrative assistant who re- produced and collated them for the TSTF. Seventeen organizational charts were submitted by task force members. The organizational charts broke down as follows:

Seven charts using one manager:

librarians—1

librarians—4

librarians—2

Adding a level of hierarchy—2

Everyone reports to the manager—5

RE-VISIONING TECHNICAL SERVICES

A UNIQUE OPPORTUNITY TO EXAMINE THE PAST, ACCESS THE PRESENT, AND CREATE A BETTER FUTURE

By Jennifer J. Leffler and Pamela Newberg

Excerpted from *Twenty-First Century Metadata Operations*

Implemented classified staff "leads"—3

Implemented classified staff pool—1

Implemented two teams—1

Three charts using two departments:

Serials/Electronic (including gov docs) and Cataloguing/Acquisitions—2

Cataloguing/Gov Pubs and Acquisitions/Serials/E-Resources—1

Six charts using three departments:

Acquisitions, Cataloguing, Electronic/Serials—4

Acquisitions/Serials, Cataloguing, eResources—1

Group model—1

One chart using team management:

Acquisitions, Cataloguing, E-Resources/Serials

Four organizational charts were chosen for continued development. These represented a two-department model, a three-department model (**Figure 1**), a team model with a single Head of Technical Services (**Figure 2**) and a three-unit model each with a librarian as unit head and the Acquisitions unit head also acting as Head of Technical Services (**Figure 3**). Tasks or functions were then divided between the entities in each chart. When this process was completed it was found that the two-department model was not feasible based on the amount of work for each librarian/manager and was scrapped.

At this point, one of the members of the TSTF moved completely out of the box and came up with a three-department model that was totally different from the traditional Acquisitions, Cataloguing, and Serials/E-Resources split. The entire spectrum of Technical Services functions were placed on a timeline and then split into three parts. This produced a Resource Procurement Unit that handled all materials, tangible and electronic, from request to purchase through order placement, a Resource Processing and Description Unit that handled all materials from receipt to placement on the shelves (or availability to patrons), and a Resource Maintenance Unit that handled all changes to materials after the initial availability to patrons (**Figure 4**). This new organizational chart and the three organizational charts listed above were then presented to faculty and staff in an open forum.

RE-VISIONING TECHNICAL SERVICES

A UNIQUE OPPORTUNITY TO EXAMINE THE PAST, ACCESS THE PRESENT, AND CREATE A BETTER FUTURE

By Jennifer J. Leffler and Pamela Newberg

Excerpted from *Twenty-First Century Metadata Operations*

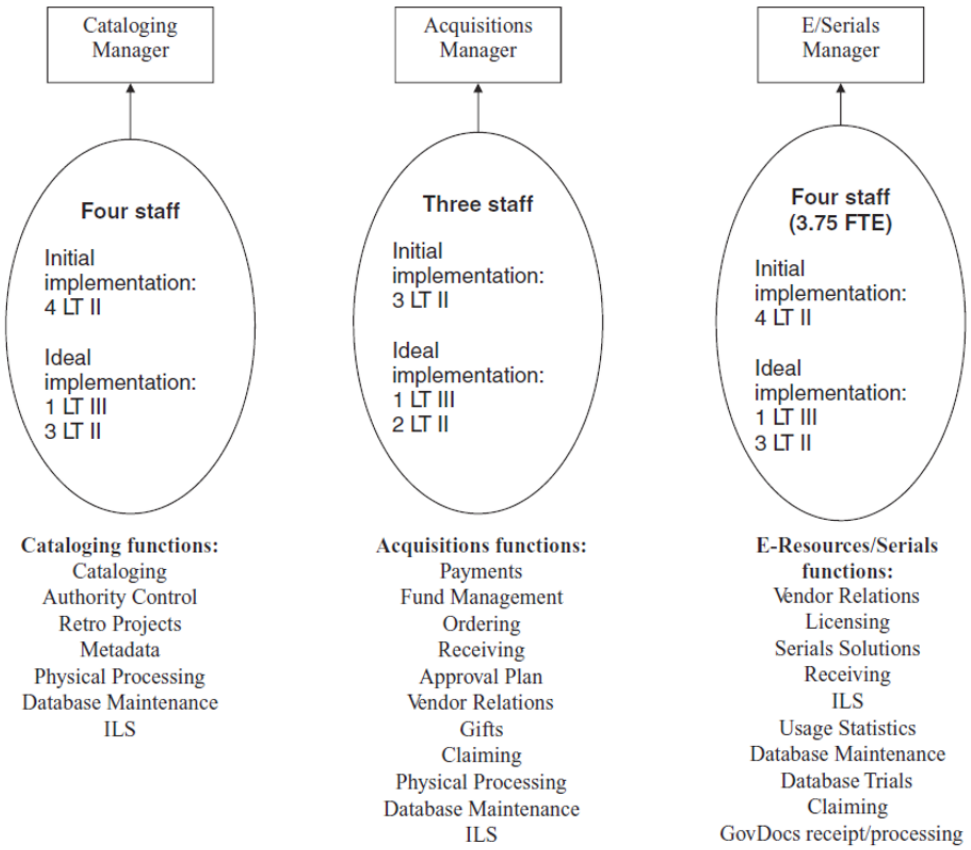


Figure 1 • Three Department Model.

The open forum was held for all libraries personnel. Each of the organizational charts was discussed, including background as to why the task force came up with each model. Questions and comments were accepted and recorded during the meeting. The TSTF asked the library administrative assistant to serve as a contact person for any comments and questions from those staff members who preferred to remain anonymous. The comments received ranged from suggestions about the existing models to entirely new organizational concepts. Some of the suggestions were incorporated, while some (i.e., adding staff members from other areas of the libraries to Technical Services) were deemed unfeasible. From the suggestions made at the open forum two additional organizational charts were created. One new organizational chart incorporated a Head of Technical Services over the three-department model and the other incorporated a three-person management team into the team model.

RE-VISIONING TECHNICAL SERVICES

A UNIQUE OPPORTUNITY TO EXAMINE THE PAST, ACCESS THE PRESENT, AND CREATE A BETTER FUTURE

By Jennifer J. Leffler and Pamela Newberg

Excerpted from *Twenty-First Century Metadata Operations*

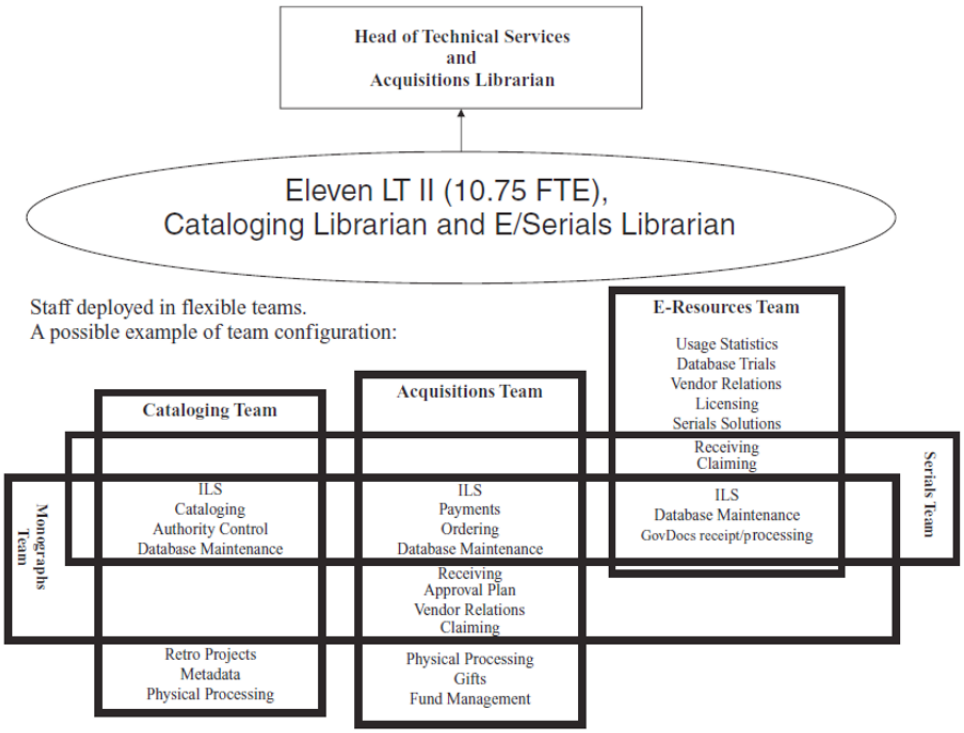


Figure 2 • Team Model.

CREATION OF FINAL PROPOSAL

The next undertaking of the task force was the compilation of a final re- port, with the intended recipient the Dean of University Libraries. The idea was that from this report the Dean would make a final decision about how to reorganize Technical Services. All meeting notes and supporting documentation were reviewed by all members of the task force. An outline was developed by the entire group, and then sections of the report were assigned to individuals to write. It was decided that the organizational charts would be included in the text of the document with explanations, instead of having them stand by themselves or to serve as appendices. The task force could not come to a consensus about which organizational structure would be the best for the libraries. Therefore, pros and cons of each were offered to the Dean in the report. The chair of the task force compiled all the sections into a cohesive document and submitted the report to the Dean.[7] Upon receipt of the final report, the Dean made it available electronically to all libraries personnel. Along with the report, the Dean gave

RE-VISIONING TECHNICAL SERVICES

A UNIQUE OPPORTUNITY TO EXAMINE THE PAST, ACCESS THE PRESENT, AND CREATE A BETTER FUTURE

By Jennifer J. Leffler and Pamela Newberg

a timeline for his decision and said that he would meet with several constituencies before a decision was reached. He also solicited feedback on the report.

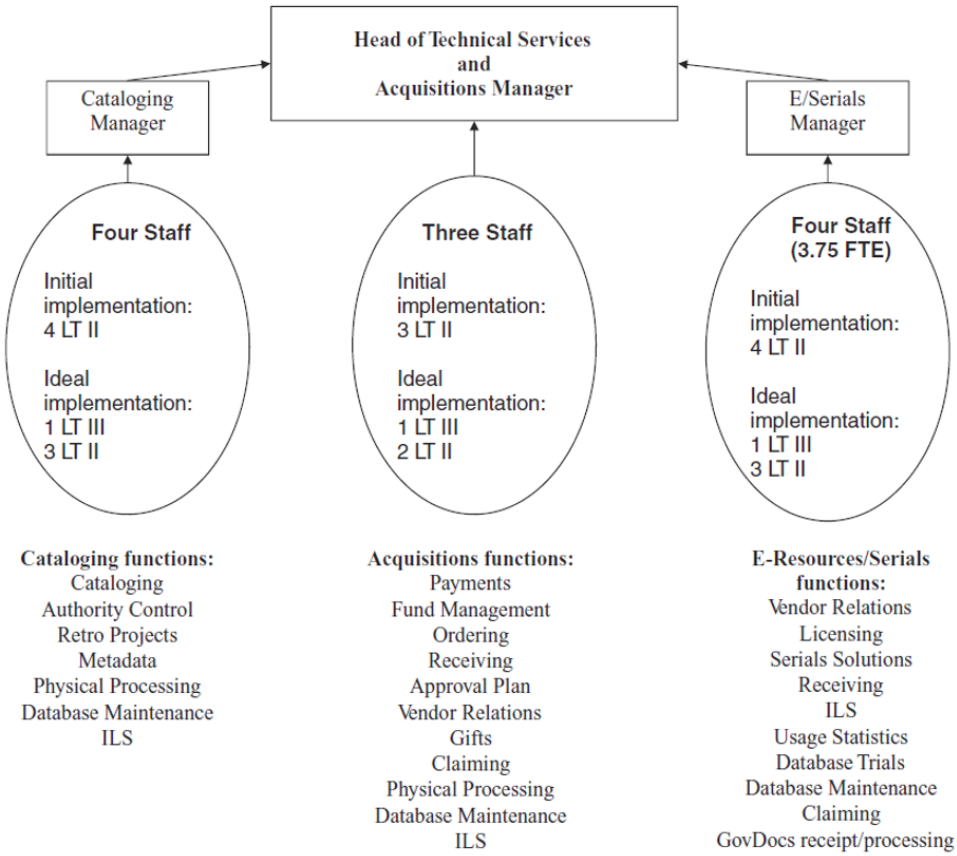


Figure 3 • Three Unit Model.

The Dean met with the other two administrators, both of whom served on the task force, to discuss the final report. The managers of departments outside of Technical Services were queried for feedback. The Dean also met with both existing Technical Services librarians to discuss the report and the task force as a whole. Based on feedback, the Dean decided that the new organizational structure for Technical Services would be the three-unit model, Acquisitions, Cataloguing, and E-Resources/ Serials, each with a librarian as unit head. In addition, the Acquisitions unit head would act as Head of Technical Services. The librarian managing Cataloguing would re- port to the Technical Services Manager as would the Manager of E-resources/ Serials.

RE-VISIONING TECHNICAL SERVICES

A UNIQUE OPPORTUNITY TO EXAMINE THE PAST, ACCESS THE PRESENT, AND CREATE A BETTER FUTURE

By Jennifer J. Leffler and Pamela Newberg

Excerpted from *Twenty-First Century Metadata Operations*

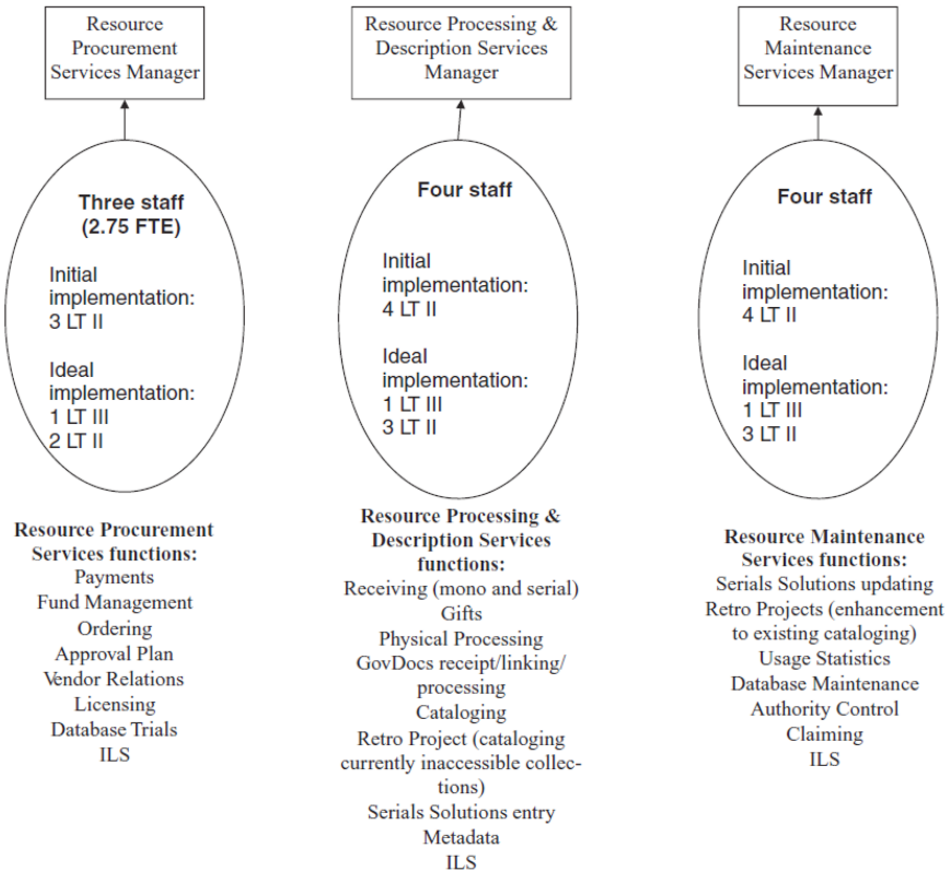


Figure 4 • Timeline Based Model.

When the Dean presented this decision to the two acting Technical Services Co-Managers, they requested time to review his decision. After discussion, the acting managers proposed a slightly different organizational structure. The managers requested that rather than using the traditional Acquisitions, Cataloguing, E/Serials spilt within the chosen structure that Resource Procurement, Resource Processing and Description, and Resource Maintenance units would be created (Figure 5). It was felt that this structure would better adapt to changing technologies in Technical Services as well as provide the opportunity for an entire re-organization of Technical Ser- vices that could serve to solve issues that were raised in survey responses. Also, since every position of Technical Services would change in some way, it was felt that everyone involved would develop a sense of buy-in to the re-organization. After careful consideration and more discussion, the Dean accepted the revised

RE-VISIONING TECHNICAL SERVICES

A UNIQUE OPPORTUNITY TO EXAMINE THE PAST, ACCESS THE PRESENT, AND CREATE A BETTER FUTURE

By Jennifer J. Leffler and Pamela Newberg

organizational structure; the new vision focusing on timeline rather than material format allows the Technical Services department flexibility in dealing with new issues and technologies as they arise.

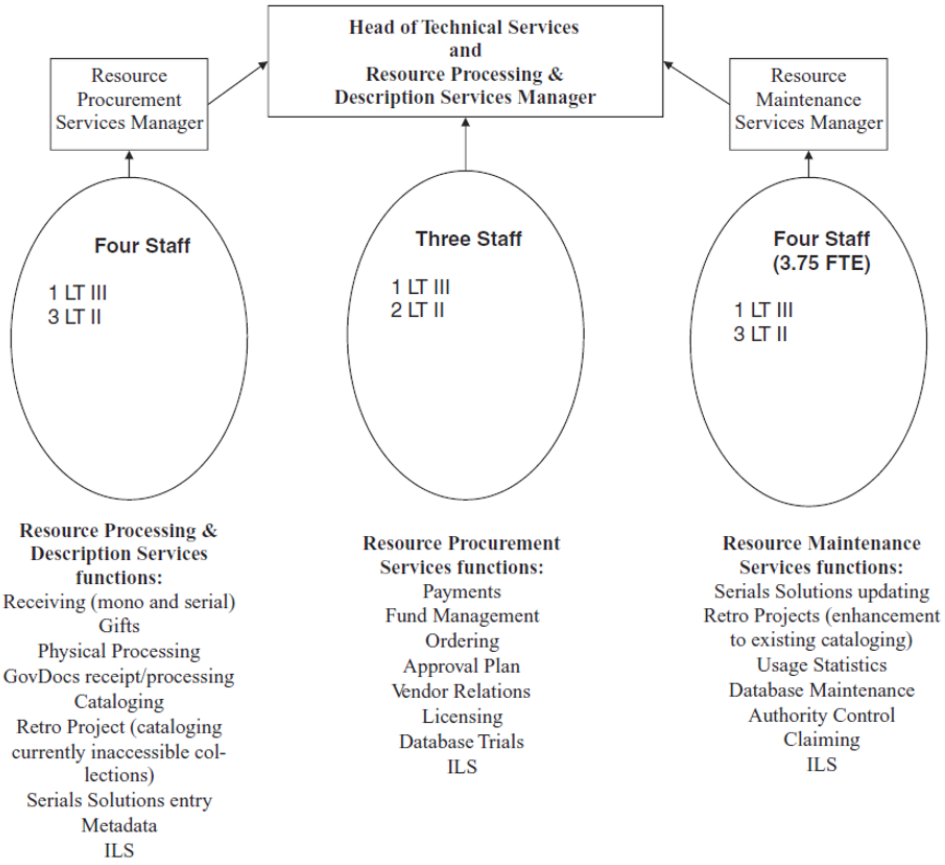


Figure 5 – Final Model.

The current managers were placed in the positions of Manager of Resource Processing and Description and Manager of Resource Maintenance as they requested. Once this decision was reached, the Dean shared the selected organizational structure with the members of the Technical Services department. Another open forum was then offered to the other members of the libraries to disseminate the decision and outline the next steps.

Since the Head of Technical Services and Manager of Resource Procurement was a new position for University Libraries, the position description was crafted by the

RE-VISIONING TECHNICAL SERVICES

A UNIQUE OPPORTUNITY TO EXAMINE THE PAST, ACCESS THE PRESENT, AND CREATE A BETTER FUTURE

By Jennifer J. Leffler and Pamela Newberg

administrators with some input from members of Technical Services. When the search committee was named for this position, it was ensured that no members of the original task force were included.

CONCLUSION

The University Libraries is now eighteen months into the re-organization of the Technical Services Department. A new Head of Technical Ser- vices/Manager of Resource Procurement has been at the University for nearly a year. Changes are still being made and re-organization is still in progress, but that is another story.

The University Libraries at the University of Northern Colorado had a unique opportunity to spend time researching the present and future of Technical Services. There was time to gather input from all members of the University Libraries and determine how change would affect each person and workflow. At times, it seemed like there would be no end to uncertainty. In the end, an organizational structure was created that maintains flexibility, efficiency, and cohesiveness.

NOTES

1. Laurie Lopatin, "Review of the Literature: Technical Services Redesign and Reorganization," in Innovative Redesign and Reorganization of Library Technical Services, ed. Bradford Lee Eden (Westport, Conn.: Libraries Unlimited, 2004).
2. Elaine Sanchez, Emerging Issues in Academic Library Cataloguing & Technical Services (New York: Primary Research Group, 2007).
3. Martha Hruska, "Where are We with the Staffing Transition from Cataloguing to Metadata Management?" Technicalities 29 (2009).
4. Vera Fessler, "The Future of Technical Services (It's Not the Technical Services It Was)," Library Administration & Management 21 (2007) and Michael Boock, "Changes in Responsibilities, Organization and Staffing within Technical Services Departments. A Report of the ALCTS Technical Services Administrators of Medium-Sized Research Libraries Discussion Group. American Library Association Mid-Winter Meeting, Seattle, January 2007," Technical Services Quarterly 25 (2008): 71–75.
5. Charting the Future: Final Report. http://www.unco.edu/pres/charting/CTFFinal Plan.pdf

RE-VISIONING TECHNICAL SERVICES

A UNIQUE OPPORTUNITY TO EXAMINE THE PAST, ACCESS THE PRESENT, AND CREATE A BETTER FUTURE

By Jennifer J. Leffler and Pamela Newberg

Excerpted from *Twenty-First Century Metadata Operations*

6. On the Record: Report of the Library of Congress Working Group on the Future of Bibliographic Control. http://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf
7. To request a copy of the final report, please contact the authors.

CHAPTER

3

21ST CENTURY LIBRARY SYSTEMS

This chapter is excerpted from

Weaving Libraries into the Web

Edited by Jay Jordan

LEARN MORE ›

21ST CENTURY LIBRARY SYSTEMS

By Andrew Pace

Excerpted from *Weaving Libraries into the Web*

ABSTRACT

Less than a decade into the 21st century, perhaps it is more fitting to describe library automation as approaching its 80th birthday, is a time to look back and carefully measure moving forward. Since the introduction of a punch card circulation system at the University of Texas in 1936, through the advent and perseverance of the MARC record, and following the ebb and flow of nearly 75 different library automation vendors, library automation has come a long way. For some, however, it has not come nearly far enough. If one were to stop the history of library automation in the mid-1990s and wish away the dominance of the Internet, libraries and patrons might have been quite content with the state-of–the- art as it existed 15 years ago. But wishing away the Internet is like envisioning a world without electricity and indoor plumbing; as such, that 1990s library automation summit is now a plateau from which many library technologists and futurists can see no launch pad to a next-generation of library software and services.

> "If you wish to make an apple pie truly from scratch, you must first invent the universe."—Carl Sagan

A SLOW START

The irony of the current stagnant situation for library systems is that libraries likely offered the public its first glimpse of computer use and database interaction. Long before ATM machines and the Web, many of the first public keyboards could be found attached to dumb terminals in libraries. These terminals were, in turn, connected to mainframes, and libraries supported workflows that either relied on data supplied from a central hub, or created stand-alone systems for local inventory control.

Those local inventory systems, built upon ordering, acquisition, and circulation of physical materials grew into the robustly functional integrated library systems (ILS) with which most libraries are now familiar. Because back office workflows were governed by electronic records and computerized inventory, libraries were able to leap forward in providing public access to those records. The displays seem quaint by today's standards, but were designed to transition patrons from card catalogues to their new electronic equivalent.

Unfortunately, libraries and their vendors were not prepared for the exponentially rising expectations that the advent of the Web would usher in. Mired in transitioning character-based telnet systems to rapidly selling graphical user interface (GUI)

21ST CENTURY LIBRARY SYSTEMS

By Andrew Pace

Excerpted from *Weaving Libraries into the Web*

systems, most vendors were ill-prepared to make another transition to the Web just a few short years later. First-generation Web-based online catalogues reflected the nascent state of Web development and lacked much of the functionality that had been available in online systems for over a decade. Faced with few alternatives, libraries suffered the pain of first generation GUI systems and took a wait-and-see approach to more sophisticated patron interfaces. Unfortunately, this strategy resulted in a wait-and-wait scenario for both end-user experience and back-office operations.

PLUGGING THE GAPS

While libraries seemingly accepted the fate that the basic functions provided by an integrated library system would not change radically, the nature of their collections and associated workflow were themselves changing rapidly. Web-based content, licensed resources, born-digital documents, and institutionally significant digital collections emerged rapidly to overtake the effort required to maintain print collections, especially in academic libraries. Traditional integrated systems proved inadequate for managing these assets despite numerous noble efforts to fit square pegs into round holes—eSerials checkin, Cooperative Online Resource Cataloguing (CORC), e-reserves scanning stations, etc.

The inadequacy of the ILS was compounded by a desire among vendors and libraries alike to build new solutions with new technologies. Electronic Resource Management (ERM), Digital Asset Management (DAM), and Institutional Repository (IR) systems would be built with 21st century technologies to aid in these new library workflows. Paradoxically, as industry expert Marhsall Breeding points out, "[The process of evaluating library workflow] may be confounded by the fact that many libraries have adapted their work- flows to match the limitations of their automation systems" (Breeding, 2007). This begs the question whether vendors have done a short-term service to libraries in the midst of a major sea-change, while doing a longer-term disservice to the efficiency of libraries.

Certainly, if automation experts were starting from scratch, they would endeavour to logically combine resource management in libraries under an umbrella of software that makes distinctions between resource format with- out unnecessarily bifurcating workflow into separate systems. A current list of essential products, of course, makes this challenge more daunting than it might seem at first glance. Many libraries might delineate a suite of services (in addition to the ILS) similar to the list provided by Mark Andrews (Andrews, 2007):

21ST CENTURY LIBRARY SYSTEMS

By Andrew Pace

Excerpted from *Weaving Libraries into the Web*

- OpenURL Link Resolver Federated search tool
- Digital archive, institutional repository, and portfolio products Electronic Resource Management (ERM)
- Compact and robotic storage systems for archived print materials
- Next-generation portal and discovery tools (for all of the above)
- A management interface (for all of the above) to determine usage and user satisfaction and allow for ad hoc reporting and statistical analysis

It's difficult to picture a library workflow, let alone a single integrated product that can handle so much. Nevertheless, there are some technical strategies, discussed below, that might make the tactical deployment of solutions adequately functional, faster to deploy and upgrade, and less expensive for libraries.

BUSINESS DISTRACTIONS

Before the demand for products capable of managing a new myriad of library content, vendors sought merely the state-of-the-art for managing print col- lections. "The hallmark of [first generation library] systems," writes Andrews, "was the struggle for 'functional completion' in an 'integrated library system'" (Andrews, 2007). By the late 1990s, the library software business had created several commodity-like applications. One vendor's offerings had become less and less distinguished from another, leading one pundit to liken the choice between ILSes to a choice between cars on a rental lot (Pace, 2004). Nevertheless, this plateau of innovation had yet to cause considerable churn within the market. Concomitant with the market saturation for integrated systems was the firm establishment of strong and loyal relationships between libraries and their vendors. In fact, an apparent paucity of new product penetration made many vendors appear less like software companies and more like relationship management companies.

Customer relations and management would get a lot trickier in the early part of the 21st century. As indicated in **Table 1**, 2000–2008 activities in the library automation space have been largely driven by mergers and acquisitions, with over 30 major activities in less than 10 years. It's no wonder that a combination of business consolidation, stunted innovation, and rapid Web application development outside the library automation space would lead to disenchantment and restlessness among libraries.

21ST CENTURY LIBRARY SYSTEMS

By Andrew Pace

Excerpted from *Weaving Libraries into the Web*

TLC acquires CARL (2000)
Auto-Graphics acquires Maxcess Library Systems (2001)
Sirsi acquires DRA (2001)
Scott Cheatham acquires EOS (2001)
Jerry Kline acquires remaining shares of Innovative Interfaces (2001)
OCLC acquires netLibrary (2001)
Geac acquires Extensity (2002)
ProQuest acquires Serials Solutions (2004)
ISACSOFT acquires Bibliomondo (2004)
Bowker acquires Syndetic Solutions (2004)
Sirsi acquires Docutek (2005)
Sirsi acquires Dynix (2005)
OCLC PICA acquires Fretwell-Downing (2005)
OCLC PICA acquires Sisis (2005)
Golden Gate Capital acquires Geac (2005)
OCLC acquires Openly Informatics (2006)
Follet acquires Sagebrush (2006)
Geac becomes Extensity (2006)
Fransisco Partners acquires Ex Libris (2006)
Infor acquires Extensity (Geac) (2006)
Fransisco Partners acquires Endeavor Information Systems (2006)
Cambridge Information Group acquires Proquest (2006)
OCLC acquires DiMeMa (CONTENTdm) (2006)
Vista Partners buys out SirsiDynix (2006)
Bowker acquires MediaLab (AquaBrowser) (2007)
Liblime acquires Katipo's Koha division (2007)
OCLC acquires remaining shares of OCLC Pica (2007)
Ronald Brisebois acquires ISACSoft (2008)
LibLime acquires Care Affiliates (2008)
Leeds Equity acquires Ex Libris (2008)

Table 1 • Mergers & Acquisitions, 2000-2008

TURNING TIDES

It's also no coincidence that the first half of this decade in which blogs became so prevalent was marked more by a clamouring and complaining about the state of library automation than by the actual development of innovative software. Twenty-first century library system development is now driven by restless customers, motivated not only by a few tireless advocates, but also by the publicly visible fruits of system development within libraries.

Open Source Software (OSS) efforts such as the Open Archive Initiative (OAI), DSpace, and Koha—just to name a few, as an exhaustive list would overwhelm the reader—challenged commercial proprietary systems, not only for market share but often in terms of sophistication and functionality. Experimentation with new so-called bolt-on catalogue interfaces such as RLG's RedLightGreen and Casey Bisson's blog-powered WPOPAC led to production efforts from several individual libraries and vendors, including, North Carolina State University Libraries, OCLC, and AquaBrowser (Antelman, Lynema, & Pace, 2006).

21ST CENTURY LIBRARY SYSTEMS

By Andrew Pace

Excerpted from *Weaving Libraries into the Web*

Challenged by relative new-comers and outsiders of the library automation space—Endeca, MediaLab, WordPress, and FAST—vendors adroitly answered the call for improved public interfaces. In fact, it is fair for vendors to decry at least some of the impatient clamouring of library IT specialists, as many of the increasingly expensive incremental changes made to legacy ILS systems were demanded by the libraries paying relatively small maintenance fees. One might argue that vendors were squandering the money of their customers doing exactly what was asked of them.

NEXT GENERATION AS A ZERO-SUM GAME

Despite the nimble reaction of many ILS vendors to fill some of the service gaps created by the inadequacy of the ILS to meet 21st century needs, the overall market for integrated library systems has not grown substantially over the last 5 years. With annual revenues estimated at $570 million, sales of new ILSes dipped 15% in 2008. These losses were partly offset by new end-user product offerings, but do little to indicate incentives to radically change or improve underlying systems.

> Several factors limited opportunities to sell traditional library automation systems this year. The higher-end market of public and academic libraries has saturated; fewer libraries have legacy systems in immediate need of replacement. Recent migrations from legacy systems have largely run to completion ... [L]ibraries considering ILS replacements are holding off, hoping better options will emerge soon, especially on the open source front. Libraries feel a sense of urgency to acquire next-generation interfaces that will allow them to cast aside library catalogues that work more like the Web of 1998 than 2008 and gain tools to manage ever- growing collections of electronic content (Breeding, 2008)

It's clear that to counter the impact of a zero-sum future for the ILS, the next generation of functional offerings must be technically compelling while providing all the functionality with which libraries are accustomed.

OPTIMISTIC FORECASTS

Two of the last three endeavours to create an ILS from scratch in the last decade have been business, if not also functional, failures. DRA's Taos system was killed after the company's acquisition by Sirsi, and Dynix's Horizon 8.0 was declared dead-

21ST CENTURY LIBRARY SYSTEMS

By Andrew Pace

Excerpted from *Weaving Libraries into the Web*

on-almost-arrival after a merger with the same company. While some might tie these failed attempts at a next-generation management system to a common corporate ownership, some might have predicted lack-lustre outcomes of the somewhat overly optimistic picture created by the newly architected systems.

A more optimistic spin might say that the second mouse gets the cheese. The third (and thus far successful) venture alluded to above is the open source ILS venture, Evergreen, now supported by Equinox, Inc. By releasing their software as open source, the Evergreen team created a new compelling reason to consider switching systems. Though it combines the functionality sought after in a new patron front-end, the system actually falls short on the full functionality of other proprietary ILS systems. Nevertheless, it is the positioning of the open source code as something new, and embraced by forward-thinking customers, that has lured customers away from more traditional solutions.

Fortunately for libraries, the freshness of the open source solution is not the only 21st century innovation to look forward to; nor is it mutually exclusive of another burgeoning trend that is likely to have an impact on a next generation of service offerings.

THE CLOUD GENERATION

Neil Howe and William Strauss are experts in evaluating the trends of generations. They write, "to anticipate what 40-year-olds will be like 20 years from now, don't look at today's 40-year-olds, look at today's 20-year olds" (Howe & Stauss, 2007). It is worthwhile, therefore, to evaluate the platforms on which younger generations are computing. This is not to suggest that Facebook, Flickr, and Wikipedia will form the basis for a next-generation library management system. It is these very services, however, that should serve as a model for 21st century data storage, software on demand, and cloud computing capabilities.

> The cloud is a metaphor for the Internet (based on how it is depicted in computer network diagrams) and is an abstraction for the complex infrastructure it conceals. It is a style of computing where IT-related capabilities are provided "as a service," allowing users to access technology-enabled services from the Internet ("in the cloud") without knowledge of, expertise with, or control over the technology infrastructure that supports them (Wikipedia, 2008).

The Gartner Group predicts that massively scalable service solutions provided by

21ST CENTURY LIBRARY SYSTEMS

By Andrew Pace

Excerpted from *Weaving Libraries into the Web*

cloud computing will be as influential as E-business (Gartner, 2008). Fast-paced improvement to IT infrastructure and the continued industrialization of IT services over the last decade has laid the ground- work for Web-based software services. Popular examples include Google- Docs, QuickenWeb, or Salesforce.com. According to Daryl Plummer, Man- aging Vice President and Gartner Fellow, "this is due, in part to the commoditization and standardization of technologies, in part to virtualization and the rise of service-oriented software architectures, and most importantly, to the dramatic growth in popularity of the Internet" (Gartner, 2008).

	18–29	*30–49*	*50–64*	*65+*
Use webmail services such as Hotmail, Gmail, or Yahoo! mail	77%	58%	44%	27%
Store personal photos	50	34	26	19
Use online applications such as Google Documents or Adobe Photoshop Express	39	28	25	19
Store personal videos	14	6	5	2
Pay to store computer files online	9	4	5	3
Back up hard drive to an online site	7	5	5	4
Have done at least one activity	87%	71%	59%	46%
Have done at least two activities	59	39	31	21

Source: Pew Internet & American Life Project April-May 2008 Survey. N = *1,553 Internet users. Margin of error is* ± *3%.*

Table 2 • Cloud Computing Activities by Different Age Cohorts. Internet users in each age group who do the following onine activities (%).

If one accepts the premise that the ILS has reached commodity status, it stands to reason that the services provided by locally installed and maintained software can and should be provided by a networked service. Of course, a higher level of trust and reliability must be achieved, and it remains to be seen whether existing vendors can put the same trust and reliability into software services that many online publishers have established with online scholarly and popular content.

Nevertheless, if this generation's 20-year-olds are the next generation's library administrators, it might be worth taking a look at the increased level of trust placed in cloud computing and data storage by younger generations. A look at usage levels according to age groups shows rising levels of trust for storing personal data on the Internet (Horrigan, 2008).

While the Pew study does not specifically address business data storage, it is easy to make extrapolations about the level of trust in those areas, and several online

21ST CENTURY LIBRARY SYSTEMS

By Andrew Pace

Excerpted from *Weaving Libraries into the Web*

businesses are banking on the future for cloud computing that Gartner, Pew, and others have predicted.

WEB AS PLATFORM

One such company banking on software-as-a-service (SaaS) and cloud computing is Bungee Labs, creators of Bungee Connect, an end-to-end environment that allows developers to build desktop-like applications from multiple Web services and databases and then instantly deploy them on Bungee's multi-tenant grid infrastructure. Services of this type are either extensions of or have been emulated by much more recognizable companies like Amazon and Google.

If such platforms—Bungee's Dave Mitchell goes so far as to call the model Platform-as-a-Service (PaaS)—were extended to library software us- age, libraries might foresee a day when large capital expenditures for hard- ware and software could be replaced by subscription-based services. Mitchell writes:

On the SaaS side of things, there have been some notable successes in the areas of [Customer Relationship Management] CRM-as-a-service, computing-as-a-service and storage-as-a-service. These are just a few examples of data, functionality and hardware as services over the network. These individual offerings represent the next logical evolution of software and computing in the cloud (Mitchell, 2008)

Technical Advantages of the PaaS Model

- Develop, test, deploy, host, and maintain on the same integrated environment
- Dramatically reduce costs of development while supporting a robust software life cycle.
- User experience without compromise: avoiding downloads, plugins, and Internet hiccups
- Built-in scalability, reliability, and security
- Multi-tenanc—the ability for an application to automatically partition state and data to service an arbitrary number of users
- Must support Web-scale use
- Built-in integration with Web services and databases
- Deep application instrumentation—see exactly how and when users are using the application (Mitchell, 2008)

21ST CENTURY LIBRARY SYSTEMS

By Andrew Pace

Excerpted from *Weaving Libraries into the Web*

It's at least time that libraries and vendors turned some of their attention from richer end-user experiences to the back-office workflows that support them. As Breeding contends, "We can't let the current focus on front-end interfaces make us complacent about the software systems that we use to automate routine library functions" (Breeding, 2007). The timing seems right to make such an effort at the creation of next-generation systems with the cloud in mind. There could come a day very soon that libraries would simply plug into the wall to receive all the required power of software services, rather than running locally deployed systems like home generators with all the associated expense, cyclical upgrades, and hardware maintenance.

The economic advantages to a service-based future for library automation should not be under-estimated. Despite a surge of online content being available to patrons, libraries will continue back-office operations for all types of materials. The more these workflows are industrialized and served by network-level applications, the more time and effort libraries can assign to other intellectual endeavours. Far too much time is spent getting systems to work at the expense of more fruitful activity.

In varied lists of technical demands made of library automation vendors, the library is poised to become part of the Web 2.0 culture, acknowledging and even supporting many Web service models. Most punditry, however, still calls for hardware independence and access to proprietary APIs; demands fall short by merely asking that local systems avail themselves of other Web services rather than establishing themselves as services in their own right. Moreover, integration with other business process systems—course management, financial services, and human resource systems—will require new thinking on a next-generation of integration. Acknowledgement that library management system will never attain dominance as college, university, com- munity, and corporate business process systems should encourage libraries to seek integration through Web-based services—a loftier goal than mere "interoperability"—so that library workflows can be managed in conjunction with other services.

THE FUTURE IS INEVITABLE

When it comes to library automation, lamenting the past is nearly as easy as predicting the future is difficult. One thing seems fairly certain, however—that the library automation landscape requires dramatic change in order to ensure its future. The landscape metaphor itself is too pessimistic, though, as shifting ground often leaves only destruction as its aftermath. Libraries require a sea-change—a dramatic

21ST CENTURY LIBRARY SYSTEMS

By Andrew Pace

departure from the status quo of library automation, solutions that will scale like typical Web solutions, technologies that will ensure our future. To date, the swelling seas of library automation have been caused by the rising tide of discontent in libraries. Going into the future, libraries, service providers, and technology experts have an unparalleled opportunity to create the swelling seas on which all boats will rise.

REFERENCES

- Andrews, M. (2007). OPAC changing markets, changing relationships: how libraries and vendors respond to the 'next generation' challenge. *Library Hi Tech*, 25, 562–578.
- Antelman, K., Lynema, E., & Pace, A. (2006). Toward a twenty-first century library catalogue. *Information Technology and Libraries*, September, 128–39.
- Bahr, E. (2007). Dreaming of a better ILS. *Computers in Libraries*, 27, 10–14.
- Breeding, M. (2007). It's time to break the mold of the original ILS. *Computers in Libraries*, 39–41.
- Breeding, M. (2008). Automation systems marketplace 2008: Opportunity out of turmoil. Library Journal. Retrieved September 30, 2008, from http://www.libraryjournal.com/article/CA6542440.html
- Fons, T., & Jewell, T. (2007). Envisioning the future of ERM systems. *The Serials Librarian*, 52, 151–166.
- Gartner Says Cloud Computing Will Be As Influential As E-business. (June 26, 2008). Press release. Retrieved September 30, 2008, from http://www.gartner.com/it/page.jsp?id707508.
- Henderson, M. (2008). Heresy and misconduct: evolution of library automation. *ILA Reporter*, 26, 4–7.
- Horrigan, J. (2008). Use of cloud computing applications and services. *Pew Internet and American Life Project*, 1–9.
- Howe, N., & Strauss, W. (2007). The Next 20 years: How customer and workforce attitudes will evolve. *Harvard Business Review*, July-August, 41–52.
- Mitchell, D. (2008). Defining platform-as-a-service, or PaaS. Bungee Connect Developer Network. Retrieved September 30, 2008, from http://blogs.bungeeconnect.com/2008/02/18/defining-platform-as-a-service-or-paas/
- Pace, A. (2004). Dismantling integrated library systems. *Library Journal, February 1, 2004*. Retrieved September 30, 2008, from http://libraryjournal.com/article/CA374953.html

21ST CENTURY LIBRARY SYSTEMS

By Andrew Pace

- Wikipedia. Retrieved September 30, 2008, from https://en.wikipedia.org/wiki/Cloud_computing.
- Wolven, R. (2008). In search of a new model. *netConnect*, January 2008. Retrieved September 30, 2008, from http://www.libraryjournal.com/index.asp?layout=articlePrint&articleID=CA6514921

CHAPTER

4

NEXT GENERATION CATALOGUING

This chapter is excerpted from

Weaving Libraries into the Web

Edited by Jay Jordan

LEARN MORE ›

NEW GENERATION CATALOGUING

By Karen Calhoun and Renee Register

Excerpted from *Weaving Libraries into the Web*

ABSTRACT

During the past several years the Online Computer Library Centre (OCLC) noted increasing concern from the library community regarding the future of cataloguing. In response to this concern and to the rapidly changing metadata environment, OCLC is taking steps to further a paradigm shift toward early acquisition of metadata in WorldCat directly from the entities responsible for content purchased by libraries—chiefly the publisher supply chain. This shift in thinking requires the acknowledgement that metadata is dynamic and will change over time and relies upon the automated capture of metadata early in the publishing cycle as well as automated processes to help make the early metadata "good enough."

The explosion of content, the expectation of rapid metadata exposure in the Web environment, user expectation of "get it (or reserve it) now," economic concerns regarding the cost of metadata creation, and the decrease in cataloguing staff make the development of more efficient, cost-effective methods for metadata creation and maintenance imperative.

LC WORKING GROUP ON THE FUTURE OF BIBLIOGRAPHIC CONTROL

In November 2006, Deanna Marcum, Associate Librarian for Library Services at the Library of Congress, convened a "Working Group on the Future of Bibliographic Control" to explore the current metadata environment and make recommendations on the future of bibliographic control in our evolving environment. The Working Group published its final report, titled "On the Record", in January 2008 (Library of Congress, 2008). One of the first recommendations contained in "On the Record" is to "Make more use of bibliographic data earlier in the supply chain." This recommendation was welcomed by OCLC since a pilot program called "Next Generation Cataloguing and Metadata Services" had been in the planning stages during 2007 and was kicked off in January 2008. This pilot was designed to explore greater use of publisher supply chain metadata in record creation for libraries and to explore the use of library data to enrich publisher supply chain metadata.

The publisher supply chain responded positively to the OCLC pilot concept. The importance of Web-based discovery and buying tools has resulted in a strong movement in the publishing industry toward exchange of metadata in electronic format, the development of standards to support metadata exchange and growing consensus on the importance of consistent application of best practices in metadata creation. The increasing evolution and adoption of the ONIX format

NEW GENERATION CATALOGUING

By Karen Calhoun and Renee Register

Excerpted from *Weaving Libraries into the Web*

(http://www.editeur.org/onix.html) for the capture and electronic exchange of title information in the marketplace provides an opportunity for the library community to break down traditional silos between library and publisher supply chain metadata.

However, the publisher supply chain also struggles with the explosion of content and the cost of metadata creation and maintenance. Publishers and vendors who serve the library market as well as the retail market must also contend with metadata needs specific to the library community such as library-defined classification schema and terminologies, need for metadata in MARC format, and rules for description that differ from standard practice in the non-library metadata arena.

NEXT GENERATION CATALOGING PILOT

OCLC's Next Generation Cataloguing pilot is designed to increase the level of interoperability between publisher supply chain and library data and add value to metadata for both through leveraging the strengths of each community. The organization of knowledge is a core competency for libraries. We excel in metadata consistency, the creation and application of classification schema and terminologies, the establishment and application of authority controlled fields. The publishing community often struggles with these aspects of metadata creation and maintenance resulting in users' failure to retrieve data that should meet their needs (loss of sales) and decreased consistency and granularity of data valuable in data mining and reporting for business intelligence.

Less than consistent and robust publisher supply chain metadata has implications for libraries as well, as this data drives our selection and acquisitions decisions and processes. For example, library vendors use this metadata to drive extraction of title information for approval lists, standing orders, opening day collections, etc.

Publishers have knowledge of forthcoming, new, and existing titles in their lists. They excel in the earliest knowledge of what will be published, close relationships with and knowledge of authors and other contributors, creation of descriptive text about the content they publish, metadata relating to publishing history of titles, metadata relating to the physical item, metadata relating to sales terms and rights, first knowledge of changes to publication dates, title changes, etc., and knowledge of reviews, awards, etc. relating to published titles.

Adding such metadata directly into WorldCat records will greatly enhance the search experience for users of library metadata—both inside and outside the library since

NEW GENERATION CATALOGUING

By Karen Calhoun and Renee Register

Excerpted from *Weaving Libraries into the Web*

library management also relies on metadata for business intelligence in the form of collection and circulation analysis.

OCLC is committed to pursuing development in this area as we are convinced there is a strong value proposition here for libraries, the publisher supply chain, and for OCLC's growth as an organization positioned to pro- vide cost-effective metadata and services for libraries and cultural heritage institutions of all types.

The value proposition for libraries can be characterized as follows:

- Provides comprehensive and reliable upstream metadata in OCLC World- Cat for use in selection, acquisition, circulation, and technical services.
- Provides early MARC record availability to the library market through OCLC cataloguing subscription services.
- Provides a mechanism for automatic receipt of record enhancements throughout the publishing cycle.
- Provides enhanced data and efficiencies in data creation or enhancement through use of OCLC tools such as FRBR, XISBN, and metadata creation and extraction.
- Reduces cost, labour, and duplication of effort in library cataloguing and streamlining library technical processes from selection to circulation.

The value proposition for the publisher supply chain can be expressed as follows:

- Reduces cost, labour, and duplication of effort in the creation, organization, enhancement, and distribution of metadata.
- Allows global visibility for available and forthcoming titles through WorldCat.
- Provides a streamlined method for data delivery to publishers and their supply-chain partners.
- Allows publishers and vendors to promise early MARC record availability to the library market.
- Enhances metadata for use in business-to-business, customer-facing, and marketing tools used by publishers and vendors.
- Provides enhanced data through data mapping from library schema to publishing industry terminology (e.g., Dewey to BISAC and BIC).

NEW GENERATION CATALOGUING

By Karen Calhoun and Renee Register

Excerpted from *Weaving Libraries into the Web*

PROCESS

The Next Generation Cataloguing process works as follows. OCLC receives files of title metadata in ONIX format from publisher supply chain partners. This is the same metadata that is routinely produced for internal and external functions relating to marketing, buying, and selling content.

OCLC crosswalks the ONIX metadata to MARC format and attempts to match to an existing WorldCat record. If an exact match is found, the WorldCat record is enriched with appropriate data from the ONIX record. Examples of WorldCat record enrichment from ONIX include: contributor biographical information, descriptions, annotations, and publishing industry BISAC subject headings.

If an exact match is found, the FRBR work set for the record is retrieved, and the matching record is enriched using appropriate data mined from the work set. The WorldCat record enrichment from FRBR work set could include additional subject headings and classification.

Where possible, new data is created from mapping between existing data elements (e.g., DDC to BISAC Subject Headings and BISAC Subject Headings to DDC). For example, if the matching record contains a Dewey of 616.07543, the BISAC Subject Code MED098000—Medical/Ultrasonography will be derived from mapping and added to the ONIX record. If the incoming ONIX has a BISAC code but the matching record has no Dewey, the inverse will occur, and the Dewey will be added to the MARC record.

Enriched records are cross walked back to ONIX, and the resulting ONIX file is returned to supply chain partners. Here are some examples of publisher metadata enrichment: authority controlled contributor names, contributor birth and death dates, Dewey call numbers, more granular BISAC subject headings derived from Dewey, LC call numbers, LCSH, NLM, and other subject schema, annotations and notes.

If no exact match is found a new record built from ONIX mapping to MARC is added to WorldCat. WorldCat attempts to FRBRize the new record, adding it to an existing work set when FRBR algorithms determine that data elements contained in the new record make it appropriate to do so. If the new record becomes part of an existing FRBR work set, records in the work set are used to enrich the newly created record. Enrichment occurs in such fields as DDC and LC Classification, authority controlled contributor names, LCSH, etc. As possible, new data is also created from mapping between existing data elements—DDC to BISAC Subject Headings and BISAC Subject Headings to DDC.

NEW GENERATION CATALOGUING

By Karen Calhoun and Renee Register

Excerpted from *Weaving Libraries into the Web*

RESULTS

A Symposium for Publishers and Libraries hosted at OCLC in early 2009 confirmed the value of these services for metadata creation and enrichment in support of both communities. OCLC is preparing to put the services outlined above into production simultaneous with pilot completion and publication of pilot results. Once final MARC and ONIX output files representing the full scope the service are delivered to library and publisher pilot partners, OCLC will compile case studies and pilot partner evaluations of the metadata created during the pilot.

OCLC is working with LC and NLM to test the viability of these processes for their cataloguing work flows, particularly in the area of CIP record creation. We believe this project is very much in line with the direction suggested by the LC Working Group on the Future of Bibliographic Control and will support the urgent need for more efficient and cost-effective methods of metadata creation, maintenance, and distribution.

OCLC is also maintains a close relationship with standards communities to ensure the services we build are in keeping with the direction sup- ported by these organizations. We commissioned a study in coordination with NISO (http://www.niso.org/home) to examine current metadata work flows for library and publisher supply chain metadata creation, maintenance, and delivery. We are in frequent communication with Editeur (http://www.editeur.org/), who develop and maintain the ONIX standard internationally, and with the Book Industry Study Group (BISG) (http://bisg.org/),

who support the ONIX standard in North America, maintain the BISAC Subject Codes and perform intensive research on publisher's supply chain practices, work flows, and economics.

CONCLUSION

OCLC is committed to supporting evolving metadata needs and the future of cataloguing by implementing the practices outlined in this article. We will routinely ingest, enhance and create metadata in WorldCat through these processes and routinely output enhanced metadata in both MARC and ONIX formats, with mechanisms for ongoing delivery of enhancements to libraries and publishers.

In support of the idea of collaborative growth and enhancement of metadata across the life cycle of titles, and in response to requests from the cataloguing community, OCLC is introducing the Expert Community Experiment, in which OCLC libraries with

NEW GENERATION CATALOGUING

By Karen Calhoun and Renee Register

Excerpted from *Weaving Libraries into the Web*

full-level cataloguing authorizations will be able to improve and upgrade many more WorldCat master bibliographic records. We hope that the experiment will result in more corrections and additions to master bibliographic records and more timely actions to correct record problems.

The experiment also allows OCLC to test a "social cataloguing" model involving the existing community of cataloguing experts who have built World- Cat record-by-record over the past four decades.

We also plan to enhance the Next Generation Cataloguing model so that we can routinely receive publisher updates to metadata across the title life cycle, including title changes, changes in publication date, and the addition of evaluative content and post-publication review and award information to WorldCat records as it is created by the publisher.

As a clearer picture of what the future of cataloguing might bring emerges, OCLC will remain actively engaged in the community and committed to developing products and services that support a 21st century vision of metadata creation, enrichment, maintenance, and distribution.

REFERENCE

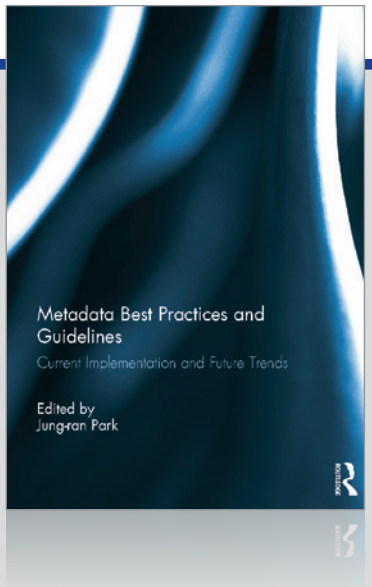
- The Library of Congress. On the Record: Report of The Library of Congress Working Group on the Future of Bibliographic Control. Retrieved from http://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf (January 9, 2008 Washington, DC).

CHAPTER

5

METADATA DECISION FOR DIGITAL LIBRARIES

A SURVEY REPORT



This chapter is excerpted from

Metadata Best Practices and Guidelines

Edited by Jung-ran Park

LEARN MORE ›

METADATA DECISION FOR DIGITAL LIBRARIES

Marcia Lei Zeng, Jaesun Lee and Allene F. Hayes

Excerpted from *Metadata Best Practices and Guidelines*

A survey on metadata conducted at the end of 2007 received over 400 answers from 49 countries all over the world. It helped the authors to identify major issues and concerns regarding metadata that should be addressed in the IFLA Guidelines for Digital Libraries. The questionnaire included a question of the roles respondents may have, and five questions of the major concerns in any project that relates to metadata, regarding design and planning of digital projects, element set standards, data contents in a record, authority files and controlled vocabularies, and metadata encoding. Findings from the survey are reported and a workflow chart is included in this paper.

BACKGROUND

In June 2005, the Librarian of Congress James H. Billington presented a proposal to UNESCO (United Nations Educational, Scientific and Cultural Organization) to establish a World Digital Library (WDL). The objectives of the World Digital Library are to: promote international and intercultural understanding and awareness, provide resources to educators, expand non-English and non-Western content on the Internet, and contribute to scholarly research. UNESCO and the Library of Congress co-sponsored an experts meeting in December 2006 with key stakeholders from all regions of the world. That meeting resulted in a decision to establish working groups to develop standards, best practices, and content selection guidelines.[1] The working groups are:

1 Selection and content working group

2 User research outreach and marketing group

3 Technical architectural working group

4 Best practices working group (IFLA Working Group on Digital Library Guidelines [WGDLG])

The IFLA (International Federation of Library Associations and Institutions) Working Group on Digital Library Guidelines was one of the four working groups recommended to be established at the conclusion of the UNESCO experts meeting; it is a standalone IFLA/UNESCO working group. The group has been supported by the WDL, which in return hopes to benefit from the results. Established in May 2007 by IFLA President Claudia Lux, the WGDLG is composed of representatives from several IFLA sections. The group's objective is to develop digital library guidelines and best practices with recommendations on the various aspects of a digital library in order to help libraries build, publish, provide access to, and share digital collections in a standardized way. The guidelines are intended to be used by libraries and other cultural institutions around the world.[2]

METADATA DECISION FOR DIGITAL LIBRARIES

Marcia Lei Zeng, Jaesun Lee and Allene F. Hayes

Excerpted from *Metadata Best Practices and Guidelines*

At the IFLA WGDLG's first meeting held at the Library of Congress in May 2007, the working group decided to include a chapter on metadata in the IFLA Guidelines for Digital Libraries. The authors of this paper, who are working-group members from IFLA Division IV (Bibliographic Control)[3] and the Library of Congress, are responsible for the metadata chapter. In preparing the chapter on metadata for the Guidelines, the authors developed a questionnaire that aimed to identify the major issues and concerns regarding metadata and controlled vocabularies that needed to be addressed in the Guidelines. The authors then conducted a survey on metadata decisions in late 2007 and analysed the survey data in 2008. This paper reports the feedback from the survey and the resultant chapter content.

REVIEW OF RELATED BEST PRACTICES AND GUIDELINES

Metadata decisions may be made at different stages of a digital library project, and intelligent decisions are integral to successful implementation of the project. Questions that arise at the beginning stages of a digital collection project can be all-important and determine the quality and consistency of all subsequent phases of metadata creation, implementation, and interoperability. Even after a digital collection is built, there may still be metadata-related questions if it is involved in further collaboration and development. Considering metadata as a unique component in a digital collection, A Framework of Guidance for Building Good Digital Collections, issued by a NISO working group (2nd edition, 2004; 3rd edition, 2007),[4] presents a set of requirements for metadata. Among them, some have long since been implemented by the conventions of library cataloguing (such as, conforming to community standards, supporting interoperability, and, the employing of authority control and content standards), while other requirements pay attention to the newer particular functions of administration, rights management, and preservation. This clearly indicates that metadata creators must have knowledge beyond the application of the rules specified by structure and content standards; they must now be involved in decisions beyond descriptive cataloguing, beginning from the very outset of a digital collection project. The Handbook on Cultural Web User Interaction, edited by MINERVA EC (MInisterial NEtwoRk for Valorising Activities in digitisation, eContentplus) Working Group Quality, Accessibility and Usability, suggests an increasing importance of metadata issues in the cultural Web world. MINERVA's seventh principle of quality states: "A good quality cultural website must be committed to being interoperable within cultural networks to enable users to easily locate the content and services that meet their needs." A related document is the MINERVA

METADATA DECISION FOR DIGITAL LIBRARIES

Marcia Lei Zeng, Jaesun Lee and Allene F. Hayes

Excerpted from *Metadata Best Practices and Guidelines*

Technical Guidelines for Digital Cultural Content Creation Programmes (2008) which has a full chapter "Metadata, standards and resource discovery." The chapter provides examples along with best practices for descriptive, administrative, preservation, and structural metadata, as well as collection-level description.[5]

Best practices provide guidance and information for the most efficient (least effort and expense) and effective (best results and function) ways of accomplishing a task and are empirically based on repeatable procedures in different settings. Project-based and metadata standard-centred best practices and guidelines have been available for some time for usage in digital collections and digital libraries, and usually include general guidelines that are related to metadata planning. The National Science Digital Library (NSDL)'s NSDL DC Metadata Guidelines, for example, covers overarching considerations and issues, background knowledge, decisions on what to describe, and appropriate levels of granularity.[6] A comparable document is the Best Practices for OAI Data Provider Implementations and Shareable Metadata, a joint initiative between the Digital Library Federation and the NSDL. It includes two best practices guides: (1) Best Practices for OAI Data Provider Implementations and (2) Best Practices for Shareable Metadata.[7] Similarly, a white paper, "Preliminary Recommendations for Shareable Metadata Best Practices" was released as part of a three-year interim project report for the IMLS Digital Collections and Content Project hosted by the University of Illinois at Urbana-Champaign.8 The recommendations emphasized sharable metadata creation, which ensures that data will remain meaningful in a broader context (regardless of the local environment in which it was created).

DATA COLLECTING

The authors created a questionnaire to identify major issues and concerns regarding metadata that should be addressed in the chapter on metadata in the IFLA Guidelines for Digital Libraries. It included

1 a question of the roles respondents may have and

2 five main questions of the major concerns in any project that relates to metadata regarding

- design and planning of digital projects
- element set standards (data structure decision)
- data contents in a record (data content decision)

METADATA DECISION FOR DIGITAL LIBRARIES

Marcia Lei Zeng, Jaesun Lee and Allene F. Hayes

Excerpted from *Metadata Best Practices and Guidelines*

- authority files and controlled vocabularies (data value decision), and
- metadata encoding (data format/technical interchange decision)

The draft questionnaire was distributed to members of the IFLA Cataloguing Section's Standing Committee at the August 2007 IFLA conference held in Durban, South Africa. The Standing Committee consists of 20 members from different countries. Based on the valuable suggestions collected during this preliminary review, the questionnaire was revised and transformed into a Web-based form utilizing Surveymonkey.com's survey tool.

A letter seeking respondents was sent through the IFLA listserv and further forwarded by IFLA members to the professional listservs in their respective countries and communities. During a one-month period (from October to November 2007) over 400 answers from 49 countries in Asia, Africa, North America, South America, Europe, and Australia were received. These included answers from individual professionals as well as collective answers from several national libraries and many institutions. In addition to respondents from the countries who are active members of IFLA Division IV, there were responses from many other countries, including Albania, Azerbaijan, Cameroon, Costa Rica, Jamaica, Lithuania, Malta, Moldova, Mongolia, and Nigeria.

Among the 417 valid questionnaire answers, a total of 413 respondents answered the question "Which of the following best describe your role in your digital collection/digital library project(s)? (Please check all that apply)." The roles of the respondents are summarized in Table 1, with a rank according to the response percentage and count.

Role	Response %	Response #
creating metadata records	47.50	196
supervising metadata and/or cataloging project(s)	45.30	187
consulting on metadata issues	40.40	167
developing policies and best practices	40.40	167
coordinating digital collection/digital library projects	39.50	163
creating and maintaining controlled vocabularies (lists of subject headings, thesauri, taxonomies, etc.) and authority files	32.00	132
teaching and training information professionals	28.60	118
consulting on vocabulary control issues	27.60	114
providing technical support to the digital library projects	21.30	88

Table 1 • Respondents' Roles in Digital Collection/Digital Library Projects (413 Answered, Each Respondent could Choose All that Apply)

METADATA DECISION FOR DIGITAL LIBRARIES

Marcia Lei Zeng, Jaesun Lee and Allene F. Hayes

Excerpted from *Metadata Best Practices and Guidelines*

About half of the respondents work directly with metadata creation as a creator and/or a supervisor. About 40% have roles beyond creating metadata, which include consulting, policy making, and coordinating for the metadata- related issues and work. Related to these, 32% of respondents' roles include creating and maintaining controlled vocabularies and authority files, and 27% have been consulted on vocabulary control issues. This indicates that vocabulary and authority control is a very important aspect during the whole metadata process. Also, nearly 29% of the respondents have been involved in the teaching and training of information professionals. This is likely because of the demands of dealing with newer metadata standards beyond MARC and a much larger and dynamic metadata creation workforce that requires more up-to-date training than ever before.

Thirty-four (8.2%) respondents chose "Other" as their response. An analysis of these answers found that half of them can be categorized into the roles of coordinating projects, technical support, and education. Additional categories include: information architecture (including interface design, portal administration, and search engine development), marketing and promoting digital libraries, funding, human resource development and management, evaluation, database analysis, and metadata schema development.

DATA ANALYSIS: RESPONSES TO FIVE "MAJOR CONCERNS" QUESTIONS

Five issues were listed under the second question, "What are the major concerns you have in your project that relate to metadata?" There were two comments indicating that the word "concerns" was not clearly defined since it could relate to "worries." They indicated that people may be "worried" about something because they do not know the best way to approach them, whereas the areas they would be paying the most attention to would be the "concerns." This may have had some impact on specific answers.

Ideally, metadata-related standards should be selected according to their purposes and their relationship to the workflow of a digital library. Therefore, after the first question regarding the overall design and planning, the questions were based on the types of standards that have been created by different communities for specific purposes. They included

- Standards for data structures. Metadata element sets are standards for data structures and semantics (e.g., Dublin Core Metadata Element Set).

METADATA DECISION FOR DIGITAL LIBRARIES

Marcia Lei Zeng, Jaesun Lee and Allene F. Hayes

Excerpted from *Metadata Best Practices and Guidelines*

- Standards for data content. Data content standards are created to guide the practices of metadata generation or cataloging (e.g., Anglo-American Cataloging Rules, Second Edition (AACR2), Cataloging Cultural Objects (CCO): A Guide to Describing Cultural Works and Their Images, and Describing Archives: A Content Standard (DACS)).
- Standards for data values (referred to as value encoding schemes in a metadata standard). These include controlled-term lists, classification schemes, standardized codes, thesauri, authority files, and lists of subject headings.
- Standards for data exchange (often referred to as formats in the context of data exchange and communication). They are standards for data exchange, separately designed or bound together with the element sets.

Question 2.1: Major Concerns—For Designing and Planning of Digital Projects

Question 2.1 intended to form a general picture of the major concerns when designing and planning a digital library as related to metadata. The suggested areas of concerns and responses are listed in **Table 2**, from the most selected to the least selected.

A majority of the areas listed under this question received responses of over 40% of concerns from 324 respondents. The six areas match, and are consistent with **Table 2** and are numerically ordered in the same way.

- to understand possible workflows
- to consider reusing existing cataloguing records by integrating them or transforming them to other formats in the new project
- to plan how search functions can be supported by metadata information
- to explore how to include various types of resources in one project
- to learn how to measure and control metadata quality
- to decide upon levels of description (e.g., item level, collection level)

The feedback reflects the changing and challenging nature of current metadata-creation work highlighting how it differs from conventional cataloguing work, which has followed only a few established rules and formats for a well-established period of time. A digital library project decision-maker has to fully understand what is really involved (and the concomitant results) when introducing new format(s) and standard(s) and whether to treat different types of resources in the same way or different ways.

METADATA DECISION FOR DIGITAL LIBRARIES

Marcia Lei Zeng, Jaesun Lee and Allene F. Hayes

Excerpted from *Metadata Best Practices and Guidelines*

Concern	Response %	Response #
above 50%		
to understand possible workflows	58.30	189
to consider reusing existing cataloging records by integrating them or transforming them to other formats, e.g., MARC to DC, a local format to EAD, etc., or any other variation in the new project	58.30	189
to plan how search functions can be supported by metadata information	56.80	184
to explore how to include various types of resources (print, web pages, images, etc.) in one project	50.60	164
above 40%		
to learn how to measure and control metadata quality	49.40	160
to decide upon levels of description (e.g., item level, collection level)	47.80	155
to find if any metadata exist already in the objects themselves that could be extracted automatically and what tools are available for this	43.50	141
to understand types of metadata (e.g., descriptive, administrative, structural, preservation, rights metadata)	43.50	141
to see examples from similar projects	41.00	133
above 30%		
to plan how metadata records will be linked with authority records	39.80	129
to plan how the metadata describing a physical object will be associated with the metadata for its digital version	38.60	125
to understand the mechanisms of harvesting protocols	36.70	119
to understand the value of controlled vocabularies	32.70	106
to understand and adopt an abstract model (e.g., Dublin Core Abstract Model, FRBR conceptual model, CCO entity-relationship model)	31.50	102

Table 2 • Major Concerns Related to Designing and Planning of Digital Library Projects (324 Answered, Each Respondent Could Choose All that Apply)

In their additional comments, respondents expressed their concerns such as: "to plan effective workflows with stable and supported tools," "to consider how to display metadata for digital projects in the OPAC," "to plan how metadata records will be linked with authority records," and "to understand the impact of metadata on the ability to build resource discovery collection structures to facilitate browsed searching of the digital collections." Since multiple options of introducing new and dynamic metadata formats have become unavoidable issues, the respondents were highly concerned about interoperability issues. Their concerns ranged from "to plan and map together various metadata templates" to "to make standards used by various communities interoperable within one discovery system."

Based on the survey, the authors decided to prepare a workflow chart to include in the Guidelines and use it to state these major areas of concern. The area of least

METADATA DECISION FOR DIGITAL LIBRARIES

Marcia Lei Zeng, Jaesun Lee and Allene F. Hayes

Excerpted from *Metadata Best Practices and Guidelines*

concern was to understand and adopt an abstract model. Therefore, abstract models were not included in the current draft of this chapter.

Question 2.2. Major Concerns—For the Decisions About Element Set Standards (= Data Structure Decisions)

Question 2.2 looked at decisions related to metadata structures that are usually defined in metadata element sets. The questionnaire provided an explanation of what "element set standards" are, because the data structure standards have been named differently in practices ("element set," "scheme," "data dictionary," and "schema" are among the most common terms). Examples of metadata standards for data structures include Dublin Core, MARC, MODS (Metadata Object Description Schema), VRA (Visual Resources Association) Core, EAD (Encoded Archival Description), and CDWA Lite. The survey encouraged checking all major concerns that might apply.

This question received feedback from 303 respondents (see Table 3). It is clear that before deciding to develop an application profile (33.00%) and make a crosswalk (41.60%), the major question would be how to find out which metadata standard should be used (62.40%). Since most projects have dealt with different types of resources and many would work with the services already in existence (e.g., library systems that use MARC 21 or UNIMARC), it would be important to learn how to employ different metadata schemes together in one project (59.40%).

Concern	Response %	Response #
to decide which metadata standard to use	62.40	189
to learn how to use different metadata schemes together in one project	59.40	180
to understand what factors influence the decision on which metadata standard to use, e.g., what sort of material they are good for	58.70	178
to find out what standards are available	47.90	145
to understand what sorts of adjustments might be made to a standard metadata schema that could result in a separate schema and/or application profile	42.60	129
to learn how to create crosswalks	41.60	126
to decide whether an application profile should be developed	33.00	100

Table 3 • Major Concerns Regarding Decisions About Data Structure (303 Answered, Each Respondent Could Choose All that Apply)

To respond to the feedback, the authors included a section in the Meta- data chapter on functional requirements. It explained various types of meta- data and indicated

METADATA DECISION FOR DIGITAL LIBRARIES

Marcia Lei Zeng, Jaesun Lee and Allene F. Hayes

Excerpted from *Metadata Best Practices and Guidelines*

that different types of metadata elements would be essential to perform different tasks. A limited list of metadata standards for data structures were presented (for 58.70% responses) according to the sort of material these standards could be applied to. A reference link provided an externally maintained list of many standards that are available (for nearly 48% responses that expressed this concern).

Question 2.3: Major Concerns—For the Decisions About Data Content in a Record (= Data Content Decisions)

The practices of metadata generation have a direct influence on the quality of metadata. For example, does a record correctly describe the resource and provide enough information? Does it consistently apply methods and format in each description? Question 2.3's intention was to find out the concerns for the decisions about data content in a record.

All areas listed under this question received over 50% ratings among the 292 respondents (see Table 4). As digital resources come into the mainstream of digital library projects, additional metadata types (e.g., administrative, technical, and use metadata) other than descriptive metadata become increasingly more important. The cost-effectiveness aspect of metadata creation was also of high concern for digital library projects (nearly 72%); this consideration would certainly affect the requirements of minimum and mandatory elements to be included in each record.

Concern	Response %	Response #
to decide which core elements should be included in all records (e.g., is RIGHTS information required), which elements are mandatory, and which are repeatable	71.90	210
to provide guides in order to ensure that metadata values will be entered consistently (e.g., for DATE, FORMAT information)	68.50	200
to decide which elements (e.g., SUBJECT, CREATOR) should use a controlled vocabulary/authority file	66.10	193
to find existing data content (i.e., cataloging) standards and best practice guides (e.g., Anglo-American Cataloging Rules (AACR), Cataloging Culture Objects (CCO), Describing Archives: A Content Standard (DACS), etc.)	53.10	155
to learn how to provide correct information in a record (e.g., where to find TITLE information from a Web site, what are the IDENTIFIERs, how many IDENTIFIERs should be included, etc.)	51.00	149

Table 4 • Major Concerns Regarding Decisions About Data Content (292 Answered, Each Respondent Could Choose All that Apply)

METADATA DECISION FOR DIGITAL LIBRARIES

Marcia Lei Zeng, Jaesun Lee and Allene F. Hayes

Excerpted from *Metadata Best Practices and Guidelines*

The library, archive, visual resources, and museum communities all have different best practices in metadata creation. They have developed and used certain content standards such as AACR2, CCO, and DACS. More and more brief, or detailed, best practices guides have also been written by various institutions in building their digital collections. However, many of these remain in silos and have not been shared beyond specific institutions and collections. Another thing to be recognized is that most of the metadata standards developed during the past twenty years have already provided various best practices guidelines (many included in the metadata element sets; some pre- pared as separate documents). In the metadata standards, they can be found under headings such as "comments," "description," "data value," "explanation," "value space," and "examples." Nevertheless, because these might be too general, a metadata creator may lack the necessary guidance in handling day-to-day problems. Therefore, application profiles designed for specialized communities would do well to provide detailed guides and examples. The suggested steps and content for application profiles have been included in the workflow chart created for the Guidelines. In the workflow chart, the authors especially emphasize the points of developing and sharing best practices and building application profiles to efficiently ensure high quality metadata.

Question 2.4: Major Concerns—For the Decisions About Authority Files and Controlled Vocabularies (= Data Value Decisions)

Question 2.4 is about data value decisions (see Table 5). Controlled vocabularies (also known as encoding schemes) and rules are usually required by metadata standards and application profiles for the values associated with subjects, media formats, resource types, audience levels, and so on.

Concern	Response %	Response #
to decide whether to use existing controlled vocabularies or authority files (e.g., LCSH, ULAN [The Union List of Artist Names], LC Authorities)	64.60	177
to develop controlled vocabularies (including controlled lists, taxonomies, thesauri, etc.)	53.30	146
to maintain our own authority files and controlled vocabularies	48.90	134
to establish our own authority files for names	35.00	96

Table 5 • Major Concerns Regarding Decisions About Data Values (274 Answered, Each Respondent Could Choose All that Apply)

METADATA DECISION FOR DIGITAL LIBRARIES

Marcia Lei Zeng, Jaesun Lee and Allene F. Hayes

Excerpted from *Metadata Best Practices and Guidelines*

Library communities already have a long history of developing and employing controlled vocabularies and authority files. In addition to using these existing value encoding schemes, developing controlled vocabularies for a specific project seemed to be of major importance (53.3%) among the 274 respondents to this question.

It is the authors' understanding that for some value spaces, a small, predefined list of terms is useful and efficient to build, especially when a particular attribute of a resource may not be accurately described by existing controlled vocabularies (which either may be too large and comprehensive, or not specific enough). A list of terms can then be predefined by those who build or implement a standard (or an application profile) to describe aspects of content objects or entities that have a limited number of possibilities. Some metadata standards (e.g., LOM, VRA Core) have provided small predefined lists of terms for particular elements' value spaces (e.g., learning object types). To respond to the concerns and introduce such an approach, "controlled term lists" was included in the metadata workflow chart and it was listed ahead of other larger and complex schemes.

Question 2.5: Major Concerns—For the Decisions About Metadata Encoding (= Data Format/Technical Interchange Decisions)

This last question in the "major concerns" category targeted decisions about metadata encoding. The question reminded the respondents that "metadata records can be represented in many syntax formats such as XML, RDF, HTML/XHTM." It is a technical question related to data format or technical interchange; therefore, only three very general questions were asked; 65% (272 out of 417) of the respondents indicated their concerns (see Table 6).

Concern	Response %	Response #
to learn about available tools for encoding and converting records	79.40	216
to understand what are the universal or widely used encoding formats	67.60	184
to see examples of encoded records	60.30	164

Table 6 • Major Concerns Regarding Decisions About Data Format and Technical Interchange (272 Answered, Each Respondent Could Choose All that Apply)

The feedback indicated how important it is for the digital library developers to learn about available tools (nearly 80%), understand encoding formats (nearly 68%), and view examples of encoded records (over 60%). In addition to the major concern given

METADATA DECISION FOR DIGITAL LIBRARIES

Marcia Lei Zeng, Jaesun Lee and Allene F. Hayes

Excerpted from *Metadata Best Practices and Guidelines*

to the available tools for encoding and converting records under this part of the questionnaire, two dozen respondents also made additional comments emphasizing the need for tools. Some comments indicated that this area had been handled by their IT departments; yet, "this is the area of major concern as we don't have the technical expertise in my department and have to rely on the systems department." This issue will be addressed in other chapters of the Guidelines.

COMMENTS IN THE OPEN QUESTIONS

For each question, the questionnaire provided an "other" option and welcomed additional comments. At the end of the questionnaire, an open question "Which of your major concerns were not addressed in this questionnaire?" was also included (see Table 7).

These comments deserve special attention because they reflected some cross-board issues. The respondents raised questions and concerns that could be found at different stages of a digital library project, affecting different parts of the collaborative effort, and relating to various procedures.

Question	Response %	Response #
1. Which of the following best describe your role in your digital collection/digital library project(s)? Other (please specify):	8.2	34
2. What are the major concerns you have in your project that relate to metadata?		
2.1 For design and planning of digital projects Other (please specify):	5.2	17
2.2 For the decisions about element set standards (= data structure decisions) Other (please specify):	4.3	13
2.3 For the decisions about data content in a record (= data content decisions) Other (please specify):	1.4	4
2.4 For the decisions about authority files and controlled vocabularies (= data value decisions) Other (please specify):	8.0	22
2.5 For the decisions about metadata encoding (= data format/technical interchange decisions) Other (please specify):	6.3	17
2.6. General comments Which of your major concerns were not addressed in this questionnaire?	17.68	73

Table 7 • Number of Responses to Open Questions (146 Answers; Each Responder Could Answer More Than One Open Question)

METADATA DECISION FOR DIGITAL LIBRARIES

Marcia Lei Zeng, Jaesun Lee and Allene F. Hayes

Excerpted from *Metadata Best Practices and Guidelines*

A few comments gave more details and can be considered as "zooming- in" on the issues and areas covered in the survey. A majority of the comments, however, can be further categorized to extend the issues and areas. In general, they encourage the authors of this article to "zoom-out," to put the metadata-related issues into a larger context. Taking a step up, or standing at a higher level, one can see the major issues at two layers (statements were selected from the comments):

LAYER 1: SIGNIFICANT ISSUES CROSSING ALL QUESTIONS

- Standardization and interoperability
  - All levels should consider using standards: structures, formats, tools, and products.
  - Levels of interoperability should be not only syndetic but also semantic (implying not only that data elements and fields be crosswalk-able but also that the values be correctly converted and exchanged).
  - Sharable data should be produced and provided, including descriptive data, subject vocabularies, and even file-naming conventions.
  - Metadata for Web archiving and publishers' metadata should be included.
- Extensibility
  - Decisions should be made whether to create extension elements or separate schemas and only extract the useful elements.
- Multilingualism
  - It is important to consider correct character sets for encoding non-Roman languages.
- Quality vs. efficiency
  - Quality of metadata, especially in the non-MARC format or input by nonprofessional's, became a clearer issue than was previously realized. Metadata creation is a costly process. Metadata production consumes enormous amounts of time.
  - It is still not clear how to calculate the hidden costs associated with different metadata decisions.
  - Metadata architecture should be studied to explore harvesting models and query models so that metadata can be shared and used efficiently and automatically.

METADATA DECISION FOR DIGITAL LIBRARIES

Marcia Lei Zeng, Jaesun Lee and Allene F. Hayes

Excerpted from *Metadata Best Practices and Guidelines*

- Among the more specific comments, respondents suggested:
  - the need to explore how to capture metadata in the most efficient way, e.g., generate values automatically, reduce the number of mandatory fields for metadata creators, harvest from other repositories
  - the need to explore ways to introduce user-generated metadata, e.g., tagging, reviews, and how best to incorporate it with traditional metadata
  - the need to get search engines to handle metadata so that users get the greatest benefit

- Staffing and Training (This area generated many comments.)
  - Good metadata creators are in high demand. Training is critical to not only nonlibrary professionals and non-cataloguers, but also the cataloguers who have been trained only in more traditional conventions.
- More open and flexible choices
  - Strongly encourage exploring non-MARC format, as emerging (and in some cases, established) standards for the creation of document structures and metadata provide greater flexibility and better integration with mainstream software applications such as enterprise-scale databases.
  - Suggest discovering "how to export and share metadata from a digital project into an aggregated environment—either our own aggregation, or as part of a larger community. Beyond OAI/Dublin Core!"

In responding to some of these concerns, four principles were included in the Guidelines to guide the decisions about metadata element sets and/or application profiles and their implementation in a digital library project: extensibility, interoperability, modularity, and multilingualism. It needs to be pointed out that communication about the functional requirements between both system designers and metadata creators is critical to the overall quality of a digital library. For a digital library to be truly successful, expertise from both teams is irreplaceable. To aim for cost-effective metadata generation, human-generated, machine-generated, publisher-produced, library- produced, professional-created, and user-contributed data should all be complementary to one another. This collective process will increase efficiency and productivity without sacrificing quality and effectiveness.

METADATA DECISION FOR DIGITAL LIBRARIES

Marcia Lei Zeng, Jaesun Lee and Allene F. Hayes

Excerpted from *Metadata Best Practices and Guidelines*

LAYER 2: SIGNIFICANT ISSUES BEYOND METADATA COMMUNITIES

- Tools for original metadata
  - "Metadata creation requires solid tools, and so far there are not any widely established systems for the creation of (for example) VRA Core records, DDI, records, etc. The digital library arena needs to evolve to develop some stable and sustainable tools to feed processing, discovery, and preservation workflows."
  - Metadata creation tools must be easy-to-use and affordable. More specifically, a responder expressed a desire to "have tools that allow users with no XML knowledge the ability to create MODS/DC/VRA Core records, etc., preferably without having to see the XML code of the record that enables metadata creators to just concentrate on the content [rather] than working with the XML codes."
  - Vendors of library-integrated systems should provide some useful tools. "Today the majority of the small-to-midrange library community is stuck in its own ILS silo using MARC ... ILS which doesn't play well with other metadata. ILS vendors need to advance more quickly or the library community will become more marginalized."
- Outreach—calling for actions to break silos
  - Spread the wealth and make this activity more mainstream and less of
  - The library integrated systems and digital libraries should be synchronized rather than being isolated and separately developed and operated: "integrating digital projects into routine work of the libraries, i.e., moving from isolated digital projects to a digital library program."
  - Stakeholders need to have increased awareness and accept the importance of controlled vocabularies and metadata.

Comments and issues in the second layer placed the metadata components into the digital library project and also equally placed those projects into a much larger and wider setting. Fortunately, in the Guidelines there are other chapters that will address those issues. In the workflow chart the authors also detailed a larger context from beginning to end of the workflow.

SUMMARY AND CONCLUSION

This world-wide survey provided a beginning for a common consensus about metadata-related issues and concerns. The feedback reflected the changing and challenging nature of current metadata-creation work that differs from conventional cataloguing work. Although the data was collected in late 2007, the continued

METADATA DECISION FOR DIGITAL LIBRARIES

Marcia Lei Zeng, Jaesun Lee and Allene F. Hayes

Excerpted from *Metadata Best Practices and Guidelines*

monitoring of the issues and trends by the authors has indicated consistency of these main concerns, especially with the growth of the digital collections and digital libraries around the world. It is important for all digital library developers to recognize that metadata element sets, content standards, and value-encoding schemes are created with the intent of guiding and ensuring the construction of high-quality metadata records. This will guarantee the correct implementation of metadata standards and will support digital library functions. These building blocks need to be used in the construction of efficient and functional information architecture through metadata services and technologies.

Based on the invaluable information from this survey the authors have incorporated as much as possible in the writing of the chapter within a six-page limit. The survey results helped to generate a concise chapter on metadata for the IFLA Guidelines for Digital Libraries that is to be released in 2010. The authors would like to use this opportunity to thank all who participated. As a token of appreciation, a current version of the workflow chart is included in this article. The final version of the Guidelines should be consulted when it becomes available.

NOTES

1 About the World Digital Library: Background. http://www.wdl.org/en/about/background.html

2 Working group on digital library guidelines meets in Washington. IFLA Journal, 33(3), 277–278 (2007).

3 IFLA Division IV, Division of Bibliographic Control Web page found at http://www.ifla.org/VII/d4/dbc.htm

4 NISO Framework Advisory Group. 2007. A Framework of Guidance for Building Good Digital Collections. 3rd ed. Priscilla Caplan et al.

5 MINERVA EC Working Group on Quality, Accessibility and Usability. 2008. Handbook on Cultural Web User Interaction.

6 NSDL DC Metadata Guidelines. http://nsdl.org/collection/metadata-guide.php

7 DLF and NSDL. [last modified July 2007]. Best Practices for OAI Data Provider Implementations and Shareable Metadata. http://webservices.itcs.umich.edu/mediawiki/oaibp/?PublicTOC

8 Jackson, Amy. 2006. Preliminary Recommendations for Shareable Metadata Best Practices, a white paper found at http://www.ideals.uiuc.edu/bitstream/handle/2142/719/shareable%20metadata.pdf?sequence=2

METADATA DECISION FOR DIGITAL LIBRARIES

Marcia Lei Zeng, Jaesun Lee and Allene F. Hayes

Excerpted from *Metadata Best Practices and Guidelines*

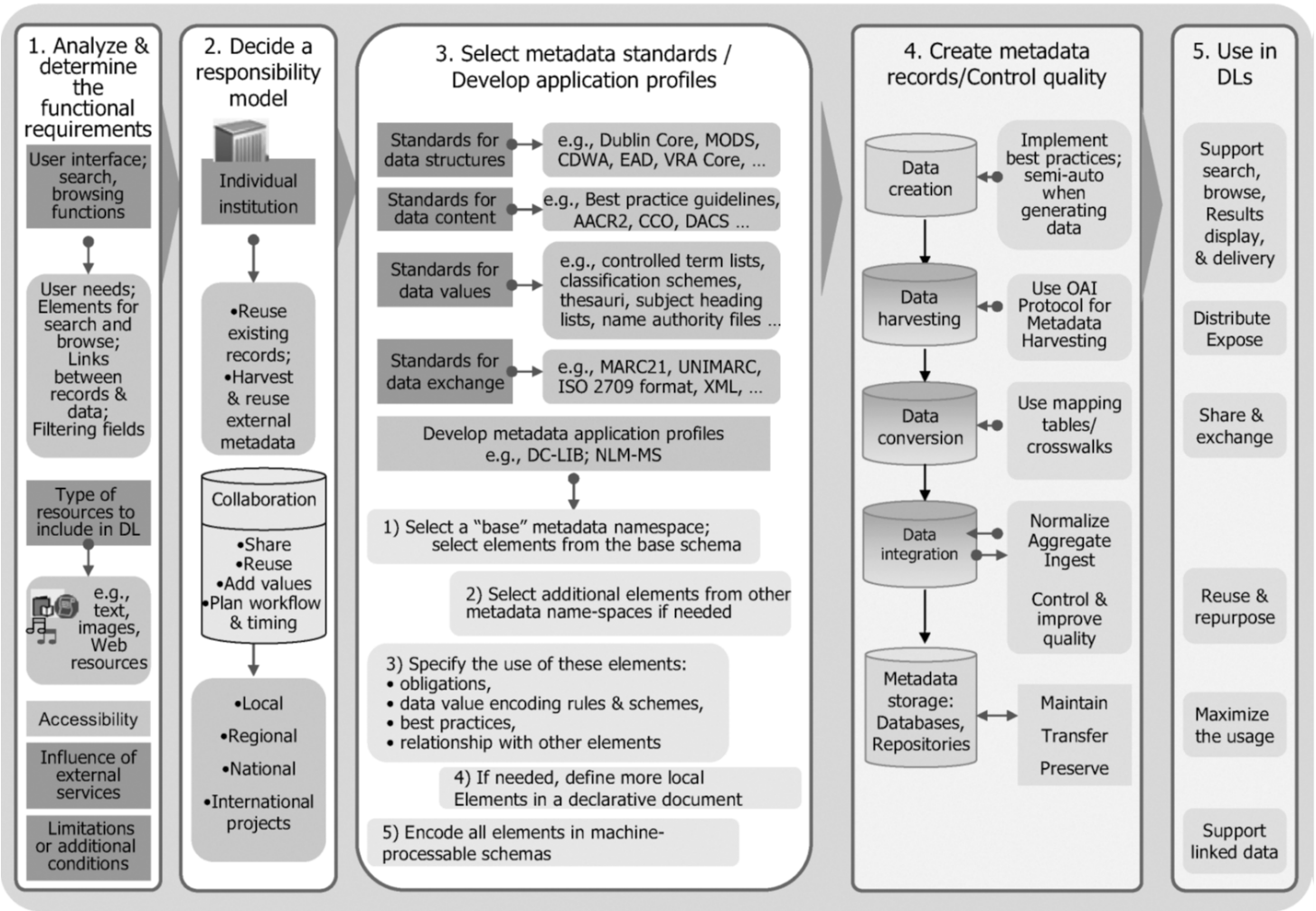
REFERENCES

- Baca, M., Harpring, P., Lanzi, E., McRae, L., & Whiteside, A., (Eds.) on behalf of the Visual Resources Association. (2006). *Cataloguing cultural objects, a guide to describing cultural works and their images*. Chicago: American Library Association.
- DLF (Digital Library Federation) and NSDL (National Science Digital Library). (last modified July 2007). *Best Practices for OAI Data Provider Implementations and Shareable Metadata*. Available at http://webservices.itcs.umich.edu/mediawiki/oaibp/?PublicTOC
- Duval, E., Hodgins, W., Sutton, S., & Weibel, S. L. (2002). Metadata principles and practicalities. *D-Lib Magazine* [Online] 8(4). Available at http://www.dlib/org/dlib/april02/weibel/04weibel.html
- Jackson, A. (2006). *Preliminary recommendations for shareable metadata best practices, a white paper*. Champaign, IL: Digital Collections and Content Project, Grainger Engineering Library, University of Illinois at Urbana-Champaign. Available at http://www.ideals.uiuc.edu/bitstream/handle/2142/719/shareable%20metadata.pdf?sequence2
- MINERVA EC Working Group on Quality, Accessibility and Usability. (2008). *Handbook on Cultural Web User Interaction*. 1st ed. Available at http://www.minervaeurope.org/publications/handbookwebusers.htm
- Nilsson, M., Baker, T., & Johnston, P. (2008). *The Singapore Framework for Dublin Core Application Profiles*. Available at http://dublincore.org/documents/2008/01/14/singapore-framework/
- NISO Framework Advisory Group. (2004). *A framework of guidance for building good digital collections*. 2nd ed. G. Agnew, L. Bishoff, P. Caplan, R. Guenther, & I. Hsieh-Yee. Available at http://www.niso.org/framework/Framework2.html
- NISO Framework Advisory Group. (2007). *A Framework of Guidance for Building Good Digital Collections*. 3rd ed. P. Caplan, G. Agnew, M. Baca, C. Fleischhauer, T. Gill, I. Hsieh-Yee, J. Koelling, C. Stephenson, & K. A. Wetzel. Available at http://www.niso.org/publications/rp/framework3.pdf
- NSDL (National Science Digital Library). [n.d.] *NSDL DC Metadata Guidelines*. Available at http://nsdl.org/collection/metadata-guide.php
- *Understanding Metadata*. National Information Standards Organization. 2004. Bethesda, MD: NISO Press. Available at http://www.niso.org/standards/resources/UnderstandingMetadata.pdf

METADATA DECISION FOR DIGITAL LIBRARIES

Marcia Lei Zeng, Jaesun Lee and Allene F. Hayes

Excerpted from *Metadata Best Practices and Guidelines*



APPENDIX A

METADATA WORKFLOW

As illustrated in the figure, the metadata process in a digital library should follow the following workflow:

1 Analyse and determine the functional requirements relating to user needs, interface and features of search and browse, types of resources to be stored, granularity levels of descriptions, limitations or conditions, accessibility features, etc.

2 Decide on a metadata creation responsibility model. Will the metadata project be in-house or part of a cooperative project? Will previous records be reused? Will data be harvested from external sources? What and how should data resources (e.g., publisher-provided, user-contributed, and auto-captured) be used?

METADATA DECISION FOR DIGITAL LIBRARIES

Marcia Lei Zeng, Jaesun Lee and Allene F. Hayes

Excerpted from *Metadata Best Practices and Guidelines*

3 Select the appropriate metadata standards and design a metadata application profile. Considerations should include the metadata element set, best practices guidelines and data content standards, data value standards, and authority files to be used to create metadata records. In an application profile, specify localized refinements, required encoding syntax rules, and recommended controlled vocabularies. Create or use crosswalks when multiple metadata element sets are involved. Application profiles should be encoded in machine-processable schemas following encoding standards in order to be implemented, registered, and exchanged correctly.

4 Create shareable metadata records and implement quality control from the beginning. Use tools for data input, data update, metadata harvesting, conversion, validation, and storage. Implement technologies to improve quality of existing metadata for maximized discovery and delivery of re- sources. Store, maintain, and preserve metadata.

5 Provide means to use, distribute, share and exchange metadata records, thereby making records available for harvesting by other organizations and aggregators. Consider metadata reuse, repurpose, and maximize their usage. Support linked data and create metadata so as to become linked data.

APPENDIX B

THE SURVEY INSTRUMENT

Brief Survey on the Metadata Decisions for Digital Libraries

Dear Library and Information Professionals,

We are collecting your suggestions to be used in preparing a chapter on metadata decisions for the Digital Library Guidelines, a task of the IFLA- World Digital Library Working Group on Digital Library Guidelines. The Guidelines will be developed for use by libraries and other cultural institutions around the world. The purpose of this survey is to investigate different issues, levels, and concerns regarding metadata and controlled vocabularies that need to be addressed in the Guidelines.

Please take 3–5 minutes to answer these questions on the survey available at: http://www.surveymonkey.com/s.aspx?smlRTMlZ 2bVEGf8 zmNCQPS3fg3d3d. Or, you can answer the same questions attached in this email and send them back to us at mzeng@.kent.edu or jslee@mail.nl.go.kr.

If you would like to know more about this research project, please call Marcia Zeng at (1) 330.672.0009 or email her at mzeng@kent.edu. This project has been approved by Kent State University. If you have questions about Kent State University's rules for research, please call Dr. John L. West, Vice President and Dean, Division of Research and Graduate Studies (Tel. 1-330.672.2704). Thank you for your participation in this survey.

Sincerely,
Marcia Zeng, Kent State University
Jaesun Lee, The National Library of Korea
Allene Hayes, Library of Congress

METADATA DECISION FOR DIGITAL LIBRARIES

Marcia Lei Zeng, Jaesun Lee and Allene F. Hayes

Excerpted from *Metadata Best Practices and Guidelines*

1 Which of the following best describe your role in your digital collection/digital library project(s)?

(Please check all that apply):

- ☐ coordinating digital collection/digital library projects
- ☐ creating metadata records
- ☐ supervising metadata and/or cataloguing project(s)
- ☐ creating and maintaining controlled vocabularies (lists of subject headings, thesauri, taxonomies, etc.) and authority files
- ☐ consulting on metadata issues
- ☐ consulting on vocabulary control issues
- ☐ providing technical support to the digital library projects
- ☐ teaching and training information professionals
- ☐ developing policies and best practices
- ☐ Other (please specify):

2 What are the major concerns you have in your project(s) that relate to metadata?

a) For design and planning of digital projects

(Please check all that apply to your major concerns)

- ☐ to understand possible workflows
- ☐ to consider reusing existing cataloguing records by integrating them or transforming them to other formats, e.g., MARC to DC, a local format to EAD, etc., or any other variation in the new project
- ☐ to understand the mechanisms of harvesting protocols
- ☐ to explore how to include various types of resources (print, web pages, images, etc.) in one project
- ☐ to plan how search functions can be supported by metadata information
- ☐ to decide upon levels of description (e.g., item level, collection level)
- ☐ to see examples from similar projects
- ☐ to plan how metadata records will be linked with authority records
- ☐ to plan how the metadata describing a physical object will be associated with the metadata for its digital version
- ☐ to find if any metadata exist already in the objects themselves that could be extracted automatically and what tools are available for this
- ☐ to understand the value of controlled vocabularies
- ☐ to understand and adopt an abstract model (e.g., Dublin Core Abstract Model, FRBR conceptual model, CCO entity-relationship model)
- ☐ to understand types of metadata (e.g., descriptive, administrative, structural, preservation, rights metadata)
- ☐ to learn how to measure and control metadata quality
- ☐ Other (please specify):

b) For the decisions about element set standards (data structure decisions)

Note: Examples of metadata standards include Dublin Core, MARC, MODS (Metadata Object Description Schema), VRA (Visual Resources Association) Core, EAD (Encoded Archival Description), CDWA Lite.

(Please check all that apply to your major concerns):

- ☐ to find out what standards are available
- ☐ to understand what factors, influence the decision on which metadata standard to use, e.g., what sort of material they are good for

METADATA DECISION FOR DIGITAL LIBRARIES

Marcia Lei Zeng, Jaesun Lee and Allene F. Hayes

- ☐ to decide which metadata standard to use
- ☐ to understand what sorts of adjustments might be made to a standard metadata schema that could result in a separate schema and /or application profile
- ☐ to decide whether an application profile should be developed
- ☐ to learn how to create crosswalks
- ☐ to learn how to use different metadata schemes together in one project
- ☐ Other (please specify):

c) For the decisions about data contents in a record (data content decision)

(Please check all that apply to your major concerns)

- ☐ to decide which core elements should be included in all records (e.g., is RIGHTS information required), which elements are mandatory, and which are repeatable
- ☐ to decide which elements (e.g., SUBJECT, CREATOR) should use a controlled vocabulary/authority file
- ☐ to provide guides in order to ensure that metadata values will be entered consistently (e.g., for DATE, FORMAT information)
- ☐ to learn how to provide correct information in a record (e.g., where to find TITLE information from a website, what are the IDENTIFIERs, how many IDENTIFIERs should be included, etc.)
- ☐ to find existing data content (i.e., cataloguing) standards and best practice guides (e.g., Anglo-American Cataloguing Rules (AACR), Cataloguing Culture Objects (CCO), Describing Archives: A Content Standard (DACS), etc.)
- ☐ Other (please specify):

d) For the decisions about authority files and controlled vocabularies (data value decision)

(Please check all that apply to your major concerns)

- ☐ to establish our own authority files for names
- ☐ to decide whether to use existing controlled vocabularies or authority files (e.g., LCSH, ULAN (The Union List of Artist Names), LC Authorities)
- ☐ to develop controlled vocabularies (including controlled lists, taxonomies, thesauri, etc.)
- ☐ to maintain our own authority files and controlled vocabularies
- ☐ Other (please specify):

e) For the decisions about metadata encoding (data format/technical interchange decisions)

Note: Metadata records can be represented in many syntax formats such as XML, RDF, HTML/XHTM.

(Please check all that apply):

- ☐ to understand what are the universal or widely used encoding formats
- ☐ to see examples of encoded records
- ☐ to learn about available tools for encoding and converting records
- ☐ Other (please specify):

f) General comments

- ☐ Which of your major concerns were not addressed in this questionnaire?

THANK YOU!

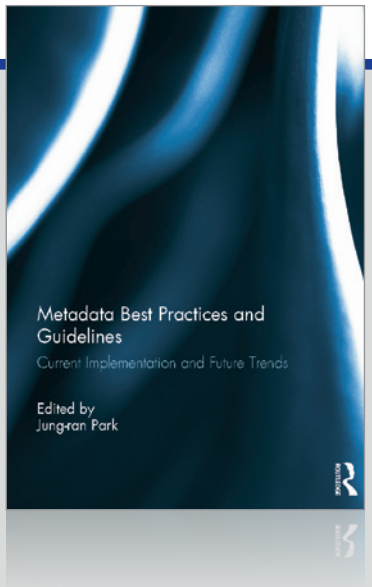
Please send your completed survey back to mzeng@kent.edu or jslee@mail.nl.go.kr.

CHAPTER

6

DOCUMENTING LOCAL PROCEDURES

THE DEVELOPMENT OF STANDARD DIGITIZATION PROCESSES THROUGH THE DEAR COMRADE PROJECT



This chapter is excerpted from

Metadata Best Practices and Guidelines

Edited by Jung-ran Park

LEARN MORE ›

DECUMENTING LOCAL PROCEDURES

THE DEVELOPMENT OF STANDARD DIGITIZATION PROCESSES THROUGH THE DEAR COMRADE PROJECT

Emily Symonds and Cinda May

Excerpted from *Metadata Best Practices and Guidelines*

During the digitization of the Eugene V. Debs correspondence col- lection, known as the Dear Comrade project, at Indiana State University, project staff developed new procedures for assigning and tracking responsibilities. This development was part of an over- all initiative to document local practices and create a more complete workflow process for projects undertaken through the Wabash Valley Visions & Voices Digital Memory Project (WV3) located at http://visions.indstate.edu. Consisting of approximately 6,000 pieces of correspondence written to and from Eugene Victor Debs, the Debs Collection is housed in Special Collections at Cunningham Memorial Library. The new work plan enabled multiple people with varying levels of expertise from different library departments to check the quality of scans and to catalogue and upload items to WV3 using the digital collection management software CONTENTdm.

INTRODUCTION TO THE DEAR COMRADE PROJECT

In early 2008, Wabash Valley Visions & Voices Digital Memory Project (WV3), located at Indiana State University, undertook the scanning and cataloguing of the Eugene Victor Debs correspondence collection (known as the Dear Comrade project) as part of an overall initiative to document local practices and create a more complete workflow process for projects. As a collaborative effort with more than 20 partners across west-central Indiana and with only three full-time staff members at Indiana State University's Cunning- ham Memorial Library for the project, WV3 was accustomed to focusing on repository-based collections, not content-based ones. What made the Dear Comrade project unique fors WV3 was that it was self-contained. We were able to make decisions about the collection prior to digitization and better organize the workflow because we knew exactly what types of materials would be in the collection. One goal in planning was to distribute the work- load so that no single person was responsible for the bulk of the project. There were varying levels of responsibility for creating the digital assets, but we wanted to structure the project so that it could continue at different stages, with either scanning, metadata creation, or record approval, while still ac- counting for the need to tend to other projects and job responsibilities. This goal was tested when the Metadata and Digital Initiatives librarian began a new position in April 2008 and again at the end of the spring semester when student work schedules changed.

Wabash Valley Visions & Voices Digital Memory Project is a collaborative effort to document and preserve in a digital format the history and cultural heritage of west central Indiana. With 24 partners across five counties, the project contains a wide range of images, documents, printed material, maps, oral histories, and audio and

DECUMENTING LOCAL PROCEDURES

THE DEVELOPMENT OF STANDARD DIGITIZATION PROCESSES THROUGH THE DEAR COMRADE PROJECT

Emily Symonds and Cinda May

Excerpted from *Metadata Best Practices and Guidelines*

video files delivered through 36 collections. Institutions that participate in WV3 as full partners pay an annual fee to support the project. In return Digital Initiatives staff based at Cunningham Memorial Library (CML) offer digitization and metadata creation support by request.

Initially each partner received a collection within WV3 to provide access to its holdings. The expansion of the collection structure of Wabash Valley Visions & Voices into more focused areas of interest in 2006 addressed the need to separate large quantities of specific material from more generalized content. The development of topical collections became an option for all WV3 contributors in 2009. This furnishes a means for the partners to provide access to assets created through grant-funded digitization projects in a more cohesive manner, facilitates browsing, and provides more precise access for researchers.

While the research benefits of digitizing the Debs correspondence were apparent, the challenge lay in structuring the digitization process for a complete collection. With other WV3 collections, outside partners provide most of the digital images and metadata, while WV3 staff perform a supervisory role and undertake quality control. Staff do take on the scanning and meta- data creation for some collections, but it is generally a small number of items at a time. In deciding to digitize the Debs correspondence, we took on a large in-house project, in addition to the maintenance of the other collections, with no additional funding, no dedicated staff, and with technological limitations. Procedures established and tested during the Dear Comrade project have been incorporated into the standard WV3 workflow and can be applied to other large-scale digitization projects completed internally and with limited resources.

HISTORICAL BACKGROUND AND PHYSICAL LOCATION OF THE DEBS CORRESPONDENCE

Born November 5, 1855, in Terre Haute, Indiana, Eugene Victor Debs strode upon the national stage during a period in U.S. history defined by the rise of industrial capitalism and pervasive social change. An eloquent orator, Debs championed the working class and the rights of each citizen through his articulation of profound beliefs rooted in democratic thought. Debs worked to organize labour unions, aided in the establishment of the Socialist Party of America, served on numerous editorial boards, spoke and published extensively, and engaged in anti-war activities during World War I for which he was imprisoned and his citizenship was revoked. He ran on the Socialist Party ticket for president no less than five times, the last campaign

DECUMENTING LOCAL PROCEDURES

THE DEVELOPMENT OF STANDARD DIGITIZATION PROCESSES THROUGH THE DEAR COMRADE PROJECT

Emily Symonds and Cinda May

Excerpted from *Metadata Best Practices and Guidelines*

being in 1920, which he conducted from his cell at the federal penitentiary in Atlanta, Georgia. Debs' public life spanned the years from 1875, when he joined the newly formed Vigo Branch of the Brotherhood of Locomotive Firemen, until his death on October 20, 1926. During this period, Debs maintained a vigorous correspondence. Debs' brother, Theodore Debs (1864–1945), served as his personal secretary and periodically purged the correspondence out of necessity for lack of storage space. The remaining correspondence represents the letters deemed important enough to save and those received after the final act of disposal by Theodore.

Marguerite Debs Cooper, daughter of Theodore Debs and niece of Eu- gene Victor Debs, donated her collection of correspondence to Cunningham Memorial Library in 1967. This initial gift of approximately 6,000 letters, telegrams, typescripts, and manuscripts from nearly 1,700 people, including Eugene and Theodore Debs, led to the creation of the Debs Collection in the Special Collections department. The Debs Collection also incorporates the Leslie family gift, consisting of correspondence between family members and Eugene V. Debs, photocopies of Debs' speeches and articles, Little Blue Books, printed texts by notable socialist and labour leaders, and an extensive collection of socialist pamphlets.

In the mid-1970s, J. Robert Constantine and Gail Malmgreen initiated the effort to microfilm the Eugene V. Debs papers, to draw together the Debs canon that resides in multiple institutions and private collections worldwide. Organized into three series, the Microfilm Edition contains correspondence (1834–1945); reminiscences, speeches, and published writings; and scrap- books (1884–1938) on 21 reels of film. Included in the Microfilm Edition is the collection of Debs correspondence housed in the Special Collections department at CML. The editors issued The Papers of Eugene V. Debs 1834–1945: A Guide to the Microfilm Edition in 1983 to facilitate access. The University of Illinois Press published Letters of Eugene V. Debs, edited by Constantine in 1990. This three-volume set contains approximately 1,500 letters extant in multiple collections and incorporates a selection of the Debs correspondence held in Special Collections.

WHY DIGITIZE THE EUGENE V. DEBS CORRESPONDENCE?

Considering that a significant portion of the extant Debs correspondence is available through microfilm or in print format, the question of why digitize the letters naturally arises. Using new technologies to facilitate research and improve access was a primary consideration in the decision to create digital surrogates and make them

DECUMENTING LOCAL PROCEDURES

THE DEVELOPMENT OF STANDARD DIGITIZATION PROCESSES THROUGH THE DEAR COMRADE PROJECT

Emily Symonds and Cinda May

available through Wabash Valley Visions & Voices Digital Memory Project (WV3). Another reason to create digital assets stemmed from the desire to add appropriate material to WV3 which already included items related to Eugene V. Debs, the Labour Movement, and the Socialist Party of America, as well as digital representations of the titles in the socialist pamphlet collection maintained by Special Collections and available through the department's Web site and the Library's online catalogue.

The "Red Special" postcards written by Debs during his 1908 transcontinental presidential campaign journey were scanned in 2004, along with photographs and pamphlets authored by Debs and added to the Cunning- ham Memorial Library Collection in WV3 as its first project. The proposed digitization of the Debs correspondence complements the writings and images of Debs already available, renders all assets searchable, and enhances the Special Collections Web site that offers historical background, collection indexes, and abstracts to the letters.

The home of Eugene V. Debs, which stands on the campus of Indiana State University, is owned and operated as a museum by the Debs Foundation. The museum is a founding partner of Wabash Valley Visions & Voices, and its WV3 collection furnishes access to nearly 200 digital photographs of personal artefacts and memorabilia. The Vigo County Historical Society Collection also contains Debs material. Bringing together all articles relating to Eugene V. Debs housed in diverse physical collections through WV3 and providing cross-collection search capabilities for researchers was deemed highly desirable. It also had the added benefit of documenting the life and career of a famous native son.

COPYRIGHT CONSIDERATIONS

The Special Collections department at Cunningham Memorial Library holds the copyright to all of the Debs family correspondence and images in the collection. The letters in the Leslie family collection formerly restricted by request were scanned for preservation purposes only and will be made available following the completion of the main body of correspondence. A good faith effort to identify copyright holders of the letters written to Debs is made by Digital Initiatives staff at CML, but many individuals remain unknown. The library welcomes additional information in this area. Considering that the majority of the correspondence has been published in other formats, it is not likely that the digital surrogates will have a negative impact on a commercial product. The printed material scanned and presented is in the public domain. The artefacts are the property of the Debs Foundation, which retains the copyright to the digital assets

DECUMENTING LOCAL PROCEDURES

THE DEVELOPMENT OF STANDARD DIGITIZATION PROCESSES THROUGH THE DEAR COMRADE PROJECT

Emily Symonds and Cinda May

but agreed to make these proxies freely available through the Internet. This is true of all items in Wabash Valley Visions & Voices. The project provides a blanket copyright statement that describes the usage rights for all collections (see http://visions.indstate.edu/visions/documentation.html). The metadata record for each item carries a copyright statement that refers back to the col- lection owner. Contributors to WV3 agree to manage copyright for the assets in their respective collections and to obtain permission from the copyright holder prior to submitting items to the repository. All digital assets are freely provided for educational and research purposes as defined by Section 107 of the U.S. Copyright Law.

DETERMINING STAFF RESPONSIBILITIES

When the Dear Comrade project was initiated at the beginning of 2008, the Digital Initiatives unit at Cunningham Memorial Library had three-full time employees, consisting of two librarians and one library associate at the rank of Library Associate III, and two student assistants who worked approximately 20 hours per week. From January to May 2008, the department also had a public history intern who could assist with the project. The physical collection of Debs correspondence is housed in Special Collections, which was solicited to assist with the scanning and metadata creation. During the period when the project was initiated and the majority of the scanning was performed (January to July 2008), the head of Digital Initiatives was also serving as the acting head of Special Collections due to a sabbatical leave. This joint role allowed for more communication between the departments as one person was ultimately in charge of everyone involved with the project. The cooperation with Special Collections provided more available staff for the project—one library associate at the rank of Library Associate IV and two student assistants who worked approximately 20 hours per week.

Weekly meetings were scheduled between the departments to discuss updates and issues. Riley and Whitsall (2005) stated that "digitizing staff in the University environment, who are often student employees, should be trained to understand enough about the digitization process to be able to catch many quality problems as they happen" (p. 40). Our weekly meetings enabled the student assistants and the intern to understand the entire process for the Dear Comrade project and not just the portion they were scanning or cataloguing. The meetings also allowed staff to discuss new issues and find solutions. For example, it was during these meetings that we decided to preselect relevant subject terms from the Library of Congress Subject Headings and how to arrange the digital items in CONTENTdm. This process followed

DECUMENTING LOCAL PROCEDURES

THE DEVELOPMENT OF STANDARD DIGITIZATION PROCESSES THROUGH THE DEAR COMRADE PROJECT

Emily Symonds and Cinda May

Excerpted from *Metadata Best Practices and Guidelines*

the recommendation of D'Andrea and (Martin, 2001), who wrote, "Understanding where a job, however tedious, fits into the grand scheme will be a benefit for students, allowing them to feel less like simple cogs and more like real contributors" (p. 23). Students divided their time between the Dear Comrade project and other assignments. From February to August 2008, one student employee in Special Collections worked 20 hours a week and focused primarily on scanning Debs materials. A second student scanned approximately 5 hours a week from February to April 2008 and from August 2008 to February 2009. The two WV3 student employees each scanned for the Dear Comrade project approximately 10 hours per week from April 2008 until September 2008. The public history intern was with WV3 from January to May 2008 but was not involved in the scanning for this project.

PLANNING THE DIGITIZATION PROCESS

With more than 6,000 letters, none of which were available in a digital for- mat, the initial goal was to scan the letters in the Debs collection before focusing on describing the materials and uploading them to the WV3 CONTENTdm site at http://visions.indstate.edu. We decided to keep the materials in Special Collections and have the student workers in Special Collections scan the collection when not performing their regular duties. This required additional training since Special Collections had different scanning specifications than Digital Initiatives. Because the digital collection would be part of Wabash Valley Visions & Voices, procedures followed WV3 guidelines, which are based on the State's Indiana Memory standards. All documentation is available at http://visions.indstate.edu/visions/documentation.html. A scanning station was already set up near the site of the Debs collection and only required software updates for Adobe Photoshop and installation of CONTENTdm's Acquisition Station. A second scanning station was set up in the same room shortly after the project began. Having the scanning take place in Special Collections also allowed the students to confer more easily with that department's library associate regarding questions about the condition of the physical materials. For materials that were deemed too fragile for handling and scanning, the photocopy, which was already in the collection, was scanned instead of the original. This was noted for the relevant items on the production spreadsheet, which will be discussed later.

The correspondence in the Debs collection was arranged alphabetically in folders based on the correspondent's last name. The student employees in Special Collections were assigned alternating letters in the alphabet so that one student scanned the letters in A and C, while the second student scanned the letters in B and

DECUMENTING LOCAL PROCEDURES

THE DEVELOPMENT OF STANDARD DIGITIZATION PROCESSES THROUGH THE DEAR COMRADE PROJECT

Emily Symonds and Cinda May

Excerpted from *Metadata Best Practices and Guidelines*

D, and so on. This was decided so that each student's duties were self-contained. They did not need to worry about what stage the previous student had stopped, and students could work concurrently when their schedules permitted. This same system was used when we began the metadata creation.

During the scanning, the students recorded in a Microsoft Excel production spreadsheet each page scanned, the date scanned, the scanner used (as the stations had different scanner models), file size, and other metadata, such as if the photocopy had been scanned. Students also entered their initials to indicate which scans they had done. Once the second scanning station was set up, the students often scanned during the same period. It was not possible to create a shared document that users could access simultaneously; therefore, the students saved the metadata to a desktop copy of the production spreadsheet and then copied and pasted the information to the master production spreadsheet at the end of their shifts.

SCANNING AND NAMING THE FILES

For files within a collection, Wabash Valley Visions & Voices already had a system using a three- or four-letter collection code and a five-digit number starting with 00001. This system was adjusted even before the Dear Comrade project in order to avoid the risk of duplicating numbers when two people were scanning and naming materials from the same collection and in order to align more closely with the standards established by Indiana Memory for Item ID, which corresponded to the file name. Indiana Memory's metadata standards (2007a) required participants to use the sequence program name- institution name-collection identifier-item number descriptor to name files (p. 4). WV3 adapted its Item ID standards to collection ID-project name-item number descriptor.file extension. The collection ID was the same three- or four-letter collection code established previously. The project name referred to the specific portion of the collection being digitized. The item number was a five-digit number starting with 00001. The descriptor was used for compound objects to indicate the title page or page number.

For the Dear Comrade project, the three-letter collection code was evd. Each project name started with letters to indicate these files were correspondence, in case other materials, such as the Debs pamphlets, were later included in the collection. Following letters was a, b, c, and so on to indicate the alphabetical section of the correspondence being digitized. For example, the file names would be assigned as evd-lettersa-00001 or evd-lettersb-00001 and continue from there. This was another

DECUMENTING LOCAL PROCEDURES

THE DEVELOPMENT OF STANDARD DIGITIZATION PROCESSES THROUGH THE DEAR COMRADE PROJECT

Emily Symonds and Cinda May

Excerpted from *Metadata Best Practices and Guidelines*

method that allowed multiple people to work on distinct portions of the collection without the risk of duplicating file names or wondering at what number the last person had left off. While the collection code was always evd and the item number al- ways began with 00001, the differences in the project name allowed project participants to establish distinct item IDs independently. This pattern for file names has become part of WV3's standards for its collections.

Items were scanned at a resolution according to WV3 guidelines, which were based on Indiana Memory guidelines. Each scan was saved at a minimum resolution of 300 ppi. Initially, students were told to save each scan to WV3's Production server and the Master TIFF server, the second of which served as dark storage for archival copies of the scans. Within the servers, folders were created in a hierarchical system starting with the letter of the alphabet, then the name of the original folder, which was the correspondent's name, and then a folder for each piece of correspondence:

[D]
[Debs, Eugene V.]
[1895-06-16]
[1895-06-18]

Due to an unidentifiable network issue, however, the students were unable to access the Master TIFF server from the Special Collections scanning stations. Instead, they saved just the one copy to the Production server. During the times when students were not scanning or using the production spreadsheet, the Metadata and Digital Initiatives librarian would copy the scans from the Production server to the Master TIFF server in order to provide two, high-resolution copies of each file. The copies in dark storage were not to be altered but were part of the Digital Initiatives unit's preservation plan. Although copying a file instead of creating two original files is not ideal, it was the easiest and most efficient solution at the time.

While quality control was always part of the project process, the necessity of the Metadata and Digital Initiatives librarian copying the TIFF files integrated the quality-control process more fully into the project procedures. We depended on visual checks instead of automated checks to confirm that files met project specifications (Riley and Whitsall, 2005, p. 40). The Metadata and Digital Initiatives librarian performed these checks daily and was able to gauge progress as well as note if any files were scanned at the incorrect resolution or if pages were skipped. After copies were saved to the Master TIFF server, the Metadata and Digital Initiatives librarian corrected the Production TIFF files. The information from the production spreadsheet was copied to

DECUMENTING LOCAL PROCEDURES

THE DEVELOPMENT OF STANDARD DIGITIZATION PROCESSES THROUGH THE DEAR COMRADE PROJECT

Emily Symonds and Cinda May

Excerpted from *Metadata Best Practices and Guidelines*

a metadata spreadsheet, which would be used to record who described and uploaded which files to the CONTENTdm site. Hull and Dreher (2001) discussed the challenge faced when creating a larger and more effective tracking system after working with multiple departments with their own systems for documenting what was complete and what was still in progress (p. 33). Although maintaining two spreadsheets for the Dear Comrade project was cumbersome and did result in the duplication of information across files, it also allowed more people to access the data and served to document the status of a project that was divided among staff, offices, and departments. Responsibility for quality control, file copying, and corrections to the scans was later taken on by the Digital Initiatives library associate, who continues to be in charge of this portion of the project.

STRUCTURING THE DIGITAL COLLECTION

During the scanning process, staff from Digital Initiatives discussed the best way to organize the digital materials and upload them to the Eugene V. Debs correspondence collection that had been created within the Wabash Valley Visions & Voices site. The physical collection was arranged in folders alphabetically by the correspondent's last name. Staff weighed the benefits of reproducing the physical structure in a digital collection by replicating the folders with the ease of describing and uploading items quickly. CONTENTdm does allow users to upload compound objects which are made up of more than one connected file, such as pages from a book. Staff knew that the collection would consist of compound objects but had not deter- mined if the uploads would take place at the item level as multiple letters or, at the collection level, as multiple folders containing the letters. We considered creating one record or compound object that would contain all the individual letters in a hierarchical arrangement. The highest level would be the name of the folder, replicated from the physical folders, and the lower levels would be the individual pieces of correspondence. For example, a record with the title "Abbot, Leonard" would exist, and within that com- pound object would be all the letters and telegrams to or from Leonard Abbott, mirroring the arrangement of the physical collection. In test uploads, however, this proved to be a time-consuming process.

While a new project can be a good opportunity to develop skills, it is best to do the research and training before initiating the project. In this case, we had proceeded far enough in one direction that it made little sense to change our approach at that point. Folders on the Production server were already created in a way that allowed for compound objects at the item level. Files would have to be moved and reorganized,

DECUMENTING LOCAL PROCEDURES

THE DEVELOPMENT OF STANDARD DIGITIZATION PROCESSES THROUGH THE DEAR COMRADE PROJECT

Emily Symonds and Cinda May

Excerpted from *Metadata Best Practices and Guidelines*

which could have been a long and risky process. This process also required more high-level experience with CONTENTdm than the project participants generally had. Each person creating metadata was responsible for uploading those items to CONTENTdm Administration for approval. By uploading compound objects at the item instead of folder level, all participating staff could work with the existing file structure. This was an instance when deciding all organizational elements before starting scanning would have been ideal, but in the end, uploading the correspondence as individual items was easier for project staff and enabled better searching of the collection by users. In addition to being more feasible to set up on the production side, it also meant that users could browse the collection by letter and not just by correspondent. Staff members creating the records were also able to exercise more autonomy and gain more experience by working with multiple smaller records instead of a small group of larger records.

CREATING METADATA

As the scanning progressed, we decided to initiate the metadata creation process. The Special Collections library associate and the public history intern in Digital Initiatives were the first project participants to describe and submit items from the Debs correspondence to the CONTENTdm system. As the project continued, the Digital Initiatives library associate became involved in metadata creation and record approval. Just as students doing scanning were assigned a letter of the alphabet, so were those describing items. Instead of having more people juggle access to the same production spreadsheet, a new metadata spreadsheet was created. This contained the information from the production spreadsheet, updates by the Metadata and Digital Initiatives librarian, and a field for additional notes. As items were described and uploaded, participants recorded the date and their initials as an external way to track which items were completed.

Before the library associate and public history intern could describe and upload materials, however, project staff had to determine what metadata elements to use. Ma (2006) noted that "the analysis of metadata requirements should take into account the goals and objectives of the project, the nature of the collection, the level of granularity that has to be supported, the formats of the materials, and user expectations and needs" (p. 7). For the Dear Comrade project, we already decided to catalogue at the item level and not the folder level. We wanted the collection to be searchable by name, to include a description of the content of each letter, and to correspond with the field properties of the other WV3 collections. This last point would make it easier for users to

DECUMENTING LOCAL PROCEDURES

THE DEVELOPMENT OF STANDARD DIGITIZATION PROCESSES THROUGH THE DEAR COMRADE PROJECT

Emily Symonds and Cinda May

Excerpted from *Metadata Best Practices and Guidelines*

search across collections. We decided not to provide full-text searching for the collection. Optical Character Recognition (OCR) would not be effective for the letters that were handwritten, and we did not have the additional staff or time to transcribe the correspondence ourselves. Following those decisions, we mapped the project's metadata elements to the corresponding Dublin Core elements and set up the collection and field properties in CONTENTdm Administration.

Similar to Hull and Dreher (2001), whose project created a master template for cataloguing by non-cataloguers (p. 32), the head of Digital Initiatives and the Metadata and Digital Initiatives librarian developed a metadata template for the Debs collection. Our template documented the procedures, guidelines, and recommendations for each project field in the collection, ensuring more consistency across the records. Project employees could make use of the Template Creator in the CONTENTdm Acquisition Station for fields that would not change from record to record. In most cases, the metadata requirements followed the standard WV3 procedures with a few exceptions. Local fields were mapped to Dublin Core elements.

Most important was the title element since that is the default field for indexing and sorting items in a CONTENTdm collection. To make browsing easier for users, we recorded the title using the correspondent's name from the physical folder, maintaining the last name–first name order to keep like items together. To distinguish multiple letters from the same person, we entered the date of the letter.

Instead of transcribing correspondence or creating our own descriptions, we referred to abstracts that were already created in Special Collections and are available at http://library.indstate.edu/about/units/rbsc/debs/ abstract.html. Ma (2006) recommends using extant metadata when possible "to utilize available resources and to avoid reinventing the wheel. ... If some kind of metadata exists and can be utilized, it is necessary to extract the existing data to an appropriate destination" (p. 10). On-demand searching of the abstracts was possible through an older version of a free-text askSam database, and HTML and text files of the abstracts were accessible; however, there was no easy way for us to extract this data for the CONTENTdm records. One issue was that correspondent, date of the letter, and the abstracts were in a single field, which contributed to difficulties when we attempted to export the data into separate fields. The result was that project staff had to copy and paste each abstract into the Media Editor in the CONTENTdm Acquisition Station. Dates were standardized by spelling out the month and using the day and four-digit year. Abbreviations were spelled out, especially for names. The abstracts served as the

DECUMENTING LOCAL PROCEDURES

THE DEVELOPMENT OF STANDARD DIGITIZATION PROCESSES THROUGH THE DEAR COMRADE PROJECT

Emily Symonds and Cinda May

Excerpted from *Metadata Best Practices and Guidelines*

description for each item. Some employees found having to refer to the abstracts, spreadsheet, and metadata template awkward. While this could be mediated by using computer stations with dual monitors, some of which are available in the department, it could not be completely avoided. Not having the abstracts in a format that allowed for easy exporting did add an extra layer of work to the project, but it was preferable to having to create new abstracts or descriptions for each piece of correspondence. This is one of the unfortunate side effects when relying on metadata from external departments, even when those departments are in the same library.

The head of Digital Initiatives and the Metadata and Digital Initiatives librarian undertook a subject analysis of the collection based on the database of abstracts and created an authority list of authors and recipients. As new names were discovered, they were submitted for approval to the lists. By using a controlled vocabulary for these two fields, we were able to ensure that the names were entered consistently, using the most complete form or the authorized form from the Library of Congress Name Authority File. Similarly, we analysed the abstracts to create a list of authorized subject headings. We did this instead of loading the standard Library of Congress Thesaurus of Graphic Materials because it contained many unnecessary terms for the col- lection and did not include Library of Congress Subject Headings that better captured the subject of the materials. A predetermined list of terms fit the content of the collection better and reduced the number of terms from which participants had to choose. The Digital Initiatives and Special Collections library associates also provided subject analysis and suggested authorized terms or relevant topics that could be matched to subject headings.

One major difference in this collection was the use of the Coverage field. Ordinarily, Wabash Valley Visions & Voices requires that Coverage note the location from which an item originated and its original date. In the case of the Debs correspondence, this would be the location from which the letter was written; however, in many cases, we did not have that information because it was not contained in the letter and the original envelopes were not part of the collection. Ultimately, we created this as a uniform field for the collection, with each item containing the general coverage scope of the collection. The location was United States—Indiana—Terre Haute, because that was the location of the collection, and the date was 1874–1977, which covered the range of all dated correspondence in the collection. Coverage was not required by Indiana Memory in its 2007 guide. This meant there were no higher-level standards to follow, only WV3's practices from earlier documentation.

DECUMENTING LOCAL PROCEDURES

THE DEVELOPMENT OF STANDARD DIGITIZATION PROCESSES THROUGH THE DEAR COMRADE PROJECT

Emily Symonds and Cinda May

Excerpted from *Metadata Best Practices and Guidelines*

An additional element that was only part of WV3's recommendations is WV3 Subject. This field has a local controlled vocabulary based, in part, on topics recommended by the Indiana Digital Library Summit. We recommended the terms "Famous Hoosiers" and "Labour" for all items in the Dear Comrade project.

To ensure consistency, with so many people working on the collection, other fields that normally would not use a controlled vocabulary were set up with one. These fields included Copyright, Repository, and Provenance. Copyright and Repository were also two fields in which information could be entered in the Template Creator through CONTENTdm. As new employees joined the project and others left, we could be sure that metadata was entered uniformly into these fields. The use of authorized terms for these fields also meant that project participants could concentrate more on analysing subjects and expanding the abstracts in order to provide better access to users.

The Key field is a hidden, unmapped field initiated by Wabash Valley Visions & Voices as part of O Miners Awake: Indiana Coal Miners, Their Families, and Their Communities project. We use the Key field to identify related materials across collections that may not contain identical fields. This allows for the grouping of like items and simplifies the custom-query process. For the Debs project, we already had the Eugene V. Debs Museum Collection, as well as some Debs postcards in the Indiana State University Library Col- lection. The term "debskey," based on the "coalkey" term established for O Miners Awake, would not show up in any other field or in any non-Debs related collection. The Key field enables us to create custom queries in CONTENTdm and display together materials across collections, as can be seen at the O Miners Awake Web site at http://visions.indstate.edu/coal.php.

Once the template was established, project participants began creating metadata and uploading items through CONTENTdm. During this period, Wabash Valley Visions & Voices was using CONTENTdm 4.2 and not the more recent 4.3. WV3 had decided not to upgrade while in the middle of providing support to the Knox County Public Library and its LSTA-funded Early Vincennes project. When undertaking Dear Comrade we decided that version 4.3 did not offer features that affected our planning. The new version did allow controlled vocabularies to be shared across collections, but the Dear Comrade controlled vocabularies were specialized for that project and would not have corresponded closely enough to the vocabularies in other collections to affect the endeavour.

DECUMENTING LOCAL PROCEDURES

THE DEVELOPMENT OF STANDARD DIGITIZATION PROCESSES THROUGH THE DEAR COMRADE PROJECT

Emily Symonds and Cinda May

Excerpted from *Metadata Best Practices and Guidelines*

PROJECT STATUS

The overall time estimates for completion of the Dear Comrade project were based on the following assumptions:

1. Special Collections supplied 6,000 items of correspondence, some consisting of two or more pages, for a total of approximately 13,500 individual scans.
2. Figuring 12 scans per hour at a rate of 10 hours per week required 112.5 weeks or 2.16 years to complete the scanning.
3. Allowing 1 hour per item to produce metadata and perform quality-control checks at a rate of 30 hours per week required 200 weeks or 3.85 years to render the digital assets. At present, 960 letters are available in Wabash Valley Visions & Voices, which represents 16% of the collection.

Four student assistants and a public history intern worked on the digitization of the Debs correspondence from January 31, 2008, to February 16, 2009, with the majority of the scanning being completed within the first seven months. The original goal to scan all the correspondence within a six-month period proved impractical as the budget lacked sufficient funds to employ four students for 20 hours per week. The public history intern graduated in May 2008 and spent a significant portion of her time working on unrelated assignments. The reduced availability of student employees during the summer months and the loss of one student at the end of August 2008 also influenced the number of hours committed to the digitization process. In addition, other work assignments within the Special Collections department and the Digital Initiatives unit impacted production. Occasionally during the first three months of the project, all four students scanned correspondence concurrently, but more frequently only one or two students scanned the letters on a weekly basis for more than 10 hours. The two student assistants assigned to WV3 ceased scanning activities in mid-September to address the digitization needs of the WV3 partners. Only one student in Special Collections continued to scan letters, but at a rate of less than 10 hours per week. In March 2009, the University Digital & Archival Services Department moved into a new space that included three networked digitization workstations. Shortly thereafter, the Wabash Valley Visions & Voices project migrated to a new server and upgraded first to CONTENTdm 5.0 and, subsequently, to 5.1.

Due to an influx of special-projects funding, a student assistant was hired in May 2009 to work an average of 30 hours per week for 12 weeks specifically to scan the Debs correspondence. At this time, all digitization of the Debs correspondence moved to the Digital Initiatives lab and folders of letters were transported from and returned

DECUMENTING LOCAL PROCEDURES

THE DEVELOPMENT OF STANDARD DIGITIZATION PROCESSES THROUGH THE DEAR COMRADE PROJECT

Emily Symonds and Cinda May

Excerpted from *Metadata Best Practices and Guidelines*

to Special Collections on a routine basis. The spreadsheet created to track progress was copied over to the new server. The student assistant recorded production data as he scanned the letters and saved the files to both the new production server and the server housing the archival master copies. The library associate continued to perform quality checks of the engendered images. The entire collection, including the Leslie family correspondence, was scanned. The final count of individual images equalled 9,475, a figure significantly lower than the initial estimate. The digitization phase of the project was completed in August; however, software bugs related to the creation of compound objects in CONTENTdm 5.0 halted metadata production as many of the remaining letters comprised multiple pages.

Over the course of the Dear Comrade project, the need to create digital assets for WV3 partners siphoned time away from the Debs correspondence in terms of scanning, metadata creation, and quality control. The departure of the Metadata and Digital Initiatives librarian necessitated a redistribution of duties and transferred the responsibility for quality control and the supervision of WV3 student assistants to the Digital Initiatives library associate. The diversion of this associate's time into quality control resulted in a significant curtailment in the number of digital assets rendered. The meta- data developed by the Special Collections library associate required careful review as the individual lacked expertise in subject analysis and failed to make consistent use of the available controlled vocabulary. However, the Special Collections associate's knowledge of the Debs Collection facilitated the identification of photographs removed from the correspondence and permitted staff to reunite them with the appropriate letters.

Two months into the project, the Library Administration merged the Digital Initiatives unit with University Archives to establish University Digital & Archival Services. The merger added more management tasks to the department head, splitting her time between two units located in different buildings. This exacerbated the staffing situation as the head possessed even less time to devote to overseeing the project. Furthermore, the Library's decision to move forward with the creation of an institutional repository led to the redefinition of the vacant librarian position transforming it into a Digital Repository Librarian position with the responsibilities focused on open source development and customization rather than metadata creation for Wabash Valley Visions & Voices, although the revised position continued to offer technical support for WV3.

Shifting priorities in Special Collections also reduced the number of hours available for metadata creation. The combined decrease in hours committed to the project in

DECUMENTING LOCAL PROCEDURES

THE DEVELOPMENT OF STANDARD DIGITIZATION PROCESSES THROUGH THE DEAR COMRADE PROJECT

Emily Symonds and Cinda May

Excerpted from *Metadata Best Practices and Guidelines*

both departments constrained production and hampered progress. The Digital Initiatives library associate continues the quality control and supervision of the unit's students. A biweekly temporary staff position is being added to the department to address the need for metadata creation support. The incumbent will devote a minimum of 20 hours per week to the project beginning in mid-September 2009, and it is estimated that an additional 720 digital assets will be rendered accessible by July 1, 2010, for a total of 1680 letters, or 28%, of the correspondence.

The upgrade to CONTENTdm 5.0/5.1 required staff instruction in the use of the Project Client that replaced the Acquisition Station. The Project Client allows users to share projects, which could possibly eliminate the need for an external spreadsheet. By sharing access, participants can see what items have been described in a project just by looking at the Project Client. Dividing the metadata creation into portions for each user may no longer be necessary. This has the potential to help streamline the entire process. While the improved functionality brought benefits to the project, the need to train personnel impacted production. Testing and stabilization of the application remained an abiding issue, but the general workflow procedures transferred well and continued to be followed, with slight modifications, to account for the new server and the removal of the digitization from Special Collections to the Digital & Archival Services.

Like the other WV3 collections, the Eugene V. Debs Correspondence col- lection has its own project page (http://visions.indstate.edu/visions/partners/evdc.html) that provides a description of the collection, contact information, and links to browse or search the collection. When this collection is complete and debskey is added to the Key field for the Debs materials in other collections, WV3 will provide a link to a custom query that will bring together all relevant items and complete the Dear Comrade project.

LESSONS LEARNED

The local practices established through this initiative can serve as guidelines for other institutions, especially those working collaboratively across locations or on multiple, ongoing projects. The key elements taken from the Dear Comrade project are the following: planning the entire project, not just one aspect of it, from the start; having weekly (or regular) meetings with all participants, including student employees; standardizing file names to avoid potential duplication and to allow more than one person to name files concurrently; documenting the scanning and metadata

DECUMENTING LOCAL PROCEDURES

THE DEVELOPMENT OF STANDARD DIGITIZATION PROCESSES THROUGH THE DEAR COMRADE PROJECT

Emily Symonds and Cinda May

Excerpted from *Metadata Best Practices and Guidelines*

creation processes through spreadsheets; selecting from a controlled vocabulary the terms relevant to the project ahead of time; and developing a metadata template to serve as a "cheatsheet" for participants creating metadata in order to reduce the risk of skipped fields or improperly assigned terms and headings. Setting fields up as controlled vocabulary fields within CONTENTdm did not follow the protocol for controlled vocabularies but did allow multiple metadata creators to select the exact term for those fields and to focus more on the description and subject for each item. For custom queries and connecting items across collections, the Dear Comrade project followed the practice established earlier for O Miners Awake and made use of a hidden, unmapped field that was assigned a specific term.

The Dear Comrade project helped emphasize the importance of pre-project planning for the Wabash Valley Visions & Voices. While this lesson may seem obvious, when schedules are full and resources limited, the initial instinct can be to jump into a new project in order to stay ahead or to focus on the first part, such as scanning, and make decisions that can have a negative impact on future parts. For CONTENTdm users, how the files are saved and grouped together can have a large influence on how the items are uploaded and displayed. Advance work of setting up computer folders to lay out the structure of the project will save time and trouble when the items are ready to be uploaded. While there will always be items to add to the partners' collections a few records at a time, for internal, long-term projects, having a plan before any scanning or metadata creation is done is key. By addressing questions such as who would be doing the work, where the work would be done, how and where the files would be saved, what metadata was needed for each item, and where that metadata could be found, we were better able to incorporate this project into the regular workflow. We were also able to eliminate confusion over who was working on what portion of the project and avoid duplicating work. The procedures we established can be applied to future special projects. The metadata template was just an example of a specific application of the WV3 digital guidelines, but it helped participants to be consistent in their metadata creation across the collection. This template, as well as the use of project-specific controlled vocabularies, can easily be adapted for other projects and can possibly be applied to current collections that contain a large amount of similar materials or items from the same source.

The experience gained through the Dear Comrade project will serve as the foundation for the ingestation of digital assets from the Newport Chemical Depot Archive into Wabash Valley Visions & Voices, a joint venture currently under discussion. Located in Vermillion County, Indiana, the New- port Chemical Depot produced the entire United

DECUMENTING LOCAL PROCEDURES

THE DEVELOPMENT OF STANDARD DIGITIZATION PROCESSES THROUGH THE DEAR COMRADE PROJECT

Emily Symonds and Cinda May

Excerpted from *Metadata Best Practices and Guidelines*

States stockpile of VX nerve agent from 1962 to 1968. The reserve was destroyed and the facility closed between 2005 and 2008. As the major employer in the county, the Newport Chemical Depot played a significant socioeconomic role in the lives of area residents. The U.S. Army Chemical Materials Agency (CMA) Public Affairs Office engaged an outside contractor to highlight and preserve the history of the depot as depicted through photographs, printed publications, and video. The contractor is working with WV3 to provide access to this rich collection. The Newport Chemical Depot Archive will be treated as a special collection, with a tailored metadata template, controlled vocabularies for repeatable fields, and a specific list of subject terms based on LCSH. Selection, digitization, and metadata creation will all be performed by the contractor according to specifications and archival standards governing WV3. Communication between the two entities will be crucial to the successful completion of this project. The Newport Chemical Depot Archive project will require significant up-front planning time and training of contractor personnel, but the majority of the work will be performed by the employees of the contractor with minimal oversight on the part of the Digital Initiatives unit except for final review and approval of the collection.

Wabash Valley Visions & Voices is built on collaboration and partner- ships between multiple organizations. With the Dear Comrade project, we were able to collaborate successfully across library departments and help provide experience to and develop new skills for student employees and full-time staff. We designed procedures that could be incorporated more fully into the department's workflow when dealing with content-based col- lections and even with the continuing repository-based ones. This project also emphasized the importance of being flexible. As the project progressed, staff decided to collapse the separate production and metadata spreadsheets into one master spreadsheet. These documents are helpful to the project only if people use them. In addition, WV3 partners have recently expressed interest in creating digital assets locally. The establishment of a metadata template and shared and controlled vocabularies facilitates training, promotes consistency, and aids in quality control for the overall repository.

DECUMENTING LOCAL PROCEDURES

THE DEVELOPMENT OF STANDARD DIGITIZATION PROCESSES THROUGH THE DEAR COMRADE PROJECT

Emily Symonds and Cinda May

Excerpted from *Metadata Best Practices and Guidelines*

REFERENCES

- D'Andrea, P., & Martin, K. (2001). Careful considerations: Planning and managing digitization projects. *Collection Management*, 26(3), 15–28.
- Hull, R. W., & Dreher, S. L. (2001). "Into the middle of the thing" (with apologies to Horace): Developing a system to manage a grant-funded digital collection project. *Collection Management*, 26(3), 29–38.
- Indiana Memory. (2007a). *Dublin Core metadata guide Indiana Memory Project* [PDF document]. Retrieved January 30, 2009, from http://www.in.gov/library/files/dig_metast.pdf
- Indiana Memory. (2007b). *Proposed digital imaging standards and best practices Indiana Memory and LSTA digitization projects* [PDF document]. Retrieved January 30, 2009, from http://www.in.gov/library/files/digimgst.pdf
- Ma, J. (2006). Managing metadata for digital projects. *Library Collections, Acquisitions, & Technical Services*, 30(1/2), 3–17.
- Riley, J., & Whitsall, K. (2005). Practical quality control procedures for digital imaging projects. *OCLC Systems & Services*, 21(1), 40–48.

APPENDIX

Template for Dear Comrade metadata

- Title: [Folder Name] [Date of Letter in YYYY-MM-DD] (if no date, use n.d.)
  - Examples: Abbate, Frank 1926-10-23
  - Abercrombie, J. H., n.d.
- Abstract: Copy and paste from online Abstracts of Correspondence from Special Collections at http://library.indstate.edu/about/units/rbsc/debs/abstract.html. Spell out abbreviations such as EVD (Eugene V. Debs) and TD (Theodore Debs). For dates, change from MM/DD/YY to MM/DD/YYYY.
- Author: Select from controlled vocabulary. If name is not listed, enter in Last name, First name format and submit for administrator approval.
- Recipient: Select from controlled vocabulary. If name is not listed, enter in Last name, First name format and submit for administrator approval.
- Repository: Repository is the same for all correspondence. Select from controlled vocabulary: Special Collections Department, Indiana State University Library, Terre Haute, Indiana, 47809.

DECUMENTING LOCAL PROCEDURES

THE DEVELOPMENT OF STANDARD DIGITIZATION PROCESSES THROUGH THE DEAR COMRADE PROJECT

Emily Symonds and Cinda May

Excerpted from *Metadata Best Practices and Guidelines*

- Date Digital: Use the date on the production spreadsheet. Enter in YYYY-MM-DD format.
- Date Original: Use the date on the piece of correspondence. Enter in YYYY-MM-DD format. If no date, enter unknown.
- Coverage: Coverage is the same for all correspondence. Select from controlled vocabulary: United States—Indiana—Vigo County—Terre Haute— 1874–1977
- Subject: Select from controlled vocabulary. One term is required; additional terms are optional. Terms may be submitted for administrator approval.
- WV3 Subject: Select from controlled vocabulary. Use Famous Hoosiers and Labour.
- Provenance: Select from controlled vocabulary. Most letters will use "Cooper, Marguerite Debs" unless they are part of the Leslie family gift or if a specific donor is named in the folder.
- Type: Select text from controlled vocabulary.
- Material Type: Select Correspondence from controlled vocabulary.
- Technical Metadata: image/tiff; [name of scanner]; [software with version]; resolution ppi. Separate elements with a semicolon. The scanner used is listed on the production spreadsheet; in most cases, it's the Epson 1640XL. Software is listed on production spreadsheet. In most cases, it's Adobe Photoshop CS2 8.0. Resolution is on production spreadsheet.
    - Example: image/tiff; Epson 1640XL; Adobe Photoshop CS2 8.0; 300 ppi
- Format of original: Identify if the scanned letter is a photocopy of the original.
- Copyright: Select from controlled vocabulary: Digital image (c) 2008, Indiana State University, Terre Haute, Indiana.
- Item ID: List file names separated by a semicolon. If more than 3 to 4 files in a compound object, include the range instead of listing all the files individually.
    - Example: evd-lettersa-00005.tif; evd-lettersa-00006.tif
- Control: Enter your initials and the date you're submitting the metadata in YYYY-MM-DD format.
- Key: Select debskey from controlled vocabulary.