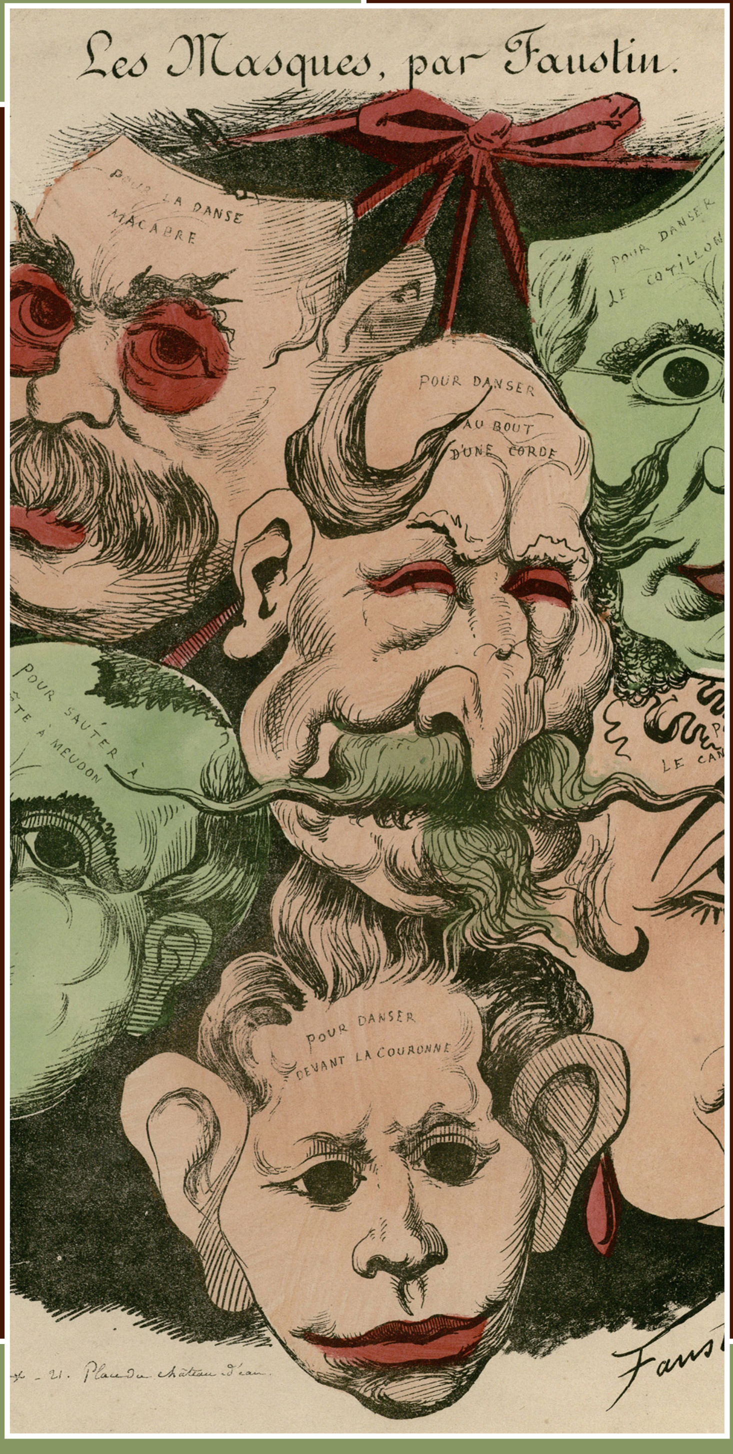
Library Resources & Technical Services

ISSN 2159-9610
April 2016
Volume 60, No. 2



The Association for Library Collections & Technical Services

60 | 2

Library Resources & Technical Services

ISSN 2159-9610 April 2016 Volume 60, No. 2

Cover image: Faustin, 1847–, The Masks, Franco-Prussian War and Paris Commune Caricatures, Special Collections, University of Houston Libraries, accessed January 13, 2016, http://digital.lib.uh.edu/collection/roman/item/664.

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

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

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

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

Association for Library Collections & Technical Services (ALCTS)

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

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

Editorial

Mary Beth Weber

Now that a new year has begun, I look forward to a new round of papers to be submitted to the journal. Each year brings different types of topics, and there are often patterns. In past years, there were more papers on a given topic, such as RDA or patron driven acquisitions. Metrics and collection analysis and digital preservation are current topics of interest.

Papers make their way to *LRTS* in a variety of ways. Some papers are unsolicited, and are submitted by authors who see the journal's scope and reputation as a good match for them. I solicit many papers by monitoring discussion lists and by reading conference programs. The *LRTS* website includes guidelines for turning a presentation into a paper (see www.ala.org/alcts/resources/lrts/authtips). I also make contact with potential authors through ALCTS activities that include the ALCTS New Members Interest Group and ALCTS 101. Members of the editorial board (current and former) also solicit content that has led to excellent papers.

Not every author who submits a paper to *LRTS* is a seasoned professional with a long list of publications. *LRTS* authors include first time authors, who may submit a paper as a single author or collaborate with experienced colleagues. An expert editorial board that represents all of ALCTS' sections reviews submissions. They provide feedback that enables authors to revise and refine their papers. A great deal of time and effort goes into reviews and feedback.

LRTS accepts two different types of papers, research papers and "Notes on Operations," which explore operational issues with value and implications for other libraries. Published "Notes on Operations" papers include "E-book Cataloging Workflows at Oregon State University" by Richard Sapon-White (v. 58, no. 2), "Metadata Makeover: Transforming MARC Records Using XSLT" by Violeta Ilik, Jessica Storlien, and Joseph Olivarez (v. 58, no. 3), and "When One Plus One Remains One: The Challenges and Triumphs of Merging Two University Libraries" by Elaine Mael (v. 58, no. 4). Both types of papers undergo a rigorous double-blind peer review, and include the following elements of a research paper: research method, results or findings, discussion and/or analysis, and a conclusion.

Book reviews are another way to contribute to *LRTS*. They are a good way for those new to the profession to get started with publishing, and also provide experienced professionals an opportunity to evaluate professional publications. In either case, book reviews provide a valuable service to the profession. Interested individuals may contact *LRTS* Book Review Editor Elyssa Sanner at www.ala.org/alcts/resources/lrts/be_reviewer.

I encourage you to consider submitting a paper to *LRTS* or writing a book review. Publishing is a very rewarding experience, particularly when you get feedback from others regarding your work or when you see your work has been cited elsewhere.

In closing, I would like to highlight the contents of this issue:

- In "Calculating All that Jazz: Accurately Predicting Digital Storage Needs Utilizing Digitization Parameters for Analog Audio and Still Image Files," Krista White considers the challenges posed to library personnel who lack computer science or audio visual training who are tasked with writing digital project proposals, grant applications, or rationale to digitization projects at their institutions.

- Myung-Ja K. Han, Nicole E. Ream-Sotomayor, Patricia Lampron, Janet Weber, and Deren Kudeki detail the challenges and expense of creating MARC data for unique and hidden collections in their paper "Making Metadata Maker: A Web Application for Metadata Production." The authors solved the problem at their institution by developing the web application Metadata Maker, which enables anyone, regardless of their familiarity with metadata standards, to create metadata in four formats, including MARC21.
- "An Analysis of Evolving Metadata Influences, Standards, and Practices in Electronic Theses and Dissertations," by Sarah Potvin and Santi Thompson, seeks to raise awareness of the differences between current practices and metadata standards and guidelines for electronic theses and dissertations. They consider the philosophies that have guided the design of several metadata standards.
- Dawn McKinnon's paper "Using Perceptions and Preferences from Public Services Staff to Improve Error Reporting and Workflows" explores the workload impact that has resulted from the ongoing transition of mostly print purchases to electronic, which has led libraries to focus on improving workflow efficiencies.

I hope you enjoy this issue of *LRTS*.

Calculating All that Jazz

Accurately Predicting Digital Storage Needs Utilizing Digitization Parameters for Analog Audio and Still Image Files

Krista White

Library professionals and library assistants who lack computer science or audiovisual training are often tasked with writing digital project proposals, grant applications or rationale to fund digitization projects for their institutions. Much has been written about digitization projects over the last two decades; digital storage has been highlighted as a central feature of any digitization project, especially the need to purchase additional storage mechanisms to house digitized collections. What is missing from the library science literature is a method for reliably calculating digital storage needs on the basis of parameters for digitizing analog materials such as documents, photographs, and sound recordings in older formats.

Much has been written about digitization projects over the last two decades, and digital storage has been highlighted as a central feature of any digitization project. What is missing from the library science literature is a method to reliably calculate digital storage needs on the basis of parameters for digitizing analog materials such as documents, photographs, and sound recordings in older formats.[1] Library professionals and library assistants who lack computer science or audiovisual training are often tasked with writing digital project proposals, grant applications or providing rationale to fund digitization projects for their institutions. Digitization projects involve purchasing additional storage mechanisms to house files for preservation and access. Digital project managers need tools to accurately predict the amount of storage for housing digital objects and estimate startup and ongoing costs for such storage.[2] To make those predictions, they must decide which standard their organization will use to create archival masters for long-term access and/or preservation because the standards they apply will affect digital file sizes. This paper provides two formulae for calculating digital storage space for uncompressed, archival master image and document files and sound files. The two formulae presented provide parameters for digitization that will also aid digitization project managers to make informed decisions regarding digitization standards and equipment purchases for their projects. Formulae for 3-D scanning and moving image (video) objects would be a valuable addition to the field, but are beyond the scope of the current study.

The first part of this paper lays out the method for the formulae for predicting the digital storage needs of analog objects, which depends on their media types and characteristics. The second section, the literature review, demonstrates

Krista White (kwhite2@rulmail.rutgers.edu) is a Digital Humanities Librarian at Rutgers University Newark.

Manuscript submitted July 20, 2015; returned to author for minor revision October 5, 2015; revised manuscript submitted October 28, 2015; accepted for publication November 6, 2015.

The author wishes to thank Isaiah Beard, Fernanda Perrone and Robert Nahory for their assistance and advice in the research and writing of this paper.

aspects of digital project management, contextualizing the environment in which librarians and digital project managers must predict digital storage needs, including costs, professional debates about digitization as a preservation tool, and varying best practices and standards documents that complicate project implementation. The third section of the paper introduces the formulae, the experiment design, and the results of testing the formulae for accuracy and reliability. In the final section, the results of the experiments and the elements of the formula for still image and document storage calculations are contextualized using experiences reformatting the transcripts for the Jazz Oral History Project (JOHP) at the Institute of Jazz Studies (IJS) at Rutgers University. The appendix at the end of the essay defines terms to help those new to digitization navigate specialized terminology used here.

The JOHP is

> a collection of audio tapes for 120 oral histories of seminal pre-Swing Era and Swing Era jazz musicians recorded between 1972 and 1983. The JOHP was initiated in 1972 by the Jazz Advisory Panel of the Music Program of the National Endowment for the Arts. Musicians sixty years and older (as well as several younger artists in poor health) were interviewed in depth about their lives and careers. The taped interviews range in length from 5 to 35 hours each and are accompanied by typewritten transcripts. They have been consulted by hundreds of scholars and writers producing articles, books and dissertations, in addition to frequent use by producers of radio and television.[3]

The process of digitizing the nearly 26,000 pages of transcripts for ingestion into RUcore, the Rutgers digital repository, is underway to make the transcripts and audio files of the JOHP publicly available online. Research on calculating digital storage needs occurred simultaneously with the JOHP digitization project because other oral history projects were being submitted to the libraries for digitization and digital storage of the JOHP needs for these projects also needed consideration. In conjunction with advice from the Rutgers University Libraries' (RUL) Digital Data Curator, the research presented here helped in the evaluation of the digitization standards and processes used for digitizing the JOHP and helped to highlit how much storage space the team saved on the institutional repository servers.

Method

The first part of the method involved a literature review, combing computer science literature on digital storage, library science and archival studies literature regarding digital libraries, and professional literature on standards and best practices for the digitization of materials. During the course of the research, the author discovered formulae for calculating digital storage on the basis of the characteristics of analog materials. These formulae surfaced in older instructional and do-it-yourself literature on multimedia object creation and in online multimedia and computer science literature designed for high school and undergraduate students. The second part of the method focused on testing the found formulae to determine whether they were reliable and accurate. The first experiment tested the accuracy of formulae for predicting file sizes when digitizing still images and documents. A second experiment tested the accuracy of the formula for predicting file sizes of digitized audio recordings. The findings on the formulae and accuracy were then applied to the work digitizing the JOHP transcripts. The JOHP example demonstrates how project managers can use the formulae discussed to make decisions about purchasing equipment and evaluate digitization standards to meet the needs of their projects and institutional goals.

Literature Review

The project began with a literature search for ways to calculate digital storage needs of digitized, analog objects. That search met with no success. This may be because librarians depend on their information technology specialists to supply such information. It is absolutely appropriate for librarians to rely on experts from areas like information technology (IT), which typically fall outside the domain of the profession, to help them calculate storage needs for digitization projects. However, IT professionals, even within library systems and IT groups, are not always familiar with digital preservation best practices. If they are familiar with digital preservation practices, IT professionals often present storage figures in absolute terms, assuming fixed values for digitization variables that may or may not be appropriate for long term preservation. In his 1997 book, *Practical Digital Libraries*, Lesk estimated thirty megabytes of storage for every hour of compressed audio, one megabyte for a page of uncompressed, plain text (bitmap format), and three gigabytes for two hours of moving image media.[4] Lesk gives no estimate for raster images such as TIFF, JPEG, or GIF images, or for the storage of uncompressed, archival master digital files; rather, he is concerned with providing figures that represent the most economical memory and storage options for delivery of objects in a digital library system. Jordan estimates storage needs for raster image files of 90 megabytes for uncompressed raster image files, 600 megabytes for one hour of uncompressed audio recording, and "nearly a gigabyte of disk space," for one minute of uncompressed digital video.[5]

In *The State of Recorded Sound Preservation in the United States*, the National Recording Preservation Board quotes a figure of 100 gigabytes (GB) of storage for 100 hours of audio tape.[6] Calculations given by these information technology professionals and standards organizations are accurate (if approximate in some cases), but they assume fixed rates for many variables in the digitization process that may not suit a particular institution's needs or the chosen digitization standard for a project. Those variables can be adjusted to alter both file size and quality, which affects the choice of digitization standard, the combination of variables used in a chosen standard and the quantity of digital storage required.

Among librarians and archivists, the issue of digital storage is taken quite seriously for digitization and digital library projects. In the newly released *Association for Recorded Sound Collections (ARSC) Guide to Audio Preservation*, Lacinak's chapter provides an overview of the issues related to digital storage, providing an in-depth example with guidance on decision making in that domain of digital initiatives.[7] Other literature in the field discusses storage as the platform for stable, long-term storage of digital assets. The section on storage in Hodge's paper on a lifecycle framework emphasizes its importance as a mechanism for long-term preservation.[8] Hooper's audio e-reserves project at Tulane features stable, digital storage for the new, master files as a cornerstone of the project, as do Pastine, Bayard, and Lang in a similar project at Temple University to create an e-reserves system for digital images for courses and general discovery.[9]

Other library science literature frames digital storage as a basic resource of digital projects whose capacity will need to be enlarged to accommodate digital library initiatives. Jones' paper on the creation of a history portal for digital objects related to the history of Michigan lays the issue out plainly, "Planning for storage needs is an ongoing task. . . . The increase in storage capacity made necessary by (Making of Modern Michigan) is only a fraction of the increase needed to handle the expansion of our own digitized collections."[10] This was the case with Pastine, Bayard, and Lang's project at Temple University, and was a feature of Maurya's paper on the challenges and hopes for digital library services in India.[11]

The suitability of migrating existing materials into digital format for preservation purposes has been contentious for more than a decade.[12] However, Arthur and her colleagues have argued that digitization be accepted as a preservation format.[13] In concert with Hodge et al., Arthur and her colleagues have highlighted the importance of digital storage as a final destination for files migrated from older, analog formats.[14] Some individuals and organizations may not consider reformatting analog materials to digital files as a long-term, viable preservation strategy, but for many projects, including the JOHP audio tape recordings, this option is the best for analog materials that have reached the end of their useful life. In these cases, library science literature demonstrates the importance of storage in digitization initiatives, but there is no guidance in any of these sources for a method to estimate the amount of storage a given project requires.

Knowing the cost of digital storage—whether starting a new project or expanding on existing storage infrastructure—is crucial to digital project planning. Costs of digital storage are generally framed in terms of dollars or cents per storage unit. Planning for storage costs, therefore, requires knowing how much storage a project will need. In an older example, Lesk's 1990 report to the Foundation of the American Institute of Conservation for Conservation at the American Institute for Conservation of Historic and Artistic Works provides a detailed chart of the cost of various digital storage formats for library materials.[15] More recently, Lazorchak compiled an excellent bibliography on "Digital Asset Sustainability and Preservation Costs."[16] Echoing Kenney and Personius, two papers featured in Lazorchak's bibliography, one by a group at the San Diego Supercomputer Center and the other by Sanett, provide an overview of the costs of different storage media plus maintenance, labor, infrastructure software licenses utilities and floor space associated with digital storage hardware.[17] Lesk's paper is the only one that provides actual numbers for magnetic hard disk costs (what is often called "server storage"), quoting a price of $4,000 per gigabyte in 1990.[18] Smith's informal but comprehensive bibliography on the cost of hard drive storage space claims the cost in 1990 was $9.00 per megabyte ($9,216.00 per gigabyte) and shows current costs per gigabyte of storage to be $6.33 as of July 2013.[19] Digital storage comes in many forms, from gold CDs to magnetic hard drive arrays connected to networks (also known as "server storage" or "cloud storage"). Costs vary by storage format and must be sustainable. Despite the downward trend in the cost per gigabyte, storage media must be periodically replaced as hardware gets old and experiences failures or the format becomes obsolete. That translates into ongoing, permanent costs for storage mechanisms in every digital project. Even small costs can be burdensome to cultural heritage institutions working with limited budgets. The ability to plan for costs related to the growth of digital assets hinges on an organization's ability to accurately estimate the amount of digital storage for current and future objects in a collection.

Further complicating the work of digital project managers and directors in estimating digital storage needs is the existence of multiple standards and best practice documents for proper stewardship of archival digital materials because standards used to digitize analog materials directly affect file sizes. The Library of Congress' Federal Agencies Digitization Guidelines Initiative includes recommendations and resources for digitizing still images and advice for preparing the digitization environment, file format comparisons, digitization workflows and overall stewardship recommendations.[20] It is a deeply technical document that institutions can use to evaluate and create their own digitization standards. The Bibliographical Research Center built upon the Colorado Digitization Project's work to create the Best Practices for Digital Imaging. Their document includes a nuanced, understandable explanation of the digitization process, recommendations for decision making for staffing, and training and software and hardware considerations for digitization projects, plus concrete parameters for digitizing analog materials.[21] The Smithsonian Institution Archives and the World Digital Library both have simple webpages detailing digitization standards for their collections.[22] The Colorado Digitization Project wrote a guide to best practices for digitizing analog audio sources and the Association for Library Collections and Technical Services (ALCTS), a division of the American Library Association whose Preservation and Reformatting Section's (PARS) mission includes, among other things, the preservation of library materials, created a guide for digitizing all types of analog objects according to format.[23] The Digital Preservation Coalition has a standard for digitizing moving image media, which differs significantly in file format, bit rate, and color recommendation from that of the standard set out by the Consortium of Academic and Research Libraries in Illinois (CARLI), while the Federal Agencies Digitization Guidelines Initiative's Audiovisual Working Group is still formulating its standards for video/moving image materials.[24] Individual institutions, especially those with digital repositories, may create their own guidelines. Rutgers has locally developed standards for digitizing analog documents, images, audio, and moving image materials that are based on independent review and testing of standards set by other bodies.[25] Utilizing standards is crucial to creating stable, long-term digital surrogates of older archival objects, but the existence of multiple standards, even when closely matched, may be confusing to the uninitiated digital project manager. Standards bodies do not provide insight into how standards affect digital storage needs or provide guidelines that would help library professionals choose appropriate settings for parameters within those standards. These variables profoundly affect the amount of digital storage necessary for a project.

Reviewing older instructional and do-it-yourself literature on the creation of multimedia objects resulted in the discovery of mathematical formulae for calculating digital storage needs for analog still images, documents, audio and moving image recordings. Many of these sources were rightly concerned with explicating the process of digitization and monitoring variables to insure quality.[26] Only a few were concerned with the practicality of determining the size of digital files in the final output of the digitization process. Three resources yielded formulae for calculating the file sizes of digitized still images and documents: Tally's *Avoiding the Scanning Blues*, Note's *Managing Image Collections: A Practical Guide* and Cunningham's formula for digital, bitmap still images.[27] Tally's formula omits the essential element of the physical dimensions of the scanned image, which is crucial in calculating file sizes for photos and documents; Both Note's and Cunningham's formulae include image or document size, scanning resolution and bit depth. The formulae in these two sources are essentially the same.[28]

Cunningham's webpage and Johnson, Gault, and Florence's *How to Digitize Video* were the two sources that elucidated formulae for calculating files sizes for digitized, analog audio recordings.[29] The formulae in both sources contained the same elements necessary for calculating storage sizes: length of the original audio recording, sampling rate, bit depth, and number of audio channels.[30]

Three resources in the literature review contained formulae for predicting uncompressed digital file sizes for moving image (video) objects. Rice and McKernan's formula seemed incomplete. They added an extra, unnecessary number for RGB color, which should be accounted for in the bit depth value and their formula lacked any variables to account for sound in the moving image recordings.[31] Cunningham's and Johnson, Gault, and Florence's formulae contained mathematical elements that included frame rate, frame size, bit depth, and length of recording.[32] The combination of Cunningham's two separate formulae for calculating uncompressed digitized moving image files and for uncompressed audio files is identical in content to Johnson, Gault, and Florence's for calculating file sizes for uncompressed audiovisual materials.[33] Johnson, Gault, and Florence combine two formulae for calculating audiovisual materials; they present one formula for the moving image portion of an audiovisual file and another formula for the sound portion of the video file.[34] Cunningham presents his moving image formula separately from the audio formula and does not make clear if they should be combined to calculate the size of digitized audiovisual materials.[35]

None of the formulae proposed by authors listed in the literature review provided supporting evidence of their effectiveness. This required experimentation to test the accuracy and reliability of the formulae. Though the literature review produced formulae for still images and documents as raster files, audio recordings and moving image recordings, the scope of the current paper is limited to testing and explicating the formulae for the reformatting of still images and documents and analog audio files into digital formats. The complexity of the processes behind digitizing analog moving image or video, plus that for scanning as-yet-unmentioned 3D objects, requires its own experimentation and analysis beyond the scope of the current work.

Also absent from the literature reviewed are formulae and experiments for predicting file sizes for born-digital media in all formats. Many tutorials with formulae are available online, created by instructors for high school and undergraduate-level computer science courses.[36] As with the formulae for calculating storage space of analog-to-digital reformatting procedures, the formulae presented for born-digital objects do not contain data on their reliability or accuracy. In the context of digital exhibits and the archival preservation of born-digital objects, calculating storage space for uncompressed, born-digital objects would be invaluable, but is beyond the scope of this study.

$$\text{File Size in Bytes} = \frac{\text{image width} \times \text{image height} \times \text{resolution}^2 \times \text{bit depth}}{8}$$

Figure 1. Formula for Calculating File Sizes of Uncompressed, Still Images

Image/document size: 2x3 inches, 8.5x11 inches, 11x17 inches.
Resolution: 150 ppi, 300ppi and 600ppi
Grayscale image bit depths: 8 bits, 16 bits
Color image bit depths: 24 bits, 48 bits

Figure 2. Experiment Variables Used to Test the Accuracy of the Still Image Formula

Experiment Design: Accuracy of the Still Image Formula

Figure 1 displays the formula for calculating the file size of uncompressed, unedited still images in bytes, suitable for use as archival master files.[37] Image scanning experiments were performed using an IBM PC with Windows 7 operating system to test the reliability of the formula. An Epson Expression 10000XL scanner and the native EsponScan software, version 3.49A were used to digitize still images. Images were captured as uncompressed TIFF files in accordance with digital, archival practices set out by standards bodies mentioned earlier in this essay. Images were scanned using a combination of variables in each scan, as shown in figure 2.

Variables were chosen on the basis of the digitization standards for still images and documents in the "BCR's CDP Digital Imaging Best Practices," "Minimum Digitization Capture Recommendations" and in "Digitizing Analog Documents and Images."[38] Any variables that do not match those standards were chosen to create atypical file combinations that would test the limits of the still image digital storage calculation formula.

Bit depth was separated into grayscale and color categories because of the fundamental difference between digital capture of grayscale versus color imagery. The two most common archival standards of 24 bits and 48 bits were used to capture color images.[39] The combination of variables resulted in twelve scans per image, with a total of thirty-six images scanned at various document sizes, bit depths and resolutions. Each file was assigned a unique ImageID that indicated its size, scanning resolution, bit depth and whether it was scanned in color or grayscale. For instance, one file was labeled "Si85x1160024C." Si indicated that it was a still image, "85x11" indicated that the original document was 8.5 by 11 inches in size, "600" indicated that it was scanned at a resolution of 600 pixels per inch (ppi), "24" indicated that it was scanned at a bit depth of 24, and "C" indicated that it was scanned for color.

Once the scans were complete, two methods were used to obtain the *measured file sizes* (labeled Ai_1, Ai_2, Ai_3, Ai_4). The first instrument used to obtain *measured file sizes* was the Windows Explorer details Pane of Windows 7 (Ai_1, Ai_3). When a user clicks on a file to highlight it in the Windows 7 operating system, the details pane displays metadata about that file, including the file size. The second instrument used to obtain *measured file sizes* was Media Info (Ai_2, Ai_4), an open source software program that displays technical and source metadata about multimedia files.[40]

All data about the scanned images and documents from the experiment were entered into a Microsoft Excel spreadsheet. Excel's calculate function was used to anticipate the *calculated file sizes* in kilobytes (Ci_1) and in megabytes (Ci_2) using the uncompressed still image file formula.[41]

The differences (Di_1, Di_2, Di_3, Di_4) between the *calculated file sizes* (Ci_x) from the formula and the measured file sizes recorded from Windows Explorer and Media Info in both kilobytes (Ai_1, Ai_2) and megabytes (Ai_3, Ai_4) were calculated in Excel.

$$Di_x = Ai_x - Ci_x$$

Looking at numerical differences between file sizes is useful, but does not provide the lay user with a sense of the value of the differences (Di_x) between the calculated values (Ci_x) and the measured file sizes (Ai_x). A pure mathematical difference would not be an informative measurement of the accuracy of the formula, since different sized files would not produce comparable, uniform variations. To that end, the Percent Difference (Pi_x) between the calculated file size and the measured file size was calculated to show the percentage of the measured file size represented by the difference (Di_x) between the measured file size (Ai_x) and the calculated file size (Ci_x).

$$Pi_x = |Di_x \div Ai_x| \times 100$$

There were discrepancies between values for some of the measured file sizes reported by Windows Explorer and Media Info. Adobe Photoshop CS6 was used as a control to

Table 1. Abbreviations for Still Image File Variables in Calculations by Instrument and Unit of Measurement

Variable	No Instrument (KB)	No Instrument (MB)	Windows Explorer (KB)	Media Info (KB)	Windows Explorer (MB)	Media Info (MB)
Calculated File Size	Ci_1	Ci_2				
Measured File Size			Ai_1	Ai_2	Ai_3	Ai_4
Differences between File Sizes			Di_1	Di_2	Di_3	Di_4
Percent Difference between File Sizes			Pi_1	Pi_2	Pi_3	Pi_4

Table 2. Calculated File Size Compared to Actual File Size of Digital Image Files. Brackets | | indicate absolute values.

Still ImageID	Size (in)	PPI	Color/ Gray	Bit Depth	Calculated File Size (MB) Ci_2	Actual File Size (MB) Ai_4	Difference (MB) Di_4	Percent Difference (MB) Pi_4
Si2x315024C	2x3	150	C	24	0.386	0.395	0.010	2.102
Si2x315048C	2x3	150	C	48	0.772	0.781	0.010	1.123
Si2x330048C	2x3	300	C	48	3.090	3.100	0.010	0.326
Si2x330024C	2x3	300	C	24	1.545	1.550	0.010	0.326
Si2x360048C	2x3	600	C	48	12.360	12.400	0.040	0.326
Si85x111508G	8.547x11	150	G	8	2.017	2.020	0.000	0.129
Si85x1115024C	8.547x11	150	C	24	6.052	6.060	0.010	0.129
Si2x33008G	2x3	300	G	8	0.515	0.516	0.000	0.124
Si2x315016G	2x3	150	G	16	0.257	0.258	0.000	0.124
Si2x31508G	2x3	150	G	8	0.129	0.129	0.000	0.124
Si11x1730024C	11x17	300	C	24	48.151	48.200	0.050	0.102
Si85x1160016G	8.548x11	600	G	16	64.564	64.600	0.040	0.056
Si11x1760048C	11x17	600	C	48	385.208	385.000	|-0.208|	|-0.054|
Si85x1130024C	8.547x11	300	C	24	24.209	24.200	|-0.009|	|-0.036|
Si85x113008G	8.547x11	300	G	8	8.070	8.070	0.000	0.006
Si11x176008G	11x17	600	G	8	64.201	64.200	0.000	|-0.002|

compare measured file sizes. The comparison between measured file sizes reported by Windows Explorer (Ai_3), Media Info (Ai_4) and Photoshop CS 6 (PS_2) in megabytes revealed Media Info to be the preferred reporter of measured file sizes because file sizes measured in Photoshop matched Media Info's measured file sizes more often than they matched measured file sizes in Windows. For the sake of consistency, only files measured in megabytes are reported in this study.[42]

Because the aim of the experiment was to test the reliability and accuracy of the still image digital storage formula, the *absolute* valuesof Di_x and Pi_x were used.[43] The rationale for this choice is that the most desirable value for determining the accuracy of the still image digital storage formula is zero. Therefore, all values produced in the experiment are evaluated as more, or less, accurate by their distance from zero. See the appendix for the definition of *absolute value*.

Still Image Experiment Results and Discussion

Table 2 compares a sample of the data from sixteen of the thirty-six total files created in the experiment. The table compares the *calculated file size* in megabytes of scanned images using the still image (Ci_2), the *measured file sizes* of the files as reported by the Media Info software in megabytes (Ai_4), the *difference* between the calculated files size and the measured file size (Di_4), and the *percentage of the measured file size* that the difference between the calculated and measured file sizes represents (Pi_4). The results in the table represent a spread of files that demonstrate the least amount of accuracy (highest Pi_4 values), results that demonstrate some error between calculated and measured file size (median Pi_4 values) and files demonstrating the least amount of error (small or no Pi_4 values). Negative numbers indicate that Ci_2 was larger than Ai_4.

The 2.0 percent and 1.0 percent errors displayed by

$$\text{File Size in Bytes} = \frac{\text{length of recording in seconds} \times \text{sampling rate in Hz} \times \text{bit depth} \times \text{number of channels}}{8}$$

Figure 3. Formula for Calculating File Sizes of Uncompressed Audio Recordings

Length of Recordings were 30 seconds, 60 seconds, 90 seconds
Sampling Rates at 44.1kHz, 48kHz, 96kHz, and 192kHz
Bit Depths of 16 bits and 32 bits, with the exception of sound recorded at 192kHz
(for which the Adobe Audition software would only record at 16 bits and not 32 bits)
Channels Single and double channels (mono and stereo)

Figure 4. Experiment Variables Used to Test the Accuracy of the Audio Formula

Si2x315024C and Si2x315048C are small and represent statistical outliers in the current dataset. Comparing the rest of the data collected demonstrates that the calculated file sizes differ less than 0.5 percent from the measured file sizes in 94.0 percent of the sample size. The large differences displayed for Si2x315024C and Si2x315048C may be the result of the relatively small file sizes of each; any Pi_4 value will represent a larger absolute percentage of Ai_4 because of the small file sizes. All indications are that the formula for calculating uncompressed, still image digital file sizes is based on original object dimensions, scanning resolution and bit depth is accurate and reliable enough for common use.

Experiment Design: Accuracy of the Audio Formula

Figure 3 displays the formula for calculating the size of uncompressed digital audio files based on recording length, digitization sampling rate, bit depth and number of channels. The experiments were performed on an IBM PC running the Windows 7 operating system. Equipment included a Sony TC-WE475 Stereo Dual Cassette Deck and the audio was transcoded from a standard, commercial-grade music cassette using a Creative Labs Sound Blaster Converter as the Analog to Digital (ATD) device. The audio feed for the single channel audio files of the experiment were captured using Audacity software downloaded for free from the Internet.[44] The dual-channel audio files were captured using Adobe Audition 3.0 software.

To test the formula for calculating uncompressed analog to digital audio conversion, the audio was recorded as uncompressed .WAV files from the same ninety seconds of a commercial music tape using the variables shown in figure 4.

This set of variables was chosen because they contain values recommended by both the Rutgers *Sound Object Archival Standards* and by many the standards bodies mentioned in this paper.[45] Any variables that do not match those standards were chosen to create atypical file combinations that would test the limits of the audio digital storage calculation formula. Sample audio files were recorded for ninety seconds and then copied and cut down those files into sixty- and thirty-second lengths for each sampling rate, bit depth and channel combination, creating a total of forty-two sound files. Lengths of recordings were chosen to provide a variety of sample sizes and because they were intervals that were easy to produce in the editing software suites used in the experiment.

Sampling rates were chosen based on the entire range of possibilities for digitizing sound recordings available in each of the two software suites used in the experiment. A sampling rate of 44.1kHz is the CD quality standard for digital audio recordings, and thus was set as the lowest sampling rate in the experiment. A higher sampling rate of 96kHz is recommended by various standards bodies. Bit depths of 16 and 32 were chosen to represent extremes. Half of the files were recorded with one channel and half of the files with two channels. This decision provided more than one value for the channels variable, but kept the number of created samples at a manageable level for the experiment.

Each file was assigned a unique AudioID which indicated the combination of variables used. For instance, one file was labeled "Au30-44-16-001." Au indicated it was an Audio file, "30" represented the number of seconds in length, "44" indicated the 44.1kHz sampling rate, "16" indicated the bit depth, and "001" indicated the number of channels.

All data about the different audio recordings were entered into a Microsoft Excel spreadsheet. The calculation function in Excel's calculation function was used to calculate the anticipated file size utilizing the uncompressed audio file formula in megabytes (Ca_1).

For the audio files produced in the experiment, all of the *calculated* (Ca_x) and *measured file sizes* (Aa_x) fell into the megabyte size range. Discussion and examples of the results of the audio file formula tests will, therefore, be limited to those measured in megabytes. Already having determined that Media Info was the preferred *instrument* for the measurement of measured file sizes in the portion of the experiment dedicated to still images, only results comparing measured file sizes as reported by Media Info are presented.

The differences (Da_1, Da_2, Da_3, Da_4) between the calculated file sizes from the formula (Ca_x) and the measured

Table 3. Abbreviations for Audio File Variables in Calculations by Instrument and Unit of Measurement

Variable	No Instrument (KB)	No Instrument (MB)	Windows Explorer (KB)	Media Info (KB)	Windows Explorer (MB)	Media Info (MB)
Calculated File Size	Ca_1	Ca_2				
Actual File Size			Aa_1	Aa_2	Aa_3	Aa_4
Differences between File Sizes			Da_1	Da_2	Da_3	Da_4
Percent Difference between File Sizes			Pa_1	Pa_2	Pa_3	Pa_4

file sizes recorded from Windows Explorer and Media Info in both kilobytes (Aa_1, Aa_3) and megabytes (Aa_2, Aa_4) were computed using Excel's calculating function.

$$Da_x = Aa_x - Ca_x$$

To provide consistency in method and presentation of the data, the Percent Difference (Pa_x) between the calculated file size and the measured file size of audio files in the experiment was calculated to demonstrate the relative accuracy of the formula.

$$Pa_x = |Da_x \div Aa_x| \times 100$$

As in the experiment with the still image digital storage formula, the results of the experiment for the audio digital storage formula will be discussed in terms of the *absolute values* of Da_x and Pa_x.

Audio Experiment Results and Discussion

Table 4 contains results comparing the *calculated file size* (Ca_2) from the formula for calculating digital storage needs from an inventory of analog audio materials, the *measured file sizes* as reported in Media Info (Aa_4), the *difference* between the calculated file size and the measured file size (Da_4) and the *percentage of the measured file* size represented by the *difference* between the calculated file size and the *measured file size* (Pa_4). A sample of sixteen files were chosen that represent results with the least amount of accuracy (the highest value of Pa_4), results that demonstrate some error (median Pa_4 values) and files that represent the least amount of error or no error (small or no Pa_4 values). Negative numbers indicate that Ca_2 was larger than Aa_4.

Of the forty-two files tested, the calculated file sizes are always less than 1 percent different from the measured file sizes. The largest absolute Pa_4 is for the 90-second clip recorded at 192kHz with a bit depth of 16 and 2 channels; the difference represents an absolute Pa_4 value of 0.88 percent of the measured file size. The file with the smallest Pa_4 value was 90 seconds long, recorded with a sampling rate of 96kHz at 32 bits with two channels; its Pa_4 value was 0.03 percent of the measured file size. The absolute median Pa_4 value of measured file sizes in the data set is 0.12 percent. The absolute mean Pa_4 value of measured file sizes as reported by in the dataset is 0.23 percent. Unlike the still image digital storage formula, there are no instances of absolute Pa_4 values that are extremely high or extremely low when compared with the absolute Pa_4 values of other files in the experiment. There were no files for which Ca_2 exactly matched the measured value.

The data show that the formula for calculating file sizes for uncompressed .WAV files from analog audio sources is extremely reliable. Examining the files with the ten highest absolute Pa_4 values and the ten lowest absolute Pa_4 values indicates that the formula is most accurate when using shorter recordings with a sampling rate in the 48kHz or 96kHz range at a lower bit depth. For the purposes of planning digital storage, the errors in the data are so small—less than 1 percent in all cases—that they are of no real concern. The trend toward slightly less accuracy with larger file sizes will only be proven or disproven with a much larger sample set.

Applications of the Still Image Formula in the Jazz Oral History Project

The high accuracy of the formulae in these experiments indicates that they can be used reliably when attempting to calculate storage needs for an audio digitization project. As the project manager for the JOHP, and as a regular consultant on digitization projects for teaching faculty at her institution, the author has found the still image and audio formulae invaluable in calculating storage needs and evaluating both digitization standards and equipment for projects.[46]

When working with a collection of analog documents, images and audio recordings, the physical dimensions or duration of objects in a collection are predetermined. The other elements of the audio and still image formulae, bit depth, resolution/sampling, color type (for still images and documents), rate and number of channels (for audio files), are variable depending on the standards used. Project managers can use the still image and audio formulae to determine how much storage space they will need to house the

Table 4. Calculated File Size Compared to Actual File Size of Digital Audio Files. Brackets | | indicate absolute values used in analysis.

AudioID	Length (sec)	Sampling Rate (KHz)	Channels	Bit Depth	Calculated File Size (MB) (Ca)	Measured File Size (MB) As_4	Difference (MB) Da_4	Percent Difference (MB) Pa_4
Au90-192-16-002	90.00	192.0	2	16	65.91	66.50	0.58	0.88
Au60-44-16-001	60.00	44.1	1	16	5.04	5.09	0.04	0.85
Au90-192-16-001	91.00	192.0	1	16	33.32	33.60	0.27	0.82
Au90-96-16-001	89.00	96.0	1	16	16.29	16.40	0.10	0.63
Au60-192-16-001	60.00	192.0	1	16	21.97	22.10	0.13	0.58
Au90-96-16-002	91.00	96.0	2	16	33.32	33.50	0.17	0.52
Au90-44-16-001	90.00	44.1	1	16	7.57	7.61	0.04	0.52
Au60-44-32-001	59.94	44.1	1	32	10.08	10.10	0.02	0.16
Au60-96-16-001	60.00	96.0	1	16	10.98	11.00	0.01	0.12
Au60-96-16-002	60.00	96.0	2	16	21.97	22.00	0.03	0.12
Au90-48-32-001	90.00	48.0	1	32	16.47	16.50	0.02	0.12
Au30-48-16-002	30.28	48.0	2	16	5.54	5.54	0.00	|-0.07|
Au30-44-32-001	30.00	44.1	1	32	5.04	5.05	0.00	0.06
Au30-44-32-002	30.00	44.1	2	32	10.09	10.10	0.01	0.06
Au60-44-16-002	60.00	44.1	2	16	10.09	10.10	0.01	0.06
Au90-96-32-002	90.00	96.0	2	32	65.91	65.90	0.01	0.03

archival quality digital surrogates in a collection. Once the amount of digital storage has been calculated, if the budget of the project disallows purchasing enough storage to house the entire collection, digital project managers can make strategic decisions about which objects would benefit the most from digitization on the basis of the original document and recording conditions, user interest, and institutional mission. They may also choose to utilize a different standard if adjusting bit depth, scanning resolution, or sampling rates would enable digitizing the entire collection. Digitizing the JOHP transcripts served as a case study which confirmed the usefulness of the still image formula.

Because the JOHP files will be housed in RUcore, the RUcore standard, "Digitizing Analog Documents and Images," was used as the guideline for the project.[47] This standard falls well within the spectrum of other digitization standards developed by bodies both national and regional. Many of the transcripts from the JOHP are on older typing paper and, in some cases, have a yellowed appearance. To capture the look and feel of the original documents to provide the user with an experience as close to handling the physical pages as possible, the RUcore standard for capturing color images was chosen. That standard requires a resolution of 600 ppi, the use of the RGB colorspace with 24-bit color, and outputting files in TIFF file format. The project began with digitization of one of the longest single transcripts, the interview with jazz great Maxine Sullivan, totaling 775 pages. This provided project staff with a robust sample for testing workflows and the digitization standard.

Plugging a height of 11.5 and a width of 9 for the page sizes (to accommodate the edges of the paper) into the still image formula results in a file size of 106.6 MB per page. It would require 80.16 GB of space to store 770 such pages, which is quite a lot of storage for a single document that is part of a larger collection. The JOHP collection contains 25,995 pages of transcripts. At the chosen bit depth and resolution for the project, the still image formula indicates a total size of 2.64 Terabytes (TB) for all JOHP transcripts. After fine-tuning the scanner settings, the scanned area for each page was adjusted to 8.82 by 11.10 inches. This results in a per-page file size of 100.84 MB and a total storage size of 2.50 TB for all 25,995 pages of transcripts.

The total storage capacity of RUcore is currently 55 TB, expandable to 15.5 Petabytes; 2.50 TB is approximately 4.5 percent of the total current storage on the RUcore servers. Each interview transcript is accompanied by approximately three to five hours of audio files, which add additional storage requirements. While the RUcore servers can handle such volume, it is always wise to try to conserve as much storage space as possible to save on maintenance, upgrade and labor costs for stewardship over the life of the data being stored.

After considering the size of the digitized Sullivan transcript, the JOHP staff determined that transcripts should be scanned at the 600 ppi required for color documents, color type and bit depth were changed to 8-bit grayscale instead of 24-bit color in the interest of lowering the size of files for each page of transcripts.[48] The desire to scan them in 24-bit color was an aesthetic choice that needed to be altered to accommodate stewardship of the entire collection of both transcript files and audio files in the collection. Providing the “look and feel” of the original transcripts in digital format would have been pleasant for users, but was set aside in favor of storage economy. Recalculating the file sizes for the entire collection with adjusted scanning parameters revealed that the changes would result in a 66 percent reduction in the necessary storage capacity for the transcript files. This reduction took the total storage needed for all 25,995 pages down from 2.50 TB to 853.21 GB or 0.83 TB, a much more manageable storage requirement for the entire collection, which demonstrated the efficacy of the formula.

Conclusion

In July 2015, AVPreserve, a consulting firm that helps institutions manage and implement digital library projects, released, “Quantifying the Need: A Survey of Existing Sound Recordings in Collections in the United States.”[49] In the report, Lyons, Chandler, and Lacinak estimate that there are 254,159,631 preservation-worthy audio holdings in US collections, and that the market cost of the digitization process for these items would be more than twenty billion dollars, “which does not include the costs that will be associated with . . . ongoing storage of digital files for preservation and access.”[50] Assuming a very conservative estimate of 5 minutes per audio recording, with CD quality bit depth of 16 and sampling rate of 44.1 kHz in stereo (two channels), using the audio calculation formula, we know that purchasing storage at the current, consumer rate of approximately one dollar per gigabyte will require an extra $12,526,408.00 for the purchase of storage media alone. AVPreserve’s survey does not indicate the average length of preservation-worthy audio recordings in its survey; the cost of storage media could be much, much higher. In an era of shrinking academic and cultural heritage budgets, purchasing digital storage to house and preserve these audio objects will be no mean feat.

The author has found that the still image and audio formulae are valuable tools for anticipating digital storage needs and for helping faculty outside the library evaluate their equipment for digitization projects. As the experiments demonstrate, the formulae for still image and audio recordings are extremely accurate. They will prove invaluable to digital archivists, digital librarians and the average user in helping to plan digitization projects, as well as in evaluating hardware and software for these projects. An understanding of the parameters of digitization contained in each formula—bit depth, color type, scanning resolution, sampling rate and audio channels—provides insight into both the quality of a digital image or sound file and provides guidelines for project managers to evaluate best practice standards and digitization equipment. Digital project managers armed with the still image and audio formulae will be able to calculate file sizes using different standards to determine which standard will suit the project needs. Knowing the parameters of the still image and audio formulae will allow managers to evaluate equipment on the basis of the flexibility of the software and hardware before purchase. Using the still image and audio calculation formulae in workflows will help digital project managers create more efficient project plans and tighter grant proposals.

Future work in the area of calculating digital storage needs to be done. Discovering or developing formulae for uncompressed, archival quality files produced by 3D image scanners and for digitizing analog moving images (video) would add significant value to the library, archival, and cultural heritage professions. Further research into predicting file sizes for born-digital objects, as well as calculating the file size savings when converting digital multimedia files from uncompressed to compressed formats would benefit the literature. Such formulae will enable even more accurate, additional projections to be made for a greater variety of projects.

References and Notes

1. Sam Brylawski et al., eds., *ARSC Guide to Audio Preservation* (Eugene, OR: Association for Recorded Sound Collections; Washington, DC: Council on Library and Information Resources, 2015), accessed May 29, 2015, www.clir.org/pubs/reports/pub164, 223.
2. Brylawski et al., 224.
3. “Jazz Oral History Project,” Institute of Jazz Studies, Rutgers University Libraries, accessed April 25, 2015, http://newark www.rutgers.edu/IJS/OralHistory.html.
4. TechTerms.com, s.v. “Megabyte,” accessed April 1, 2015, http://techterms.com/definition/megabyte; TechTerms.com, s.v. “Gigabyte,” accessed April 1, 2015, http://techterms.com/definition/gigabyte; Michael Lesk, *Practical Digital Libraries: Books, Bytes, and Bucks* (San Francisco: Morgan Kaufmann, 1997), 81.
5. Mark Jordan, *Putting Content Online: A Practical Guide for Libraries* (Oxford: Chandos, 2006), 100, 107, 93.
6. Ron Bamberger and Sam Brylawski, *The State of Recorded Sound Preservation in the United States: A National Legacy at Risk in the Digital Age* (Washington, DC: Council on Library and Information Resources, 2010), 81–82, www.clir .org/pubs/reports/pub148.
7. Chris Lacinak, “What to Do After Digitization,” in *ARSC Guide to Audio Preservation*, edited by Sam Brylawski et al. (Washington, DC: Council on Library and Information Resources, 2015), 127–52, www.clir.org/pubs/reports/pub164.

8. Gail M. Hodge, "Best Practices for Digital Archiving: An Information Lifecycle Approach," *D-Lib Magazine* 6, no. 1 (2000), www.dlib.org/dlib/january00/01hodge.html.
9. Lisa Hooper, "Building an E-Audio Reserves Program at Tulane University: Methods and Procedures," *Journal of Interlibrary Loan, Document Delivery & Electronic Reserve* 22, no. 3-4 (2012): 181–96, http://dx.doi.org/10.1080/1072303X.2012.723671; Maureen Pastine, Ivy Bayard, and Carol Lang, "Digital Diamond: Temple University Libraries' IMLS Grant," *The Bottom Line* 14, no. 2 (2001): 76–84.
10. Ruth Ann Jones, "Empowerment for Digitization: Lessons Learned from The Making of Modern Michigan," *Library Hi Tech*, 23, no. 2 (2005): 205–19.
11. Pastine, Bayard, and Lang, "Digital Diamond"; Ram Nath Maurya, "Digital Library and Digitization," *International Journal of Information Dissemination & Technology* 1, no. 4, (October 2012): 228–31; Jones, "Empowerment for Digitization."
12. Peter J. Astle and Adrienne Muir, "Digitization and Preservation in Public Libraries and Archives," *Journal of Librarianship & Information Science* 34, no. 2 (2002): 67–79; Yan Quan Liu, "Best Practices, Standards and Techniques for Digitizing Library Materials: A Snapshot of Library Digitization Practices in the USA," *Online Information Review* 28, no. 5 (2011): 338–45.
13. Kathleen Arthur et al., "Recognizing Digitization as a Preservation Reformatting Method," *Microform & Imaging Review* 33, no. 4 (2004): 171–80.
14. Arthur et al., "Recognizing Digitization as a Preservation Reformatting Method"; Hodge, "Best Practices for Digital Archiving"; Hooper, "Building an E-Audio Reserves Program"; Lacinak, "What to do After Digitization."
15. Michael Lesk, *A Report of the Technology Assessment Advisory Committee to the Commission on Preservation and Access* (Washington, DC: Commission on Preservation and Access, 1990), http://cool.conservation-us.org/byauth/lesk/lesk#digital.
16. Butch Lazorchak, "A Digital Asset Sustainability and Preservation Cost Bibliography," *The Signal: Digital Preservation* (blog), June 26, 2012, http://blogs.loc.gov/digitalpreservation/2012/06/a-digital-asset-sustainability-and-preservation-cost-bibliography.
17. Richard L. Moore et al., "Disk and Tape Storage Cost Models," San Diego Supercomputer Center, (University of California San Diego: La Jolla, 2007), http://users.sdsc.edu/~mcdonald/content/papers/dt_cost.pdf; Shelby Sanett, "The Cost to Preserve Authentic Electronic Records in Perpetuity: Comparing Costs Across Cost Models and Cost Frameworks," *RLG Diginews* 7, no. 4 (2003), http://library.oclc.org/cdm/singleitem/collection/p267701coll33/id/366.
18. Lesk, *A Report of the Technology Assessment Advisory Committee*.
19. Ivan Smith, "Cost of Hard Drive Storage Space," Nova Scotia's Electric Gleaner, April 3, 2014, http://ns1758.ca/winch/winchest.html. Note: This source is quoted in other, online articles about digital storage, and, despite its unprofessional appearance, diligently collates and tracks its sources from a wide variety of literature including primary sources such as computer hardware catalogs and computing magazine articles.
20. Don Williams and Michael Stelmach, eds., *Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files* (Washington, DC: Federal Agencies Digitization Guidelines Initiative, revised August 2010), www.digitizationguidelines.gov/guidelines/FADGI_Still_Image-Tech_Guidelines_2010-08-24.pdf.
21. Colorado Digitization Project Digital Imaging Best Practices Working Group, "BCR's CDP Digital Imaging Best Practices Version 2.0," (Aurora, CO: Bibliographical Center for Research, 2008), http://mwdl.org/docs/digital-imaging-bp_2.0.pdf.
22. "Digitization Standards for Still Images," Smithsonian Institution Archives website, accessed June 5, 2015, http://siarchives.si.edu/services/digitization; World Digital Library, "WDL Digital Image Standards" (Washington, DC: World Digital Library), accessed June 5, 2015, http://project.wdl.org/standards/imagestandards.html.
23. Colorado Digitization Project Digital Audio Working Group, "Digital Audio Best Practices Version 2.1," March 25, 2015, http://sustainableheritagenetwork.org/content/digital-audio-best-practices-version-21; Ian Bogus et al., "Minimum Digitization Capture Recommendations" (working paper, Preservation and Reformatting Section, Association for Library Collections and Technical Services, American Library Association, 2013), www.ala.org/alcts/resources/preserv/minimum-digitization-capture-recommendations.
24. Richard Wright, *Preserving Moving Pictures and Sound* (Helsington, York, UK: Digital Preservation Coalition, 2012): 17; CARLI Digital Collections Users' Group, Standards Subcommittee, "Guidelines for the Creation of Digital Collections: Digitization Best Practices for Moving Images" *Digital Collections: Best Practices for Digital Collections* (Champaign: Consortium of Academic and Research Libraries in Illinois, 2014), 4, www.carli.illinois.edu/sites/files/digital_collections/documentation/guidelines_for_video.pdf.
25. Isaiah Beard et al., *Sound Objects: Recommended Minimum Requirements for Preservation Sampling of Audio* (New Brunswick, NJ: Rutgers University Libraries, 2010), http://odin.page2pixel.org/standards/latest/RUcoreStandards-Audio.pdf; Isaiah Beard, *Digitizing Analog Documents and Images: Recommended Minimum Standards for Archival and Presentation Datastreams* (New Brunswick, NJ: Rutgers University Libraries, 2010), http://odin.page2pixel.org/standards/latest/RUcoreStandards-ScannedImagesDigitalSurrogates.pdf; Isaiah Beard et al., *Video and Moving Image Objects: Recommended Minimum Standards for Archival and Presentation Datastreams* (New Brunswick, NJ: Rutgers University

Libraries, 2010), http://odin.page2pixel.org/standards/latest/RUcoreStandards-Audio.pdf.

26. Alan P. Kefauver and David Patschke, *Fundamentals of Digital Audio* (Middleton, WI: A-R Editions, 2007); Arch C. Luther, *Using Digital Video* (Orlando, FL: Academic Press, 1995); Francis Rumsey, "Appendix I: From Analog to Digital and Back," in *Digital Audio Operations* (Boston: Focal Press, 1991), 215–29.
27. Taz Tally, *Avoiding the Scanning Blues: A Desktop Scanning Primer* (Upper Saddle River, NJ: Prentice Hall, 2001), 67; Margot Note, *Managing Image Collections: A Practical Guide* (Oxford, UK: Chandos, 2011), 54; Forrester High School Computing, "Bitmapped Graphic Calculations," accessed April 2012, http://forrestercomputing.wikispaces.com/Bitmapped+Graphic+Calculations.
28. Techopedia, s.v. "Resolution," accessed April 1, 2015, www.techopedia.com/definition/2743/resolution.
29. Forrester High School Computing, "Calculations (sound)," accessed April 30, 2012, http://forrestercomputing.wikispaces.com/Calculations+%28Sound%29; Nels Johnson, Fred Gault, and Mark Florence, *How to Digitize Video* (New York: Wiley, 1994), 28, 31. Note: Despite being a resource for digitizing video, this book contained a formula for calculating digital file sizes for audio recordings as part of the calculations for video recordings.
30. Cambridge Dictionaries Online, s.v. "Track," accessed April 1, 2015, http://dictionary.cambridge.org/dictionary/british/track; Brylawski et al., *ARSC Guide to Audio Preservation*, 229.
31. John Rice and Brian McKernan, eds., *Creating Digital Content: Video Production for Web, Broadcast, and Cinema* (New York: McGraw-Hill, 2002), 18.
32. Johnson, Gault, and Florence, *How to Digitize Video*, 28; Forrester High School Computing, "Calculations (Video)," accessed April 30, 2012, http://forrestercomputing.wikispaces.com/Calculations+%28Video%29.
33. Forrester High School Computing, "Calculations (Video)"; Johnson, Gault, and Florence, *How to Digitize Video*, 28.
34. Johnson, Gault and Florence, *How to Digitize Video*, 28–29.
35. Forrester High School Computing, "Calculations (Sound)"; Forrester High School Computing, "Calculations (Video)."
36. Eddie Woo, *Calculating File Size: Audio*," MPEG4 YouTube video, 10:32, posted June 4, 2013, www.youtube.com/watch?v=0ctX3z7yzuU; Eddie Woo, *Calculating File Size: Images*, MPEG4 YouTube video, 7:06, posted June 4, 2013, https://www.youtube.com/watch?v=6jIhaAkzOvo; Eddie Woo, *Calculating File Size: Video*, MPEG4 YouTube video, 4:21, posted June 4, 2013, https://www.youtube.com/watch?v=lju4RzSIkqA ; Kenneth R. Koehler, "Storage Requirements," in *Elementary Computer Mathematics*, 2002, http://kias.dyndns.org/comath/44.html.
37. TechTerms.com, s.v. "Byte," accessed April 1, 2015, http://techterms.com/definition/byte; The still image formula has been presented previously at AUTHOR, "Digital Imaging Specification and the Management of Digital Storage Needs" (presentation, Image Resources Interest Group, ALA Annual Conference, Las Vegas, NV, June 28, 2014), http://dx.doi.org/doi:10.7282/T3MG7MXQ; and at Krista White, "Calculating All That Jazz: Linking Technical Specifications to the Management of Digitization Projects" (poster presentation, Research Data Symposium, Columbia University, New York, New York, February 27, 2013), http://hdl.handle.net/10022/AC:P:19180. The formula, as presented was adapted from Tally, *Avoiding the Scanning Blues*; Note, *Managing Image Collections*; and Cunningham, "Bitmapped Graphic Calculations."
38. Colorado Digitization Project Digital Imaging Best Practices Working Group, "BCR's CDP Digital Imaging Best Practices"; Ian Bogus et al., "Minimum Digitization Capture Recommendations"; Isaiah Beard, *Digitizing Analog Documents and Images*.
39. TechTerms.com, s.v. "Bit," accessed April 20, 2015, http://techterms.com/definition/bit; Brylawski et al., 224.
40. "Media Info," MediaArea, 2015, accessed August 3, 2012, https://mediaarea.net/en/MediaInfo.
41. TechTerms.com, s.v. "Kilobyte," accessed April 1, 2015, http://techterms.com/definition/kilobyte.
42. All data and the full description of the experimental design and results are available at URL TBA.
43. Oxford English Dictionary, s.v. "absolute value," accessed October 14, 2015, www.oed.com/view/Entry/679?redirectedFrom=absolute+value#eid4714829.
44. "Download," Audacity, accessed June 27, 2012, http://audacityteam.org/download.
45. Beard et al., *Sound Objects*; Colorado Digitization Project Digital Audio Working Group, "Digital Audio Best Practices"; Bogus et al.,"Minimum Digitization Capture Recommendations."
46. Only the still image formula is used as an example here; the digitization of the audio files was completed in 2005, some seven years before the author began digitizing the JOHP transcripts.
47. Beard, *Digitizing Analog Documents and Images*.
48. This compromise for the standard was made at the suggestion of Isaiah Beard, the Digital Data Curator, who has been an integral part of the JOHP work.
49. Bertram Lyons, Rebecca Chandler and Chris Lacinak, "Quantifying the Need: A Survey of Existing Sound Recordings in Collections in the United States" (New York: AVPreserve, 2015), https://www.avpreserve.com/papers-and-presentations/quantifying-the-need-a-survey-of-existing-sound-recordings-in-collections-in-the-united-states/.
50. Ibid., 19–20.

Appendix. Definition of Terms

Absolute value: "the value of a real number disregarding its sign" where a sign indicates negative or positive value.[1]

Analog/Born analog: a device or system that represents media as continuously variable physical quantities. Analog media cannot be displayed on a computer or uploaded as files without transferring them into digital format.[2]

Audio: sound recordings of any variety.

Bits: "A bit (short for "binary digit") is the smallest unit of measurement used to quantify computer data. It contains a single binary value of 0 or 1."[3]

Bit depth: a unit that measures the amount of information recorded for each pixel in a still image or each sample in an audio file. Bit depth indicates the amount of information about the color of a pixel in an image or the sound level of the wave in a sound file.

Born digital: any recording (or file) that was digitally encoded at the point of creation.[4]

Bytes: "A byte is a unit of measurement used to measure data. One byte contains eight binary bits, or a sequence of eight zeros and ones."[5]

Channels/Tracks: "a part of a magnetic strip onto which sound can be recorded, with several tracks on one magnetic strip."[6] In digital sound, tracks are referred to as "channels." In analog and digital recordings, multiple tracks or channels are usually "mixed down" to create mono (one channel/track) or stereo (two channels/tracks) in the final version of a recording.

Gigabytes (GB): a unit of measurement for data equal to 1024 megabytes.[7]

Kilobytes (KB): a unit of measurement for data containing 1,024 bytes.[8]

Megabytes (MB): a unit of measurement for data equal to 1024 kilobytes and containing 1024^2 or 1,048,576 bytes.[9]

Raster graphics: "Computer graphics employing pixels as the display elements, storing data regarding the component pixels for a given image."[10]

Resolution: "In the computer and media industry, resolution refers mostly to display resolution and the number of picture elements (pixels or simply dots) that can be displayed both horizontally and vertically by a screen. Resolution in this case will then refer to how many pixels the display can produce horizontally (width) and vertically (height). This measure also applies to digital images."[11]

Sampling rate: "how many times per second a continuous (analog) signal is sampled during the digitization process."[12]

Still image/Document: objects such as photographs, letters or manuscripts.

References

1. *Oxford English Dictionary*, s.v. "absolute value," accessed October 14, 2015, www.oed.com/view/Entry/679?redirectedFrom=absolute+value#eid4714829.
2. Sam Brylawski et al., eds., *ARSC Guide to Audio Preservation* (Eugene, OR: Association for Recorded Sound Collections; Washington, DC: Council on Library and Information Resources, 2015), accessed May 29, 2015, www.clir.org/pubs/reports/pub164, 223.
3. TechTerms.com, s.v. "Bit," accessed April 20, 2015, http://techterms.com/definition/bit.
4. Brylawski et al., *ARSC Guide to Audio Preservation*, 224.
5. TechTerms.com, s.v. "Byte," accessed April 1, 2015, http://techterms.com/definition/byte.
6. Cambridge Dictionaries Online, s.v. "Track," accessed April 1, 2015, http://dictionary.cambridge.org/dictionary/british/track.
7. TechTerms.com, s.v. "Gigabyte," accessed April 1, 2015, http://techterms.com/definition/gigabyte.
8. TechTerms.com, s.v. "Kilobyte," http://techterms.com/definition/kilobyte.
9. TechTerms.com, s.v. "Megabyte," accessed April 1, 2015, http://techterms.com/definition/megabyte.
10. Art and Architecture Thesaurus Online, s.v. "Raster Graphics," accessed June 16, 2015, www.getty.edu/vow/AATFullDisplay?find=raster&logic=AND&note=&english=N&prev_page=1&subjectid=300215280.
11. Techopedia, s.v. "Resolution," accessed April 1, 2015, www.techopedia.com/definition/2743/resolution.
12. Brylawski et. al., *ARSC Guide to Audio Preservation*, 229.

Making Metadata Maker

A Web Application for Metadata Production

Myung-Ja K. Han, Nicole E. Ream-Sotomayor, Patricia Lampron, Janet Weber, and Deren Kudeki

Cataloging and metadata operations in academic libraries are focusing on original cataloging of their unique and hidden collections that have not been available to users because of a lack of metadata. However, creating MARC format metadata is an expensive process; libraries need professional catalogers with appropriate experience and knowledge or must train staff to do the work. To improve the cataloging and metadata creation workflow, the University of Illinois at Urbana-Champaign Library developed a web application, Metadata Maker, which allows anyone to create metadata in four different formats, including MARC21 for an online public access catalog, regardless of their familiarity with metadata standards or systems that utilize the metadata. Released as an open source application, Metadata Maker supports diacritics and Unicode non-Roman language encoding, and creates metadata records that ensure discovery and access of unique library collections.

As more resources purchased by libraries come with vendor provided cataloging records or via other libraries through cooperative cataloging initiatives, cataloging and metadata operations in academic libraries are focusing on processing more unique materials and hidden collections that have not been available to users because of a lack of metadata. To provide metadata, libraries generally employ professional catalogers with subject knowledge and appropriate cataloging experience to make these hidden and possibly valuable library collections searchable and discoverable in a timely manner. Recent budgetary issues and a shift in library priorities led to these positions being downsized or eliminated, and libraries must find alternatives to facilitate metadata creation.[1] At the University of Illinois at Urbana-Champaign (UIUC) Library, the loss of professional cataloging positions due to retirements and the increasing volume of resources that need catalog records has imposed a change in cataloging workflows that rely on temporary staff with little to no experience in metadata creation. While the UIUC Library has worked to train these staff to create metadata in MARC format and to use appropriate cataloging software, such as integrated library systems and shared cataloging systems (e.g., OCLC's Connexion), this training is an intensive and time-consuming process. Because of the temporary nature of many of these staff members, the professional catalogers are in a constant state of training new employees, monitoring their work, and providing appropriate feedback to protect

Myung-Ja K. Han (mhan3@illinois.edu) is a Metadata Librarian, **Nicole E. Ream-Sotomayor** (ream2@illinois.edu) is the Foreign Language Cataloging Coordinator, **Patricia Lampron** (lampron2@illinois.edu) is a Metadata Services Specialist, **Janet Weber** (janetw@illinois.edu) is Cataloging Services Coordinator, and **Deren Kudeki** (dkudeki@illinois.edu) is Project Programmer at the University of Illinois at Urbana-Champaign.

Manuscript submitted March 4, 2015; returned to authors May 13, 2015 for revision; revised manuscript submitted June 17, 2015; returned to authors for minor revision September 30, 2015; revised manuscript submitted October 9, 2015; accepted for publication November 9, 2015.

metadata quality. Another option that libraries may consider is the outsourcing of cataloging work to vendors, but that also incurs a substantial cost to accomplish the task.

To improve the productivity of their cataloging and metadata creation workflow, the UIUC Library sought to develop a web application, Metadata Maker, that would allow anyone to create metadata in various formats, regardless of their familiarity with metadata standards or systems that utilize the metadata.[2] With information that is readily available from the item in hand, a user can create quality metadata that ensures the discoverability of resources in the library's various asset management systems including the online public access catalog (OPAC). This paper discusses the emerging need for libraries to have a metadata creation tool that enables metadata creation in different formats, and shares a detailed description of the project development process and the initial user testing results.

Literature Review

The need for an efficient metadata creation tool for backlogs of library materials is steadily increasing as library budgets and experienced cataloging positions are being reduced, and libraries are shifting priorities to electronic resources and digitization efforts. Boydston and Leysen studied the changing nature of the roles and responsibilities of the cataloging librarian by examining thirty-two responses to their survey on the topic.[3] While the study discusses in depth the shift in priorities and skills toward electronic resources and new descriptive standards including non-MARC metadata, it also reports that 60.71 percent of responders indicated that cataloging positions had been eliminated. They cite "budget, reorganization, and retirements" as top reasons for the elimination of these positions.[4] Ithaka S+R Library Survey 2013 results, with responses from 499 academic libraries, indicate that many library directors are experiencing constraints because of limited budgets, while their focus is shifting toward digital preservation and special collections, leaving the purchase and processing of print materials on the back burner.[5] The survey also shows that almost 30 percent of respondents predict a reduction of staff in "technical services, metadata, and cataloging."[6] Of the sixteen categories of "staff resources," the metadata and cataloging staff category was predicted to see the largest decrease.[7]

A decrease in experienced catalogers also brought a change in the roles and responsibilities of cataloging librarians. Sapon-White found that traditional cataloging work has shifted to "paraprofessional staff" and graduate students who required regular cataloging training by catalogers.[8] Sapon-White asserted that while the training program organized for the Oregon State Library was successful, it required one-on-one training and then weekly follow-up training, which spanned a two year period, and only in one aspect of cataloging: subject analysis.[9]

Compounding the issue of staffing and resources for cataloging print materials is the prevalence of backlogged print materials among academic libraries. Jones discusses in depth the problem of hidden collections and inaccessible resources due to the lack of metadata.[10] Not only do unprocessed materials go undiscovered by library users, they also cause errors or duplication in acquisitions, can be lost or stolen, and lead to poor donor relations resulting from "not making collections available in a timely fashion."[11] Citing the results of the 1998 ARL Survey by Panitch, the white paper states that "15 percent of collections on average remained unprocessed or uncataloged" and suggests that libraries must develop policies and workflows for levels of access, while taking into account characteristics of the individual item or collection.[12]

Metadata Maker is designed to allow staff with minimal training in the intricacies of cataloging and metadata standards, but possibly with more subject or language knowledge, to create metadata describing these backlogged materials to provide adequate access to both library users (for research) and staff (for resource maintenance). In its attempt to provide access to library materials in a timely manner, the UIUC Library's Metadata Maker project team had to determine the minimal standard of metadata quality acceptable for bibliographic metadata produced by the application. This minimal set of information was guided by the main goal of the project: to produce bibliographic metadata that facilitates discovery and access of the library resource both by users and librarians who use and manage them.

There is great debate regarding what constitutes quality bibliographic metadata across the library community. However, there is a general consensus that accuracy is important, plus the presence of appropriate access points and subject headings and the usefulness of the record in terms of search and retrieval. Snow reported on academic library catalogers' perceptions of quality catalog records from her study.[13] Referring to interviews and questionnaires completed by cataloging librarians, Snow described aspects of cataloging that determine quality. For the purposes of the study, "quality cataloging" is broken into four categories: technical details of the bibliographic record, adherence to standards, the cataloging process/workflow/staff, and impact on users/accessibility.[14] Of these four, technical details of the bibliographic record were discussed the most when describing quality cataloging, though almost 80 percent of respondents described aspects that represent two or more categories.[15] When asked to rank MARC data fields and subfields, the catalogers surveyed chose the top three data fields as 245$a (Title Proper), 100 (Personal Name), and 650 (Topical Subject Heading), indicating that access points are perceived as a component of a quality record.[16] Roy Tennant's visualization of MARC fields used on book records further supports Snow's findings, which shows that the 040 (Cataloging Source), 245 (Title Statement), 260 (Publication), 300 (Physical Description), and 650 (Subject Added Entry—Topical Term) are the most commonly appearing MARC data fields for such records.[17] Calhoun et al. addressed the concept of "quality data" by surveying both users and librarians on their expectations for search results.[18] One of the key findings of the study highlights librarians' emphasis on duplicate record merging and clean up, indicating the need for bibliographic metadata that have adequate information for matching resources to their proper record, which played an important role in the design of Metadata Maker.

Project Development

The Content Access Management (CAM) unit of the Technical Services Division in the UIUC Library has gained valuable experience in creating metadata for digital resources and special collections during the last several years by implementing various information technologies, notably Extensible Markup Language (XML), Extensible Stylesheet Language for Transformations (XSLT), and simple programming languages including Python and JavaScript. Metadata creators for the Emblematica Online project, a digital portal for all digitized emblem books, are not trained in any metadata standards or systems.[19] An emblem book contains a collection of emblems, which are described as "a symbolic and often enigmatic image," and the image's accompanying text.[20] To provide access to a book and the emblems contained in it, metadata must be created for each emblem. Students and scholars who create emblem-level metadata use a Microsoft Excel spreadsheet as their input form, adding information about the individual emblem to designated rows and columns. Once the spreadsheet describing all the emblems in a book is complete, it is saved in XML Spreadsheet format. Using XSLT, the metadata in the XML Spreadsheet is then transformed to the SPINE metadata standard, a schema specially designed for describing emblem books and their emblems.[21] With the new workflow in place, the library does not need to conduct training on the metadata standard, and system access permissions are not required for each individual working on the project.

Relying on their previous experience with automated metadata creation workflows, a group within CAM, consisting of the Senior Metadata Librarian, the Foreign Language Cataloging Specialist, the Manager of Cataloging Services, and the Metadata Specialist, started the Metadata Maker project with support from an Innovation Grant from the University Librarian in October 2014 that allowed the group to hire a part-time research programmer for the project. The project team sought to build a web-based application that allows anyone to create metadata in MARC21, MARCXML, Metadata Object Description Schema (MODS), and/or HTML page marked up with Schema.org semantics. While the project originally started as MARC Maker, the addition of non-MARC metadata output prompted a change in the project name. The application, currently in version 1.1, cues the user to record information about a resource, such as *Title*, *Name*, and *Keywords*, plus other bibliographic information, and provides options for the user to choose one or more of the output formats listed above. The data are then transformed into the selected format(s) and downloaded to the user's own computer. The data can subsequently be reviewed by another staff member with cataloging experience or a professional cataloger and ingested into the appropriate library system to allow for resource discovery.

Metadata Design

The initial goal of the project was to create minimal-level bibliographic metadata in MARC21 format. Starting with the Library of Congress (LC)'s minimal-level record for books example and Snow's findings, the project team decided that the MARC record should include the data fields 1XX (Main Entry, if applicable), 245 (Title Statement), and keywords used to provide subject access.[22] In addition, data fields 260/264 (Publication, Distribution, etc. (Imprint)), and 300 (Physical Description) are added because this information is readily available from the item. For Metadata Maker to create rich metadata, each subfield is separated from its parent data field in application display so that users can add individual data attributes appropriate for each subfield into a separate element in the application. For example, in the Metadata Maker web form, the Title statement is divided into *Title* and *Subtitle*, and the Imprint statement is divided into *Place of publication*, *Name of publisher*, *Date of publication*, and *Copyright date*. This ultimately enables the application to encode information in each element within the proper MARC subfield with International Standard Bibliographic Description (ISBD) punctuation for the MARC21 format metadata. The application also provides sections to record other information easily gleaned from the item that is being cataloged, such as *Number of pages/volumes* (options are selected from a dropdown menu), *Language*, *ISBN*, and *Edition statement*. Also included in the application are questions that determine whether the item is a work of fiction or literature (*Is this item literature?*), and whether it includes illustrations (*Does the item include illustrations?*).

Because of the user testing findings, additional exploration of controlled vocabulary services, and the UIUC Library's cataloging needs, the elements available in Metadata Maker have changed since its initial release as version 1.0. In version 1.0, the application included names only for

Table 1. Elements Included in the Metadata Maker Application ((R) identifies an element as repeatable)

Element Name	Required/Optional
Title	Required
Subtitle	Optional
ISBN	Optional
Edition statement	Optional
Language	Required
Names (R)	Optional
Name of publisher	Optional
Place of publication	Optional
Country of publication	Optional
Date of publication	Optional
Copyright date	Optional
Number of pages/volumes	Optional
Dimensions	Required
Is this item literature? (If yes, choose from dropdown)	Optional
Does the item include illustrations?	Optional
Keywords (R)	Required
Note to the Cataloger	Optional

authors, and the element was labelled as *Author*. However, the user testing results revealed that testers added additional names available on the item into the *Note to the Cataloger* element. In version 1.1, the *Author* element label was changed to *Name*, and now users can select role information for the name from one of the six roles provided in the dropdown menu: artist, author, contributor, editor, illustrator, and translator. Depending on future needs, additional role values may be added to the dropdown menu. Also changed for version 1.1 is the implementation of Faceted Application of Subject Terminology (FAST) headings in the *Keywords* element, which are mapped to 6XX data fields accordingly. By facilitating FAST headings, users can now choose keywords from the already established controlled subject headings in addition to using any uncontrolled terms. Table 1 shows all available elements a user can input in Metadata Maker.

The application automatically adds default information into the metadata in MARC21 and MARCXML formats. For example, because the application was developed for monographic materials, MARC data fields 336 (Content Type), 337 (Media Type), and 338 (Carrier Type) are populated with the appropriate designators for a book and are added automatically during the transformation process. Data field 040 (Cataloging Source), Leader position 07 (Bibliographic Level), and Leader position 18 (Descriptive Cataloging Form) are also added automatically (see table 2). In the end, the application allows users to create a metadata record that is close to a full-level record if all information is available within the item and recorded accordingly.

Web Application Design

The application is presented as a web form (see figure 1) comprising the elements shown in table 1. There is a question mark icon next to each element that provides a short description with an example when the cursor hovers over it.

Among the sixteen elements, *Title*, *Language*, *Dimensions*, and *Keywords* are required. *Number of pages/volumes* requires a number to be entered under "pages" or "volumes," or the "unpaged" box to be checked, as shown in figure 1. The remaining eleven elements are optional. Though Metadata Maker can record a range of descriptive elements, to make the majority of these elements required would exclude metadata creation for a great deal of material. As the goal of Metadata Maker is to facilitate a more efficient workflow for metadata creation, the project team sought to make the tool flexible enough to work with any materials, and so only the four elements mentioned above are required. Two elements, *Name* and *Keywords* are repeatable—a new box is added when the "+" sign located after the first text box is clicked. All other elements have a check box named "unlisted" beneath the text box, which the user can check when information for that element is not available in the item. During transformation into metadata, if an element is not populated, and "unlisted" is not checked, a message will appear as part of the validation process, as a measure of verifying that the information is not included in the resource. Values for three elements were designed to be chosen from the element's dropdown menu: *Language*, *Country of publication*, and *Type of literature* to ensure consistent and quality metadata. Values for *Language* and *Country of publication* are based on controlled vocabularies available from LC.[23] The dropdown menu for *Number of pages/volumes* allows the user to record either the number of *pages* (for a single volume) or *volumes* (for multivolume sets). The *Number of pages/volumes* element also includes an "unpaged" box, which the user can check if the piece lacks page numbers.

A *Note to the Cataloger* field is provided to allow input of any additional information that should be included in the metadata or requires the cataloger's attention. Information added in this element is initially transformed into the data field 500 (General Note) as a note for the cataloger who will review the metadata in OCLC Connexion. The decision to use the data field 500 instead of the 590 (Local Notes) was based on local practice and workflow. It was noted during testing that the *Note to the Cataloger* field was used for various types of information, such as presence of multiple languages, translations, or series, which may belong either in a public note or other MARC data fields. Since all catalog

Table 2. Mapping from the elements in the application to MARC, MODS, and HTML page marked up with Schema.org semantics (Rows in grey are added into MARC format metadata during transformation process.)

Element in Application	MARC	MODS	Schema.org
Title	245 1? ‡a	<titleInfo> <title>	name (Thing)
Subtitle	245 1? ‡b	<titleInfo> <subTitle>	name (Thing)
ISBN	020 _ _ ‡a	<identifier type="isbn">	isbn (Book)
Edition statement	250 _ _ ‡a	<originInfo> <edition>	
Language	008/35-37	<language> <languageTerm type="code" authority="iso639-2b">	inLanguage (CreativeWork)
Name (author)	100 1_ ‡a	<name type="personal"> <namePart type="given"> <namePart type="family">	Author
	100 1_ ‡e	<role> <roleTerm type="text" authority="marcrelator">author</roleTerm> <roleTerm type="code" authority="marcrelator">aut</roleTerm>	
	100 1_ ‡4		
		<name type="personal"> <namePart type="given">	
Other Names	700 1_ ‡a	<namePart type="family"> <role> <roleTerm type="text" authority="marcrelator">role term</role-Term>	
role="editor" role="contributor"	700 1_ ‡e	<roleTerm type="code" authority="marcrelator">role code</roleTerm>	editor contributor
role="illustrator" role="translator"	700 1_ ‡4		illustrator translator
Name of publisher	264 _1 ‡b	<originInfo> <publisher>	publisher (CreativeWork)
Place of publication	264 _1 ‡a	<originInfo> <place type="text">	n/a
Country of publication	008/15-17	<originInfo> <place type="code" authority="marccountry">	n/a
Date of publication	008/06 008/07-10 264 _ 1 ‡c	<originInfo> <dateIssued>	datePublished (CreativeWork)
Copyright date	008/06 008/11-14 264 _ 4 ‡c	<originInfo> <copyrightDate>	copyrightYear (CreativeWork)
Number of Pages/volumes	300 _ _ ‡a	<physicalDescription> <extent>	numberOfPages (Book)
Dimensions	300 _ _ ‡c	<physicalDescription> <extent>	
Is this item literature?	008/33	<genre>	genre (CreativeWork)
Does the item include illustrations?	008/18 300 _ _ ‡b	<physicalDescription> <note>	
Keywords	653 _ _ ‡a	<subject> <topic>	keywords (CreativeWork)
Content Type	336 _ _ ‡a, b, 2	n/a	n/a
Media Type	337 _ _ ‡a, b, 2	n/a	n/a
Carrier Type	338 _ _ ‡a, b, 2	n/a	n/a
Cataloging Source	040 _ _ ‡a, b, e, c	n/a	n/a
Descriptive Cataloging Form	Leader 18	n/a	n/a

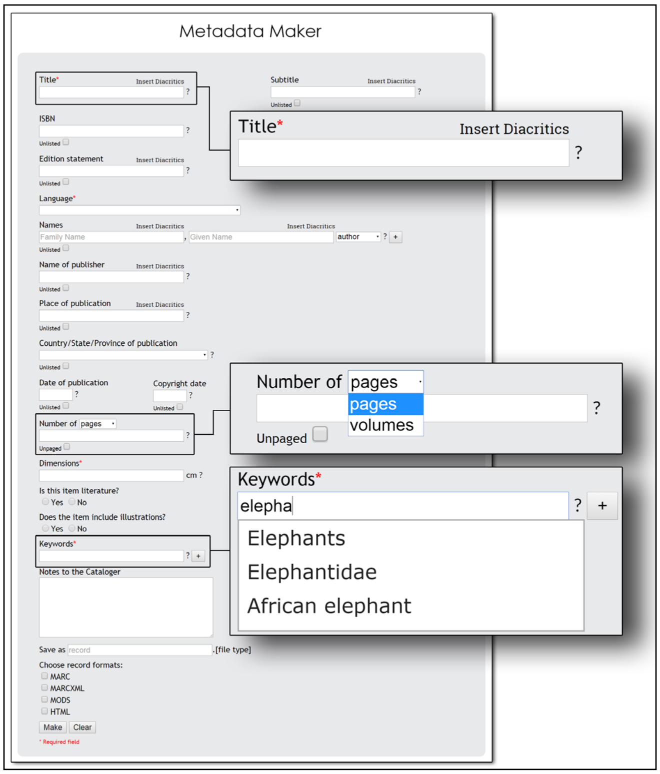


Figure 1. User interface of Metadata Maker, a web application for a metadata creation, with illustrations of required elements, drop down menu options, and FAST suggested terms.

records created by the application will be reviewed by a cataloger, using the data field 500 for all notes works for UIUC's local purposes and makes the application more user friendly for non-catalogers. The MARC field designation can be changed depending on each institution's implementation plan and practice; however, the *Note to the Cataloger* element should be used only when a professional cataloger will review metadata after non-catalogers create the metadata.

After all the available information is added, the user chooses one or more metadata output formats (MARC21, MARCXML, MODS, and/or HTML page marked up with Schema.org semantics), and the metadata downloaded to the user's computer. The application allows the user to name the file, or applies a default file name if one is not provided.[24] Figure 2 shows metadata in MARC format created by staff using Metadata Maker, and figure 3 shows the metadata after the professional cataloger enhanced it.

Metadata Formats Output

Because libraries work with many different metadata standards, the application currently creates metadata in four formats: MARC21, MARCXML, MODS, and HTML page marked up with Schema.org semantics. MARC21 was the first format chosen for output because the majority of metadata created locally from the application will be ingested into OCLC Connexion and Voyager, UIUC Library's integrated library system (ILS). It was decided that MARCXML, MODS, and HTML page marked up with Schema.org semantics should also be offered as output formats by the application. Currently, more metadata services require MARC records in XML format, and having MARCXML without the need to use another transformation process streamlines the workflow. The UIUC Library's digital preservation system requires MODS as its bibliographic metadata standard, so MODS was also selected as one of the options. The HTML page marked up with Schema.org

000 00663cam a2200205ki 450
001 7757597
005 20150602104540.0
008 150602t20122012caua 000 u ful d
040 __ **|a** UIU **|b** eng **|e** rda **|c** UIU
100 1_ **|a** Kamara, Mamadou, **|e** author. **|4** aut
245 10 **|a** Togue's alphabet.
264 _1 **|a** California : **|b** Friends of African Villages Libraries, **|c** [2012]
264 _4 **|c** ©2012
300 __ **|a** 38 pages : **|b** illustrations ; **|c** 21 cm
336 __ **|a** text **|b** txt **|2** rdacontent
337 __ **|a** unmediated **|b** n **|2** rdamedia
338 __ **|a** volume **|b** nc **|2** rdacarrier
500 __ **|a** The item is in Pular.
650 _7 **|a** Pular dialect **|2** fast **|0** (OCoLC)fst01083771
650 _7 **|a** Alphabet books **|2** fast **|0** (OCoLC)fst00805964

Figure 2. MARC21 format record created by staff using the Metadata Maker.

000 00997cam a2200289li 450
001 7226405
005 20150602100244.0
008 130509t20122012caua 000 0 ful d
035 __ **|a** (OCoLC)ocn899141392
040 __ **|a** UIU **|b** eng **|e** rda **|c** UIU **|d** UIU
050 _4 **|a** PL8184.Z9 **|b** P8553 2012
049 __ **|a** UIUU **|o** skip
100 1_ **|a** Kamara, Mamadou, **|e** author. **|4** aut
245 10 **|a** Togue's alphabet / **|c** created by Mamadou Kamara with help from Togue.
264 _1 **|a** California : **|b** Friends of African Villages Libraries, **|c** [2012]
264 _4 **|c** ©2012
300 __ **|a** 38 pages : **|b** illustrations ; **|c** 21 cm
336 __ **|a** text **|b** txt **|2** rdacontent
337 __ **|a** unmediated **|b** n **|2** rdamedia
338 __ **|a** volume **|b** nc **|2** rdacarrier
546 __ **|a** In Pular.
650 _7 **|a** Pular dialect **|2** fast **|0** (OCoLC)fst01083771
650 _7 **|a** Alphabet books **|2** fast **|0** (OCoLC)fst00805964
650 _0 **|a** Pular dialect **|x** Alphabet **|v** Juvenile literature.
650 _0 **|a** Pular dialect **|x** Vocabulary **|v** Juvenile literature.
655 _0 **|a** Alphabet books **|z** Africa.
994 __ **|a** C0 **|b** UIU

Figure 3. The same record after post processing—enhanced by a cataloger with a statement of responsibility, subject headings, classification, and others.

semantics was added as part of the UIUC library's ongoing contribution to Linked Open Data development work. The UIUC library has experimented with linked data by transforming 5.5 million bibliographic records and associated holdings data to linked data using Schema.org semantics that align with OCLC's linked data work.[25] For this experimentation, most MARC data fields and subfields were mapped to Schema.org semantics and transformed by Metadata Maker accordingly.

Transforming Information to Metadata

To transform information submitted to the application into the four metadata formats, three different mappings were created, with one mapping used for both MARC21 and MARCXML (see table 2). For MARC21, since the metadata will be imported into OCLC Connexion, the output format for MARC21 metadata are an MRC (machine-readable) file. Transforming application data to metadata in MARC21 format is more challenging than other output metadata formats because some mappings utilize the information entered to determine the data in the fixed fields. However, because MARC21 and MARCXML have the same structure and the underlying schema is the same, transformation of these two metadata formats are done through similar code. Preparing transformation of information to MODS and HTML is simpler than MARC21 and MARCXML for two reasons: mapping to MODS and HTML do not require conditional mapping (for example, in a MARC record, the first indicator of the data field 245 is determined by the value in the *Names* element); and MODS and HTML are not affected by the white spaces in fixed field 008 and the leader in MARC21 and MARCXML. For MODS and HTML, a template was created for each metadata format and the application adds the supplied information to the corresponding element in the template, according to the mapping. Mappings from the elements provided in the application to MODS and Schema.org semantics were relatively easy due to the application's simple set of elements, each with clearly defined meanings.

Technical Considerations

The application was created in JavaScript, which can handle enough complexity to fulfill the library's needs, and is easy to modify for additional functionality. The JavaScript code runs on the user's computer (the user's web browser must have JavaScript enabled to use the application) and is computationally simple enough that it can run quickly on any computer. The application works best in Google Chrome.

Currently the web application is hosted on one of the web servers in the UIUC Library, which can be used by anyone who knows the URL.[26] All source files and a simple instruction document are available in GitHub with an MIT license, a standard license for open source software, so any institution can use and modify the code for their needs.[27] For institutions that do not have programmer support, Metadata Maker now has an entrance page.[28] By adding institution specific information on the entrance page, Metadata Maker populates this information in the form and adds it into the output metadata as a default value. Modifying the source code is simple enough that after the initial development of Metadata Maker, CAM was able to customize the application for cataloging backlogs of theses and dissertations, data sets, and government documents.[29] Another department in the UIUC Library that is cataloging backlogs of serials has been modifying the code to allow the application to work with serial materials.

Testing the Application

After version 1.0 of the application was stable, the Foreign Language Cataloging Specialist conducted user testing with six CAM staff members. The user testing was focused on the MARC21 format metadata creation since it is the most used metadata format in the UIUC Library. Testers included two student workers, whose daily work is physical processing of materials, two graduate assistants with copy cataloging experience, and two hourly staff who have experience in copy cataloging and catalog maintenance work. Testers were given a variety of monographs in Western European languages, including English, which required original cataloging. With little instruction, they were asked to complete the form for each item and create metadata in all formats. Staff were asked to follow a standard naming convention for their records, and they wrote the file name on a streamer placed in the corresponding book. The files created by each user were collated in a shared network drive for the project. More than 240 records were created as part of this initial test; five of the users created between 5–8 records per hour, while one user created more than 10 records per hour.

After records were created, the Foreign Language Cataloging Specialist evaluated a sample of 88 records created in MARC21 format. It was decided at this point that only MARC21 records would be evaluated because of the format's complexity. The Foreign Language Cataloging Specialist imported all MARC21 records created by testers to a cataloger's local save file in OCLC Connexion and reviewed each record against the corresponding item. To preserve the UTF-8 symbols (for example, a copyright symbol) and non-Roman characters, the import record character set was selected as UTF-8 in OCLC Connexion. For all records, the Foreign Language Cataloging Specialist corrected any incorrect information, fixed errors in both coding and transcription, created subject headings based on the provided keywords, and controlled name headings if authorities were available in the LC Name Authority File

(LCNAF). All records were assigned appropriate classification based on the item, for example, Dewey Decimal, Library of Congress, or a local classification designed for special collections.

Because of the variety of resources used for the test, some materials needed more enhancements to create a complete record than others. On a case-by-case basis, the Foreign Language Cataloging Specialist decided if a record required additional information and upgrading to a full-level record. Upgrading of records was limited primarily to newer acquisitions and included the addition of various fields including statements of responsibility (subfield $c of the data field 245), information about bibliographies and indexes (data fields 504 and 500), language notes (data field 546), etc. After the records were edited, they were added to OCLC's WorldCat database and exported to the library's ILS for normal cataloging workflow.

Table 3. System Errors Identified through Testing

MARC Field	System Error
008/06, 008/07-14	Lack of coding for both publication and copyright dates when dates are the same
Descriptive fields	Lack of diacritic support
100	Incorrect punctuation Relationship designator in ‡c instead of ‡e
245	No support for non-filing indicators
264 _1 ‡a, b, c	Element not mapped if "unlisted" in form
264 _1 ‡c	No implied publication date if only copyright date provided
300 ‡a	No support for unpaged materials
700	Incorrect indicators

Table 4. User Errors Identified through Testing

MARC Field	User Error
008/35-37	Incorrect language selected
008/33	Incorrect literary form selected
008/15-17	Did not select state/province when appropriate
020	ISBN recorded with hyphens
100	Incorrect form of name
245	Title recorded from cover instead of title page Incorrect capitalization Typographical errors
250	Edition statement not recorded Edition statement translated into English
264 _1 ‡a, b	Incorrect information recorded
264 _1 ‡c	Date taken from CIP information Date implied from copyright date
300 ‡a	Lacked preliminary page sequences Incorrect information
300 ‡c	Incorrect measurements for oblong materials
700	Lack of additional names for editors and illustrators

Identified System and User Errors

The Foreign Language Cataloging Specialist maintained a spreadsheet to track the changes made to each record reviewed and noted if the change was made to fix a system or user error. If the change was because of user error, notes about possible training to address the error were included. After review of the records, the list of errors by MARC field were collated in the spreadsheet and divided by error type (e.g., system or user). The system errors can be found in table 3 and the user errors in table 4.

All system errors have been addressed within the application, including support for nonfiling indicators for English and French language materials. Support for all articles of foreign languages listed in appendix C of the RDA Toolkit has been discussed, but there are no plans for implementation soon.[30] Articles of foreign languages not included in the list will be added to the tool when they are identified. Coding of dates has been modified to supply a publication date in brackets in subfield $c of data field 264 if only a copyright date is available, and the fixed fields 008 positions 06 and 07-14 are coded to include both dates. Diacritic support has been added with the use of a pop-up window that allows the user to select the diacritic to be added to the preceding character, similar to the method used in OCLC Connexion. If a user checks the "Unlisted" box for any publication information, the appropriate "[. . . not identified]" phrase is now mapped to the corresponding subfield in data field 264, and an "unpaged" box has been added to the extent element that maps the phrase "1 volume (unpaged)" to subfield $a of data field 300.

The majority of user errors could be addressed by basic training. While the ideal is to keep training to a minimum, simple instructions focused on choosing the preferred source of information and how to record information would increase the quality of metadata created. New users should also be walked through the application to point out the options for literary form (fixed field 008 position 33), to note the difference between the "Place of publication" and "Country of publication" elements and the state/province exception in the latter, and to answer any preliminary questions the user may have. While each element in the application includes help text, providing users with some instruction up front ensures that they know the basic expectations and allows them to easily refer to the help text when a refresher is needed. As with any project or workflow of this type, time spent training and providing context at the outset produces better results and a more engaged user.

A surprising result of the test was a problem that arose when users were asked to create metadata for materials in languages they did not know. While many catalogers have basic bibliographic knowledge of numerous Western European languages, it was incorrect to presume this same level of knowledge in users who will use Metadata Maker. Lack of language expertise resulted in errors with forms of names, transcription (based on the idea that one with language knowledge could more easily catch typographical errors), lack of or incorrect edition and publication information, and keywords that were not useful for the assignment of subject headings. The prevalence of such errors identified during the test phase underscores the need for libraries to select users with language expertise appropriate to the materials to be cataloged with the application.

"Note to cataloger" Field

Along with tracking changes and errors, the Foreign Language Cataloging Specialist also kept a record of the information added to the *Note to cataloger* field, which was used during the test in a variety of ways. When testers were asked to handle materials in unfamiliar languages, they often entered data indicating that they were unsure about a certain piece of information. The element was also used to list additional names appearing on the item and their roles, which some testers did not think should be entered into the *Author* element. While it was expected that additional contributors would be entered under *Author*, which includes the option to add multiple values for *Author*, this use emphasized the need to reconsider the element name and the addition of dropdown boxes to allow the user to select the appropriate role.

In some cases the *Note to cataloger* element was used to enter information that added to the overall quality of the metadata but was not addressed elsewhere in the form, including information about related works, translations, the presence of multiple languages, and multiple places of publication. Because the data in this element is currently mapped to the MARC21 data field 500 (General Note), an institution or library could decide to use the element more systematically and give users instruction on the type of notes to be entered. This could serve to add to the quality of metadata without adding additional elements and mappings to the different metadata formats. Depending on the scope of a given project, these notes could be used by the cataloger to upgrade records, or they could be left in the data field 500 as additional information for both patrons and other catalogers.

Conclusion

Metadata Maker was developed for anyone to create metadata for the discovery and access of hidden collections or backlogs that currently lack descriptive metadata. Initial testing has revealed that the application is easy to use and creates metadata that supports resource discovery, access, and management. The application also supports diacritics and Unicode non-Roman language encoding that would greatly help foreign language cataloging workflows. If the item in hand has the appropriate information, users could potentially create full-level metadata using Metadata Maker in four different formats.

As more metadata and cataloging authority sources make their resources available through application programming interfaces (APIs), the Metadata Maker project team is also working to improve the application by identifying available services and exploring ways to add them in new versions. As a first step, Metadata Maker version 1.1 includes suggested FAST headings as an option for the *Keywords* element, which allow users to add more controlled subject terms into the metadata, greatly improving discoverability through subject-based faceted browse services. The addition of the Virtual International Authority Files (VIAF) service is also being considered for all names included in the application, so that users can add authorized forms of names directly from the Metadata Maker web form. Both of these improvements also allow for future linked data capabilities. Additionally, the WorldCat Metadata API can be added for direct import of metadata into WorldCat, if the UIUC Library decides to take that route in the future. Because the application was created as an open source project, any institution can modify the codes as needed. The project team hopes that Metadata Maker will grow into a tool that is used and improved upon by the library community, and that it will benefit the community as well as library users.

References

1. Jeanne M. K. Boydston, and Joan M. Leysen, "ARL Cataloger Librarian Roles and Responsibilities Now and In the Future," *Cataloging & Classification Quarterly* 52, no. 2 (2014): 229–50; Mattew. P. Long, and Roger C. Shonfield, "Ithaka S+R US Library Survey 2013," accessed May 19, 2015, www.sr.ithaka.org/wp-content/uploads/2015/08/SR_LibraryReport_20140310_0.pdf.
2. Myung-Ja Han et al., "Final Project Report for Making Metadata Maker," accessed March 1, 2015, www.library.illinois.edu/committee/exec/innovation_fund/innovation_fund_proposals/2014-2015/Innovation_Grant_Metadata_Maker_Final_Report.html. See also, http://quest.library.illinois.edu/marcmaker; OCLC is in development of the *Low*

Barrier Metadata Creation tool that also supports minimal-level metadata creation. The tool is being tested by several academic libraries and vendors as of June 2, 2015.

3. Boydston and Leysen, "ARL Cataloger Librarian Roles and Responsibilities Now and In the Future," 229–50.
4. Ibid., 235.
5. Long and Shonfield, *Ithaka S+R US Library Survey 2013*, 29–31.
6. Ibid., 30.
7. Ibid.
8. Richard Sapon-White, "Subject Analysis Training for Cataloging Paraprofessionals: A Model for Ongoing Learning and Support," *Technical Services Quarterly*, 26 no. 3 (2009): 183–93.
9. Ibid., 186–88.
10. Barbara M. Jones, "Hidden Collections, Scholarly Barriers: Creating Access to Unprocessed Special Collections Materials in North America's Research Libraries: A White Paper for the Association of Research Libraries Task Force on Special Collections," (white paper, Association of Research Libraries Task Force on Special Collections, 2003), www.arl.org/storage/documents/publications/hidden-colls-white-paper-jun03.pdf. See also Barbara M. Jones, "Hidden Collections, Scholarly Barriers: Creating Access to Unprocessed Special Collections Materials in North America's Research Libraries," *RBM* 5, no. 2 (Fall 2004): 88–105.
11. Ibid., 4.
12. Judith M. Panitch, "Special Collections in ARL Libraries: Results of the 1998 Survey Sponsored by the ARL Research Collections Committee," (Washington, DC: Association of Research Libraries, 2001), accessed June 2, 2015, www.arl.org/storage/documents/publications/special-collections-arl-libraries.pdf; Jones, "Hidden Collections, Scholarly Barriers," 4.
13. Karen Snow, "A Study of the Perception of Cataloging Quality among Catalogers in Academic Libraries" (PhD diss., University of North Texas, 2011).
14. Ibid., 74–75.
15. Ibid., 98, 129, 160.
16. Ibid., 114.
17. Roy Tennant, "MARC Usage in WorldCat: Zoomable Starburst of all MARC Tags—By Formats: Books," accessed October 1, 2015, http://experimental.worldcat.org/marcusage/viz/starburst.html.
18. Karen Calhoun et al., *Online Catalogs: What Users and Librarians Want* (Dublin, OH: OCLC, 2009), accessed May 29, 2015, www.oclc.org/reports/onlinecatalogs.en.html.
19. University of Illinois at Urbana-Champaign, "Emblematica Online: Resources for Emblem Studies," accessed October 24, 2014, http://emblematica.grainger.illinois.edu.
20. Timothy W. Cole, Myung-Ja K. Han, and Jordan Vannoy, "Descriptive Metadata, Iconclass, and Digitized Emblem Literature," *Proceedings of the 12th Annual Joint Conference on Digital Libraries* (Washington, DC: ACM, 2012): 111–20.
21. "SPINE Metadata Schema," XMLSPY, last updated October 2, 2011, accessed May 29, 2015, http://diglib.hab.de/rules/schema/emblem/emblem-1-2.xsd.
22. Library of Congress, Network Development and MARC Standards Office, "Appendix C—Minimal Level Record Examples," accessed October 24, 2014, www.loc.gov/marc/bibliographic/bdapndxc.html#book; Snow, "A Study of the Perception of Cataloging Quality among Catalogers in Academic Libraries," 114.
23. Library of Congress Network Development and MARC Standards Office, "MARC Code List for Languages," last updated September 8, 2011, www.loc.gov/marc/languages; Library of Congress, Network Development and MARC Standards Office, "MARC Code List for Countries," last updated April 4, 2008, www.loc.gov/marc/countries/.
24. Myung-Ja K. Han, "How to Use Marc Maker," last modified April 9, 2015, https://github.com/dkudeki/metadata-maker/blob/version-1.1/How%20to%20use%20Metadata%20Maker.docx.
25. "University of Illinois Library Catalog Datasets," University of Illinois at Urbana-Champaign Library, last updated October 10, 2014, http://catalogdata.library.illinois.edu; OCLC, "OCLC Adds Linked Data to WorldCat.org," press release, June 20, 2012, www.oclc.org/news/releases/2012/201238.en.html.
26. "Metadata Maker," University of Illinois at Urbana-Champaign Library, last updated on August 27, 2015, http://quest.library.illinois.edu/marcmaker.
27. "MIT License," accessed June 4, 2015, http://choosealicense.com/licenses/mit; GitHub Metadata-Maker, accessed June 4, 2015, https://github.com/dkudeki/metadata-maker.
28. "Metadata Maker Entrance," University of Illinois at Urbana-Champaign Library, last updated on September 23, 2015, http://quest.library.illinois.edu/marcmaker/entrance.
29. "Metadata Maker for Theses," University of Illinois at Urbana-Champaign Library, last updated on September 28, 2015, http://quest.library.illinois.edu/marcmaker/theses; "Metadata Maker for Data Set," University of Illinois at Urbana-Champaign Library, last updated on April 21, 2015, http://quest.library.illinois.edu/marcmaker/dataset; "Metadata Maker for Government Documents," University of Illinois at Urbana-Champaign Library, last updated on April 21, 2015, http://quest.library.illinois.edu/marcmaker/govdocs.
30. "RDA Toolkit, Appendix C: Initial Articles," American Library Association, August 11, 2015, http://access.rdatoolkit.org/rdaappc_rdac-27.html.

An Analysis of Evolving Metadata Influences, Standards, and Practices in Electronic Theses and Dissertations

Sarah Potvin and Santi Thompson

This study uses a mixed methods approach to raise awareness of divergences between and among current practices and metadata standards and guidelines for electronic theses and dissertations (ETDs). Analysis is rooted in literatures on metadata quality, shareable or federated metadata, and interoperability, with attention to the impact of systems, tools, and practices on ETD date metadata. We consider the philosophies that have guided the design of several metadata standards. An examination of semantic interoperability issues serves as an articulation of the need for a more robust ideal moving forward, rooted in lifecycle models of metadata and concerned with the long-term curation and preservation of ETDs.

Sarah Potvin (spotvin@library.tamu.edu) is a Digital Scholarship Librarian at Texas A&M University Libraries. **Santi Thompson** (sathompson3@uh.edu) is the Head of Digital Repository Services, University of Houston.

Submitted August 7, 2015; returned to authors for revision November 3, 2015; revised manuscript submitted December 18, 2015; accepted for publication January 4, 2016.

The authors benefitted tremendously from conversations with our colleagues on the Texas Digital Library Electronic Thesis and Dissertation Metadata Working Group: Kara Long (Baylor University), Colleen Lyon (University of Texas), Monica Rivero (Rice University), and Kristi Park (Texas Digital Library). We are thankful for their expertise and insight; this article would not have taken shape without them. Hannah Tarver (University of North Texas) and Shelley Barba (Texas Tech University) provided invaluable, generous feedback on our draft, as did Amy Rushing (University of Texas at San Antonio) and Brian Surratt, chairs of prior TDL metadata working groups. Many thanks to *Library Resources and Technical Services* reviewers and to Mary Beth Weber for their suggestions and careful editorial guidance.

As theses and dissertations have evolved in format from shelved print resources to electronic files housed in institutional repositories, recordkeeping practices have been developed to account for the description of theses' content and their administration across a lifecycle marked by institutional approval, deposit, publication, and preservation.[1] These practices are based in standards and recommendations issued at institutional, regional, national, and international levels. As Pargman and Palme have argued, "What can and what cannot be expressed when it comes to electronic communication is, in the end, determined by the underlying and in many respects invisible infrastructure of standards that enables (and, at the same time, constrains and restricts) such communication."[2] This paper attempts to raise the visibility of the standards and infrastructure, philosophies and practices that enable and constrain the expression of electronic theses and dissertations (ETDs) as records.

Development and Application of ETD Metadata Standards

The development and application of ETD metadata standards, and the resulting quality, consistency, and interoperability of the metadata produced and exchanged, incur major implications for the discovery and long-term preservation of these unique student works. As Arms et al. asserted, "The goal of interoperability is to build coherent services for users, from components that are technically different and managed by different organizations." They noted that "This requires agreements to cooperate at three levels: technical, content and organizational." Here, we focus primarily on content and organizational aspects of interoperability, aspects that emphasize semantic agreement (content) and "ground rules for access, preservation of collections and services, payment, authentication" (organization).[3] We observe, in particular, failures of semantic interoperability, which distort the meaningful, consistent interpretation of metadata values associated with particular elements.

What forces have proliferated inconsistent metadata, further complicating interoperability? Broadly, we argue that the failures of interoperability, particularly for date-related metadata, are exacerbated by divergent philosophies about the role of metadata, viewed either as primarily descriptive or as a distinct component in the lifecycle management of electronic documents, and are shaped by the constraints enforced by the systems and tools developed to shepherd ETDs. This argument is an attempt to reconcile how philosophies and tools have restrained and expanded metadata practices, and to document the incongruities between reality and ideal. A view of the issue that considers recent history, coupled with close analysis of standards, positions us to identify gaps in the sociotechnical infrastructure, and to understand forces, whether decisions, compromises, or trends, that have separated practice from ideal.

This paper uses a mixed methods approach to illustrate divergent metadata philosophies and the impact of systems, tools, and practices on ETD date metadata. First, we review the historical developments of ETDs and ETD metadata. This analysis of the guiding principles of the ETD movement highlights how practices have changed over time. We then conduct a meta-analysis of various ETD standards and guidelines, designed to show areas of agreement and confusion across these ideals, as well as to indicate the distinctive goals and philosophies underpinning these approaches. Next, we sample data from selected Networked Digital Library of Theses and Dissertations (NDLTD) institutions to better understand the current quality and consistency of ETD date metadata. Finally, we consider how tools affect metadata standards and practice, using DSpace repository software and the coevolution of tools and standards produced by one state consortium as examples. Attention to date metadata is prompted by ETD stakeholders' confusion over the quantity and meaning of dates provided in ETD metadata, and in the interest of analyzing an aspect of metadata with descriptive, technical, administrative, and preservation implications.

By raising awareness of these shortcomings, and the forces behind them, we hope to begin to move closer to and engage with approaches that consider the long-term curation and preservation of ETDs. Our goal, too, is to promote a deeper understanding of the tradeoffs incurred in emphasizing a union catalog model for the discovery and administration of ETDs. These tradeoffs are suggested by the union catalog model itself, which privileges metadata over full-text, as well as by the application of this model for ETDs. Generally, in the union catalog approach, metadata are emphasized as the basis for resource discovery: whether in its traditional form (aggregating records contributed by member institutions into a central database) or its more recent incarnation (aggregating records from multiple repositories automatically, via harvesting protocols), union catalogs unify multiples source into a single record set.[4] In the particular case of ETDs, the union catalog approach has been tremendously successful in enabling search and discovery of ETDs across repositories and countries, providing a low barrier to entry for institutions contributing metadata and users searching across metadata records. The dominant metadata standard for ETD exchange, NDLTD's ETD-MS, is a relatively lean standard, designed to emphasize ease of inclusion.[5] It follows that institutional approaches to ETD metadata that reify NDLTD compliance as the ultimate objective, rather than the most basic format of exchange, may forfeit the rich affordances of these digital objects, including discovery and information retrieval enabled by full-text search.[6]

Additionally, because the union catalog model for ETD exchange emphasizes descriptive metadata and largely ignores administrative, technical, and preservation metadata, a lack of awareness of the limits of the model may threaten ETDs' long-term survival. Administrative, technical, and preservation metadata document the structure of an object and trace its provenance throughout the object's lifecycle. Popular preservation metadata schema, like Preservation Metadata: Implementation Strategies (PREMIS), frequently contain information on the composition of an object (including file size and formats), chronicle important actions and decisions made over time to extend access to an object (including decisions to migrate file formats), and outline specific rights management issues that can determine an object's accessibility.[7] Maintaining this information helps build trust in records by ensuring that they are authentic and reliable. "The Society of American Archivists' *Glossary of Archival Terminology* defines authenticity as 'the quality of being genuine, not a counterfeit, and free from tampering' and reliability as 'the quality of being consistent and undeviating.'"[8] Because the union catalog model focuses primarily on descriptive metadata, it might lack the evidence needed (found in administrative, technical, and preservation metadata) to ensure that libraries have maintained authenticity and reliability. The need for metadata beyond descriptive becomes apparent in some real world scenarios: for example, as institutions migrate content from one repository to another, the descriptive metadata frequently privileged by the union catalog model may prove insufficient in capturing the structure of complex objects, in explaining metadata decisions developed to meet specific system requirements or functionality, or in accounting for an object's administration over time (including, for example, when an embargo expired). Failure to account for technical, administrative, and preservation metadata incurs the risk of limiting functionality in the new system or losing the ability to faithfully render the object. Alternatives to the union catalog model are addressed in greater detail in subsequent sections of the paper.

Metadata Concerns with the Emergence and Growth of ETDs: A Recent History

The roots of the ETD movement extend to experiments in dissertations produced in Standard Generalized Markup Language (SGML) markup in the 1980s, growing out of discussions between UMI (later ProQuest) and the Virginia Polytechnic Institute and State University (Virginia Tech). As Fox described it, meetings in the late 1980s and 1990s brought together the Coalition for Networked Information, the Council of Graduate Schools, UMI, Virginia Tech, SoftQuad, and Adobe (then testing their Portable Document Format), with coordinated efforts and progress with community building and technology.[9] Virginia Tech initiated a requirement that students submit ETDs, rather than printed documents, in 1997, the first institution to do so.

As ETDs moved from theory to practice, the literature emphasized two key areas of promise and innovation in the transition away from print: expression and access.[10] The former considered the possibility that students, now unrestricted by print format requirements, could more fully express their creative and scholarly vision.[11] This hope was wedded to the more pragmatic idea that graduate education would be enhanced by students' mastering those digital production tools necessary to author even a basic ETD.[12] In the latter scenario, the format of ETDs is linked to possibilities of access, and to works distributed, aggregated, and made available worldwide, to wider audiences than bound, shelved volumes had permitted.

The new formats and promise of ETDs posed a challenge to libraries. As Virginia Tech librarian Gail McMillan observed, "Theses and dissertations as electronic files transferred from the Graduate School to the Library may well be the first major source of electronic texts that many libraries and their catalogers will regularly encounter," and the "first significant body of electronic materials [that] regularly requiring cataloging."[13] McMillan identified two goals, based on quality and efficiency, that developed in Virginia Tech's initial efforts to process ETDs: (1) ensuring that "access would be at least as good as it is for a hard copy" and (2) developing workflows and practices to "derive cataloging information from the electronic text and avoid rekeying as much as possible."[14]

These concerns about access and avoidance of redundant labor were taken up in an extensive subsequent literature examining efficiencies in creating bibliographic records for ETDs and developing workflows. The literature reflects an anxiety surrounding the shift from bibliographic records created by expert catalogers to metadata records supplied by ETD authors. Particular attention was paid to the enhancement of author-contributed metadata and cost-benefit analyses of expert-assigned subject headings.[15] As full-text electronic documents associated with bibliographic records, ETDs represented a significant shift from a machine-readable record serving as surrogate for a separately located print item. Lubas observed, "ETDs are full-text searchable in DSpace and other repository systems, so the need for a metadata quality control process or application of a controlled vocabulary may not appear paramount."[16] Yet the union catalog model of ETD discovery, promoted by groups such as NDLTD, continues to rely on metadata, not full-text search, in aggregated discovery environments.[17]

Part of the challenge of cataloging ETDs was specific to the genre of theses and dissertations, rather than the electronic format. As unique items, theses and dissertations, even before the advent of ETDs, prompted special considerations for catalogers. Repp and Glaviano explained in a 1987 article, "As Library of Congress priorities preclude cataloging of even depository copies of dissertations submitted for copyright, no LC cataloging for dissertations appears on the bibliographic utilities, and full responsibility for bibliographic control falls to the degree-granting institution."[18] Local responsibility for creating records, where abundant information was relegated to local fields, took its toll. As McMillan observed in the mid-1990s, "Even the full MARC record for a dissertation is not very robust and often has a local twist, presenting valuable information in a unique format that can be seen only at the originating institution because it is masked to users of OCLC or other centralized cataloging repository."[19] In the pre-ETD era, scholars interested in viewing graduate works either traveled to the holding institution, requested a print copy via interlibrary loan, or viewed a UMI reproduction. Repp and Glaviano described significant barriers to discovering dissertations, barriers that were lessened for the intramural scholar, who was likely to have access to records tailored for local access, locally maintained indexes, or "special shelving arrangements, amenities lost to the extramural scholar."[20]

Irregularities, idiosyncrasies, and local conventions for cataloging theses and dissertations have contributed to ongoing metadata interoperability issues for union catalogs and other shared records. These challenges were magnified and significantly altered as graduate works moved into the sphere of digital delivery and non-MARC metadata.

Models for ETD Metadata: Discovery and/or Curation

As the management of theses and dissertations evolved from print to the electronic environment, those responsible for ETDs focused on creating policies, tools, and workflows address the deposit, access, and preservation of these documents. These included the capacity to manage file formats, support categories of metadata (including descriptive, administrative, structural), assert the rights of authors or publishers, and elucidate access policies. Accounting for data that assists in the management of ETDs was a change in practice for libraries accustomed to emphasizing retrieval and access when generating cataloging records. Greenberg argued that strides have been made toward conceptions of "metadata as structured data about an object that supports functions associated with that object" and noted that repositories, with their connection to "archival or recordkeeping practices," may diverge from goals and metadata types and functions that dominate in libraries.[21] This shift reflects potentially divergent philosophies of metadata: one was founded in a simplified vision of library cataloging approaches and theories grounded in print, seen as emphasizing the record as descriptive surrogate; the second moved toward managing electronic and networked objects and a pressing need to consider long-term access and curation.

As noted above, McMillan observed in the mid-1990s, "Even the full MARC record for a dissertation is not very robust and often has a local twist, presenting valuable information in a unique format that can be seen only at the originating institution because it is masked to users of OCLC or other centralized cataloging repository."[22] Both cataloging and metadata practices are aimed at resource description to facilitate discovery and access. Approaches to ETD metadata that focus exclusively on adherence to the NDLTD union catalog model are the equivalent of cataloging approaches attentive only to OCLC exchange, stripped of the administrative information related to a work's acquisition, circulation, preservation, and access requirements. As discussed in an earlier section, cataloging practices provided the foundation for metadata creation for the first ETDs. In this section, we explore the influence of lifecycle records management in relation to the development of ETD standards and guidelines and address the distinctive goals of describing items and curating ETDs.

The record lifecycle model, popularized by researchers examining the collection, description, and preservation of records, recognized that objects are not static, but are born, change and evolve as they age, and eventually die.[23] Building on this metaphor, the lifecycle model traced important events that took place while the document ages. As technology shaped how records were created, shared, and preserved, information professionals adapted the broad lifecycle model to fit new record keeping challenges. Some frameworks, like the Digital Curation Center's Lifecycle Model, illustrate the iterative roles that curation and preservation play in the long-term maintenance of digital objects (see figure 1).

Researchers have argued for the explicit application of a lifecycle model to metadata, helping us to both understand metadata and create metadata models that complement and embody the lifecycle approach to digital resource management. As Greenberg explained,

> A key reason for using lifecycle concepts for repositories is that digital resources are more mutable and sharable than their physical printed counterparts; and the mutable nature presents a seemingly *organic* object . . . like the digital resource, metadata—in digital form—is more mutable and sharable than traditional cataloging records printed for library card catalogs, or maintained in closed databases.[24]

The lifecycle model, she argued, "not only [has] appeal, but a proven applicability."[25]

Literature on ETD management has also aligned with the lifecycle model. According to the *Guidance Documents for Lifecycle Management of ETDs*, this model has sought to "study and document the progression of digital objects through stages of creation, dissemination, use, update and re-use, storage retention or archiving, and sometimes destruction or disposal, of digital objects."[26] Because of its expansive scope and iterative approach, the lifecycle model approach is well suited to facilitate the processes of acquiring, administering, providing access to, and preserving ETDs. Since the model focuses on an object from creation to either its destruction or disposition in a repository for long-term access and preservation (and further evaluation for retention in the future), it incorporates all of the stakeholders who play a role in the ETD process, including the student/creator, faculty committee, graduate school, university library, and university information technology.[27] The model also accommodates a complex workflow that can allow for simultaneous actions from different contributors.

Review of Standards: Treatment of Dates

As ETD management embraces the lifecycle management approach, ETD standards are developing recommendations that better account for key dates in an ETD document's lifecycle. While we argue that capturing dates in the work's lifecycle is integral to any robust method for administering these materials, ETD standards have not always supported this approach. The earliest ETD standards, which predate the dominance of the lifecycle management model, focused on a philosophy of metadata that emphasized data exchange

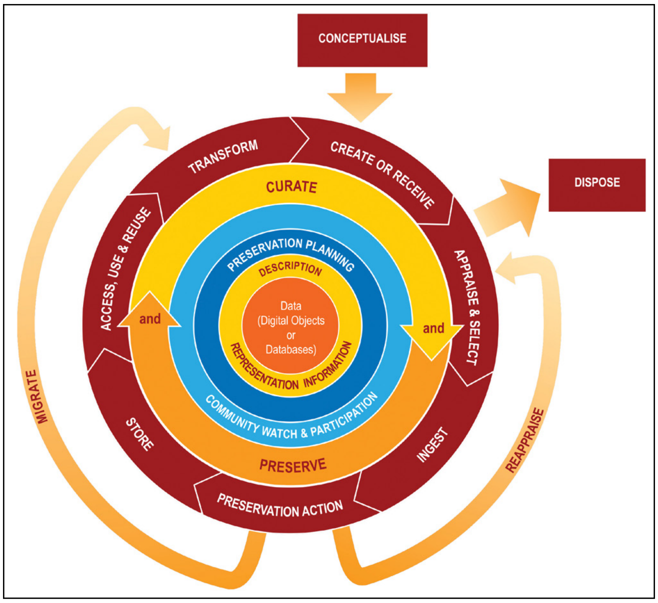


Figure 1. Digital Curation Center Curation Lifecycle Model. Source: Digital Curation Center, University of Edinburgh, "DCC Curation Lifecycle Model," last modified 2014, www.dcc.ac.uk/resources/curation-lifecycle-model.

and discoverability. As such, these standards focused on descriptive metadata elements, such as identifying title, author, and subjects.

These standards allocated one or several fields for capturing date information. For example, NDLTD's ETD-MS: an Interoperability Metadata Standard for Electronic Theses and Dissertations, first published in 2001, served primarily to promote exchange of metadata and the creation of a union catalog among NDLTD member institutions. Early, ambitious attempts by that organization to build an XML DTD standard for encoding the full text of an ETD had been met with resistance from members. Instead, ETD-MS "emerged as a flexible set of guidelines for encoding and sharing very basic metadata regarding ETDs among institutions."[28] ETD-MS identifies one date category that should be recorded, mapped to the DC element date and requiring the user to capture the date "that appears on the title page or equivalent of the work."[29] Created in 2009, the Electronic Thesis Online Service (EThOS) metadata standard, used in the United Kingdom as the basis for a national union catalog, outlined two date fields to capture: the date the thesis is awarded and, if applicable, the date that an embargo on the document ends.[30]

Standards and guidelines evolved to incorporate more than the date of creation or publication; many of these standards embraced another philosophy of metadata that began to emphasize the management and preservation of these objects as records. As such, these standards and guidelines paid greater attention to administrative dates. The 2014 *Guidance Documents for Lifecycle Management of ETDs* identifies four key areas where dates should be recorded: a general date (ideally, publication or graduation date); a date when an embargo ends; birth and death years of the creator to track copyright issues; and dates to track preservation work on the document.[31] In 2014, OhioLINK, a consortium of academic libraries in Ohio, which hosts an ETD Center, established a standard for recording ETD metadata in RDA. Like the *Guidance Documents*, the OhioLINK standard identified four key dates to capture, including copyright date, production (or publication) date, the date the degree was awarded, and the date that any restricted access on the document ends.[32] In 2015, the Texas Digital Library, a consortium of academic libraries in Texas, which hosts a shared metadata repository for ETDs and the Vireo thesis management tool, issued updated metadata guidelines that included an expansive set of dates to capture and publish, including copyright date, graduation date, date of repository ingest, date made public in the repository, date of embargo lift, and author birth date. These guidelines recommended that date fields "be revised and enhanced with increasing reliance on provenance fields to supply additional context for ambiguous date values. Given the likelihood of fields to change meaning over time, explicit encoding of meaningful lifecycle dates in dc.description.provenance fields will help administrators make sense of the myriad dates associated with an item."[33] We consider the coevolution of standards and tools maintained by the Texas Digital Library in a subsequent section of this paper.

While the Thèses Électroniques Françaises (TEF) standard used in France does not explicitly reference the lifecycle model, the standard is exceptional in its articulation of eight areas where dates should be captured. Created to ensure that ETD metadata were both recorded and transferred in the differing contexts and applications used to administer the documents, the TEF guidelines address the holistic approach needed to generate important dates about metadata throughout the workflow. According to TEF, ETDs "reflect three dimensions that characterize the whole theses," including information that documents the "academic work validated by peers," "intellectual work subject to the law of intellectual property," and "an administrative document that governs the grant of a national diploma."[34] The dates captured by the standard reflect both descriptive and administrative metadata. The types of dates associated with this standard include: date of defense, date of publication, author birth date, date of record creation, date of record modification, date that embargo ends, and temporal coverage of the thesis.

Analyzing the variety of date fields reflected in ETD standards and guidelines reveals inconsistencies between the types, definitions, and granularity of dates to be captured by

ETD stakeholders, platforms, and tools. These inconsistencies are shaped by the differing philosophical approaches to metadata promoted in ETD standards. Standards such as ETD-MS focus on broad dates that represent the beginning of a document's lifecycle. This approach makes little data available for the long-term management of ETDs. Other standards leave the interpretation of the date being captured to the creator or ETD administrator (for example, reflecting the date shown on the cover page of the document, the date the document was submitted by the student to the Graduate School, or the date of the student's graduation—which may or may not be the same date depending on institutional policies and specific contexts). The lack of semantic clarity may create values that do not correspond between documents and impede interoperability. Still other standards vary in the amount and the detail of dates to be captured. For example, the TEF standard has specific fields for the date of the thesis defense and the date of thesis approval. Divergent standards guide the production of inconsistent metadata, which impede both the management and discoverability of ETDs.

A Snapshot of Metadata Quality and Consistency: NDLTD Member Institutions

The quality of ETD metadata, including fields associated with dates, presents another barrier to interoperability. Regional and national digital library consortia, like Open Access Theses and Dissertations and the Digital Public Library of America, rely heavily on metadata aggregation to bring disparate collections together into one user interface. For content to be discoverable in an aggregated environment, the metadata must be robust enough to include information queried by the user. Furthermore, records must contain similar fields and valid values in those fields. These properties require records creators to have standardized data entry practices and to use common guidelines for describing content.

With metadata driving how objects are discovered and reused, concerns about maintaining quality metadata have increased.[35] Information professionals have developed categories for analyzing metadata to evaluate its quality. The literature frequently cites Park's metadata quality measurement criteria as one of the most practical benchmarks in metadata evaluation.[36] Park identifies three core categories of metadata quality: completeness, accuracy, and consistency. Completeness of a metadata record "can be measured by full access capacity to individual local objects and connection to the parent local collection(s)."[37] Park notes that completeness does not necessarily correlate with populating a high number of elements with values that describe an object. Instead, it "can be measured by full access capacity to individual local objects and connection to the parent local collection(s). This reflects the functional purpose of metadata in resource discovery and use."[38] Accuracy focuses on the "correctness" of an object's descriptive representation and can address spelling, formatting, and intellectual content.[39] Consistency, according to Park, accounts for data values at the "conceptual" and the "structural" levels. At the conceptual level, consistency "entails the degree to which the same data values or elements are used for delivering similar concepts in the description of a resource."[40] At the structural level, it addresses "the extent to which the same structure or format issued for presenting similar data attributes."[41] Date values, expressed in a variety of ways, including natural language (January 1, 2015) and ISO 8601-compliant (2015-01-01), are examples of structural-level consistency. Collectively, these criteria provide information professionals with a framework to assess existing data and descriptive practices.

Drawing on this framework, we analyzed the consistency and accuracy of date elements across institutions to evaluate the quality of ETD metadata. We conducted an environmental scan of metadata records from sixteen NDLTD members. We harvested records from institutions' digital asset management systems, including DSpace, Digital Commons, and homegrown repositories, using OAI-PMH requests. We documented the categories, frequency, and purposes of dates being captured and made accessible by NDLTD member institutions. Our approach relies on sampling to provide insight into the current state of metadata quality related to dates. This approach requires close interpretation to match dates in records with their semantic meaning. Because ETD records are typically produced using tools that assure regularity, the dates included in these random samples are likely to be repeated across collections. However, readers should not assume that the information reflected in figure 2 necessarily reflects the practices of each institution.

The environmental scan revealed a diverse array of dates being captured by NDLTD member institutions. We analyzed one record per institutions. The number of date fields used by institutions varied from as little as one to as many as twelve.

Most dates conform to ETD-MS, including dc.date, dc.date.available, and dc.description.provenance. Complying with this standard promotes consistency among key dates in the ETD lifecycle, allowing high quality metadata (at least in relation to consistency) to be shared among numerous systems and libraries.

Divergences from the most common elements referenced in the previous chart occurred partly because repository systems generated different date fields over time. Two date fields, dc.date.accessioned and dc.date.issued, were used interchangeably to denote the date that content was deposited in a particular repository. DSpace metadata recommendations note that versions before 4.0 supported dc.date.issued for tracking an object's entry into the

Table 1. Comparison of Date Fields in ETD Metadata Standards and/or Guidelines

Source	Date Field	Field Definition
Networked Digital Library of Theses and Dissertations ETD-MS 1.1 (2009)	dc.date	The date "that appears on the title page or equivalent of the work"
EThOS UK ETD (n.d.)	dcterms:issued	The date the thesis was awarded
	uketdterms:embargodate	The date that an embargo on a document ends
Guidance Documents for Lifecycle Management of ETDs (2014)	date	Publication date. Graduation date.
	embargo lift date	"the metadata should include information sufficient to allow a repository system to know the date upon which the embargo is lifted."[i]
	creator's birth and death years	"Knowing the birth and death dates of the creator and the year in which the ETD was created will help to calculate and determine the copyright status."[ii]
	preservation event date/time	The date when objects are altered by administrators
OhioLINK Standard for Cataloging ETDs in RDA (2014)[iii]	264 #4 $c © [year]	"Copyright date, if available. (RDA 2.11). Optional if there is a publication date."[iv]
	264 #1 $A $c [year]	Publication date
	500 ##$a [year]	"Quote 'Year and Degree' information from OhioLINK ETD Center website."[v]
	502 ## $d [year]	Degree granted date ("calendar year in which a granting institution or faculty conferred an academic degree on a candidate")
	506 ## $a Full text release delayed at author's request until [year month day]	Restriction on access—Full date that an embargo on the document ends
Thèses Électroniques Françaises 2.0 (2006)	dcterms:dateAccepted	Date of thesis defense
	dcterms:issued	Date of publication
	tef:dateNaissance	Author birth date
	dcterms:temporal	Temporal coverage
	metsRights:ConstraintDescription	Date that embargo lifts
	mets:metsHdr CREATEDATE mets:dmdSec ID="CREATED"	Date of record creation
	mets:metdsHdr LASTMODDATE	Date of record modification
TDL Descriptive Metadata Guidelines for ETDs 1.0 (2008)	mods:dateCreated	"The date the student graduates or the date the degree is conferred"[vi]
	mods:dateIssued	"The date the ETD is released to the public."[vii]
	mods:nametype="personal" mods:namePart type="date"	Birth year of author
	mods:nametype="personal" mods:namePart type="date"	Birth year of advisor
	mods:nametype="personal" mods:namePart type="date"	Birth year of committee member
	mods:recordCreationDate	"month, year, and day of the creation date of the record"[viii]
	mods:recordChangeDate	"month, year, and day of the change date [of the record]"[iv]

i. Alemneh, et. al., *Guidance Documents*, 6–3.
ii. Ibid.
iii. While not explicitly stated in the standard, an appendix to the standard, "ETDs in RDA template, as of Oct. 2014; KSU example," includes the dates (including day, month, and year) that the record was entered and replaced.
iv. OhioLINK, "Standards for Cataloging Electronic Theses and Dissertations."
v. Ibid., note included in "ETDs in RDA template, as of Oct. 2014; KSU example" appended to standard.
vi. Texas Digital Library, "Descriptive Metadata Guidelines," 12.
vii. Ibid.
viii. Texas Digital Library, "Descriptive Metadata Guidelines," 17.
ix. Ibid.

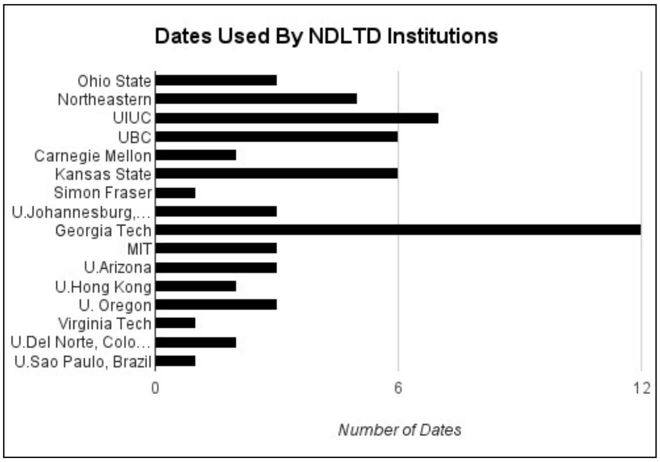


Figure 2. Dates Used by Selected NDLTD Institutions

repository, while DSpace versions at 4.0 or higher supported dc.date.accessioned.[42] This change has direct implications on the quality of ETD date information. If institutions migrate to a newer version of DSpace but fail to transfer values from dc.date.issued to dc.date.accessioned, they store and disseminate inconsistent date elements throughout their repository. These inconsistencies decrease metadata quality.

Additionally, divergences existed because local metadata practices incorporated the usage of unique date elements to describe ETD temporal content. Popular dates among the ETD lifecycle received unique fields for some institutions. These fields included the following:

- dc.date.graduation
- dc.date.graduationmonth
- dc.date.published
- dc.date.updated
- dc.dateAccepted
- "Available in [name of repository]"
- "Date Deposited"
- "Embargo Period" portion of metadata record header
- date stamp in metadata record header

Multiple instances of date fields for graduation date, embargo date, and the content approval date suggest that these kinds of dates present unmet needs among metadata creators, digital repositories, and/or metadata guidelines. Because common date elements (like those ones listed in table 2) may not adequately address the rationale for these unique fields, future metadata guidelines should identify ways to accommodate some of the temporal data being captured in these local fields. Until this occurs, the proliferation of local date elements fosters inconsistent and inaccurate uses of temporal fields and compromises the overall quality of ETD metadata.

Understanding the consistency and accuracy of ETD date information becomes more complicated when analyzing the relationships between the types of dates captured by NDLTD institutions and the frequency with which they are used. Table 4 tracks the type of date used by NDLTD institutions, how often the institution used each type of date, and the date element where they recorded the temporal information. The table divides the latter information into two categories: common uses of the elements (used by over half of the sixteen NDLTD member institutions) and "localized" uses (used by fewer than half of the NDLTD institutions surveyed).

Inconsistent practices between even the most frequently used date type (the date an embargo ended, the date an object is published to the digital repository, and the date an object is submitted) suggest that future metadata guidelines should address some of the more specific ETD temporal data to promote more consistent and accurate uses of date elements. The varying ways that institutions convey the copyright date (dc.rights, dc.date.copyright, dc.description) also complicates description and accessibility, as some institutions repurpose the value in this date for other important administrative functions (including determining embargo start and end dates). Finally, the lack of guidance for graduation date continues to lead to the creation of localized fields, which further impede consistency across NDLTD institutions.

How Tools have Shaped de facto Standards

We have alluded to the influence of tools and systems such as repositories in the production of metadata. Our study of metadata standards and resulting practices would be incomplete without an examination of the influence of tools and systems in the development of de facto standards. In this final section, we consider the coevolution of tools and standards, concluding with the case of the Texas Digital Library (TDL).

Access platforms, the digital asset management systems or repositories into which documents and records are ingested, serve as influential factors in the creation and management of ETD metadata. These systems shape the de facto metadata standards for ETDs through automated processes of metadata creation and assignment, even as they are integrated into a wider system of Internet standards and protocols (like OAI-PMH) for discovery, persistence, and aggregation. Given our observation that the lack of definitional clarity in standards may create values that do not correspond between documents and impede interoperability, how do the systems used to ingest, manage, and steward ETDs reinforce, shift, or ameliorate these issues? How do the constraints of tools shape ETD management?

In some cases, it proves impossible to square the ideal of platform-neutral standards with the reality of platform

Table 2. Common Date Elements Used by NDLTD Institutions

Metadata Field	Definition[i]
dc.date	A point or period of time associated with an event in the lifecycle of the resource.
dc.date.available	Date (often a range) that the resource became or will become available.
dc.date.copyright	Date of copyright [dateCopyrighted].
dc.date.created	Date of creation of the resource.
dc.date.issued	Date of formal issuance (e.g., publication) of the resource.
dc.description	An account of the resource.
dc.description.provenance	A statement of any changes in ownership and custody of the resource since its creation that are significant for its authenticity, integrity, and interpretation.
dc.identifier.bibliographicCitation	A bibliographic reference for the resource.
dc.rights	A legal document giving official permission to do something with the resource.

i. "Section 2: Properties in the /terms/ namespace," Dublin Core Metadata Initiative, http://dublincore.org/documents/dcmi-terms/#H2.

Table 3. System-Generated Date Elements

Metadata Field	Definition
dc.date.accessioned	Date the repository took possession of the item.[i]
dc.date.issued	Date of formal issuance (e.g., publication) of the resource.[ii]

i. See DuraSpace, "Metadata and Bitstream Format Registries," DSpace 4.x Documentation, 2014, accessed February 8, 2016, https://wiki.duraspace.org/display/DSDOC4x/Metadata+and+Bitstream+Format+Registries.
ii. See "Section 2: Properties in the /terms/ namespace," in DCMI Usage Board, "DCMI Metadata Terms," Dublin Core Metadata Initiative website, accessed August 7, 2015, http://dublincore.org/documents/dcmi-terms/#H2.

constraints. Metadata manuals specify that, when developing a metadata application profile, one must consider the repository or content management.[43] Institutions make design decisions and select standards based on repository functionality. Yet researchers have argued for "the importance to reliable digital preservation management of . . . the practice of packaging digital objects in a repository-independent manner."[44] These decisions are particularly problematic when the repository-based access copy is the basis for the digital preservation copy. The adjustments made in metadata creation to conform to repository functionality belie the promise of repository-independent digital packages.

Inevitably, the dates associated with ETDs are shaped by the tools used to manage them, as the Vireo ETD submission system and DSpace demonstrate. DSpace, in its function as a core component of ETD management and publication, has contributed to the development of de facto standards that rely on DC and the ETD-MS Thesis schema.[45] As of 2015, DSpace only supports flat, non-hierarchical metadata schema. As the TDL case study will illustrate, this constrained functionality hastened the abandonment of MODS as the TDL ETD schema, particularly as TDL moved to a reliance on OAI-PMH for harvesting metadata into a portal of TDL ETD metadata, and sought compliance with the ETD-MS metadata standard.

But what are the broader implications of DSpace's emphasis on flat metadata, which has, since its 2004 launch, centered around a DC-dominated library application profile? In considering what level of description was adequate to enable discovery or administration, experts have continually expressed doubt about DC, but ease of use and functionality have hastened adoption. In a generalized critique from 2003, Martin Dillon described "three majors causes that can be adduced for the less than enthusiastic adoption of the library world of the Dublin Core": its "incompleteness," the lack of documentation or agreed-on standards for filling the fields, and "slow adoption."[46] Yet the use of unqualified DC for ETDs, Lubas argued, proliferated because of institutional repositories and OAI-PMH.[47] She observed,

> While during the early days the use of a simplified metadata element set such as Dublin Core may have seemed limiting, over the course of a decade of experience with electronic theses and dissertations metadata reveals that blending the use of qualified Dublin Core with harvesting and crosswalks, plus creating tools to encourage better results from author-generated metadata have proved useful.[48]

On a more granular level, DSpace has affected the dates that are included in object-level metadata. DSpace versions

Table 4. Type, Frequency, and Uses of Date Metadata at Selected NDLTD Institutions

Type of Date	Instances (*N* = 16)	Common Metadata Uses (> half of instances):	Local Metadata Uses (< half of instances):
Date embargo ended	12	dc.date.available	dc.date “Embargo Period” portion of metadata record header “Available in [name of repository]”
Date object published in the digital repository	12	dc.date.issued dc.date.available dc.description.provenance	dc.date dc.date.accessioned dc.date.published date stamp in metadata record header
Date object submitted (including to the digital repository)	10	dc.date.submitted dc.date.accessioned dc.description.provenance	“Date Deposited”
Date of degree or graduation	6		dc.date.created dc.date.graduation dc.date.graduationmonth dc.date.published dc.identifier.bibliographicCitation
Date of copyright	5	dc.rights dc.date.copyright dc.description	
Date of approval	2	dc.description.provenance dc.description	
Date of metadata record creation	2	dc.date.created	dc.date.submitted
Date object accepted by academic department	1	dc.dateAccepted	
Date of license agreement	1	dc.description.provenance	
Date of metadata record modification	1	dc.date.updated	
Date object withdrawn	1	dc.description.provenance	

before 4.0 automatically assigned values to dc.date.issued if items lacked values for that element, indicating prior publication. Confusion over this automatic assignment was not limited to ETDs. In 2013, Google and GoogleScholar alerted DuraSpace, the organization that oversees DSpace community development, that the automatic assignment of dc.date.issued (intended as the formal date of publication) as the value of the date of ingest was causing their web crawlers to inaccurately index publication dates. Google reported seeing “repositories, where 30–50 percent of their items all have the same ‘dc.date.issued,’ as those items were all imported on the same date.”[49] Rodgers noted that the automatic assignment of ingest dates for dc.date.issued was built into the system with a rationale in mind: “The bedrock use-case for DSpace was not published articles, but ‘grey lit’ (born digital content from an institution that was **not** in the official scholarly record): for this sort of content, appearance in the IR essentially **is** the equivalent of publication.”[50] With DSpace 4.0, the software stopped automatically assigning the date of accession as dc.date.issued, and the documentation specified that dc.date.issued, defined as date of publication, should be supplied by metadata creators.[51] At the time of publication, the DSpace community had yet to resolve the thorny issue of consistency in dc.date.issued for collections of items ingested before and after 4.0. As an outstanding card indicates: “we still need a way to help individual DSpace sites to locate any existing, possible inaccurate ‘dc.date.issued’ values.”[52] For some ETDs, the date of ingest into DSpace would constitute a date of publication; for others, ingested into DSpace under embargoes, the date of publication is distinct from ingest.[53] By relying on an automated feature of DSpace, those standards that equated ingest with publication failed to anticipate and account for these use cases.[54]

The Texas Digital Library, which develops and hosts the Vireo ETD thesis submission tool and an ETD metadata standard, furnishes an interesting case study in coevolution. The TDL ETD descriptive metadata standard was first developed as the basis of a union catalog of ETDs from TDL member institutions, introduced in the form of a shared ETD metadata repository. In 2005, TDL tasked a working group with “developing a common [descriptive] metadata standard that would allow members to share metadata in the TDL repository.”[55] The working group, rejecting the

Dublin Core expression of ETD-MS as flawed, issued an ETD-specific application profile for the Metadata Object Description Schema (MODS), which brought ETD-MS elements into MODS.[56] Rationalizing this decision, authors of the recommendation specifically referenced the limitations inherent to DSpace repositories and OPACs; convinced that the tools to manage MODS would soon be developed and adopted, the working group emphasized that more robust, structured schema.[57]

TDL used the MODS application profile as the basis for Vireo, which, in addition to supporting ingestion, verification, and publication of ETDs, would channel institutions into the production of more consistent metadata.[58] Vireo was—and is—typically used to generate deposits into the popular open-source repository platform DSpace, where materials are published, embargoed, stored, and integrated into preservation systems. But Vireo's application was limited by its reliance on MODS, DSpace did not natively support; in DSpace, MODS files were inactionable, stored as bitstreams. TDL had developed both the standard and the tool needed to support the standard, but the shared ETD metadata repository, based in DSpace, operated primarily on flat Dublin Core. In practice, then, active metadata did not align with either the TDL or NDLTD standards.

In 2008, TDL organized a new working group to address the gap between the MODS application profile and members' increasing reliance on DSpace and to bring TDL metadata into compliance with NDLTD. The resulting guidelines focused primarily on the bibliographic elements needed to support aggregation of ETD metadata among the various member institutions.[59] While continuing to position MODS as the canonical schema, the 2008 standard introduced mappings to ETD-MS, expressed as Dublin Core and a flat "thesis" extension. The guidelines dictated that the DC schema mappings "are provided only to assist participants in meeting DSpace requirements, and are not a recommendation to provide qualified Dublin Core as the primary descriptive metadata schema."[60] New, too, in the 2008 guidelines were explicit references to the metadata that would be generated by or necessitated by DSpace.[61]

Over the intervening six years, Vireo was further developed for the management of complex submission and approval workflows. Interestingly, metadata related to dates began to proliferate, in violation of TDL guidelines. When, for example, a student clicked through to approve a license, that action and date were stored as metadata and included as part of the item record upon publication. The tool increased the number of dates generated and retained during the student's submission of the ETD, its approval by committee members, graduate offices, and other required stakeholders, and its ingest into the institution's digital repository. Some of these fields provide supplemental information that can aid the ETD curation process, including the student's submission date, the student's license agreement date, the approval date from the student's committee chair, and embargo beginning and end dates. Vireo generates this metadata, which is largely administrative in nature. Additionally, several institutions that used Vireo observed that the tool had, at some point, stopped generating MODS files and had changed the way that date fields were populated.[62] Vireo's metadata output, consistent as it was, constituted a de facto standard.

In 2014, in recognition of a growing divergence between its tool and standard, TDL commissioned a new working group to analyze and update the standard. Guidelines issued in 2015, while continuing to emphasize descriptive metadata, were increasingly attentive to lifecycle concerns and advocated for more robust technical, administrative, and preservation metadata.[63] As with the 2008 iteration of the standard, the 2015 guidelines advocated for repository-neutrality while tailoring recommendations to DSpace.[64] In a departure from the 2005 and 2008 guidelines, these guidelines did not include a MODS application profile.[65] Continued work is underway to align Vireo development with these new standards.

The tools applied over the course of an ETD's lifecycle are not neutral: they were developed by particular groups, with specific use cases, stakeholders, and goals in mind. These tools, which may be influenced by divergent metadata or stewardship philosophies or reflect design decisions made by those who commissioned, built, or guided their development, constrain and shape ETD metadata. In instances where formal standards proved an awkward fit with available tools, we have observed the development of displacing de facto standards, which complicate existing concerns around interoperability.

Conclusion

In this paper, grounded in literatures on metadata quality, interoperability, and standards, we have coupled research into the history of ETDs and the recent evolution of the Texas Digital Library's ETD standard and tool with close readings of institutional metadata records and a meta-analysis of ETD standards. In so doing, we have sought to initiate a conversation around the generation, maintenance, and evolution of ETD metadata. Our findings highlight distinctions between ETD metadata standards—and the philosophies and goals that underpin these standards—and provide insight into the ETD metadata produced at NDLTD institutions. This exercise has identified a proliferation of fields, without standard definitions, whose interpretation requires close human intervention. Given the erosion of meaningfulness that accompanies diverse and sometimes dissonant metadata standards and practices, we need ways for dates to "speak" and relay their meaning. Possibilities include (1) implementing clearer field and display labels in repository user interfaces; (2) adding clarifying comments in OAI exports; (3) making institutional application profiles more clearly accessible; (4) developing narratives around dates and placing them in description elements; (5) integrating meaningful local fields that are crosswalked into DC, ETD-MS, or other namespaces; and (6) adjusting existing schema and standards to incorporate commonly used or needed date fields.

We have emphasized that our examination of metadata practices serves as a snapshot. Larger-scale or longitudinal investigations are needed to establish statistical significance, which could inform data-driven decisions around the variety, meaningfulness, and interoperability of dates we capture.

Our analysis has shown that ETD metadata has been shaped by forces related to differing philosophies of metadata and the tools and systems that frequently assist in the process of acquiring, managing, and disseminating ETDs. Dominant standards have emphasized a union catalog model, with descriptive metadata as the basis for federated search. ETD-MS is a lean exchange standard that serves as the basis for the NDLTD union catalog; the standard was formulated as "a flexible set of guidelines for encoding and sharing very basic metadata regarding ETDs among institutions."[66] Institutions seeking to optimize the management and description of ETDs must look to more robust standards and models, from which ETD-MS metadata can be derived. We hope, here, to have provided an argument toward a lifecycle metadata model—a model more attuned to the long-term curation of these unique digital objects.

References and Notes

1. The authors refer to both theses and dissertations as "theses" or ETDs (electronic theses and dissertations).
2. Daniel Pargman and Jacob Palme, "ASCII Imperialism," in *Standards and Their Stories: How Quantifying, Classifying, and Formalizing Practices Shape Everyday Life*, edited by Martha Lampland and Susan Leigh Star (Ithaca: Cornell University Press, 2009), 191.
3. See William Y. Arms et al., "A Spectrum of Interoperability: The Site for Science Prototype for the NSDL," *D-Lib Magazine* 8, no. 1 (January 2002), accessed August 7, 2015, www.dlib.org/dlib/january02/arms/01arms.html; As Witten, Bainbridge, and Nichols note, interoperability is both created and thwarted by sociotechnical forces, networked systems "did not come about by accident; they required the development of common understandings about the nature of data formats. It is these communication protocols that allow the reach of digital libraries to extend across our networked world and to interoperate"; Ian H. Witten, David Bainbridge, and David M. Nichols, *How to Build a Digital Library* (Burlington, MA: Morgan Kaufmann, 2010), 343. The authors observe: "Creating and sharing quality metadata is not a straightforward task. . . . Although the technologies may be relatively simple, this is only a *necessary* condition for success, and without the associated human support it will not be sufficient" (350).
4. Mary S. Woodley, "Crosswalks, Metadata Harvesting, Federated Searching, Metasearching: Using Metadata to Connect Users and Information," in *Introduction to Metadata*, 2nd edition, ed. Murtha Baca (Los Angeles: Getty Research Institute, 2008): 38–62; Caplan described union catalogs as enabling search interoperability between multiple sets of records, whether through the ongoing compilation of a stable "central searchable catalog," the maintenance of a unified union catalog from which local records are copied, or the generation of a "pseudo-union catalog" that builds a virtual index over multiple sets of records. See Priscilla Caplan, *Metadata Fundamentals for All Librarians* (Chicago: American Library Association, 2003), 34.
5. As a press release explains, the NDLTD Global ETD Search, launched in July 2015, "allows researchers to find ETDs based on keyword, date, institution, language and subject." "NDLTD Announces Global Electronic Thesis and Dissertation Search," (July 6, 2015), accessed August 7, 2015, www.ndltd.org/events/news/ndltdannouncesglobalelectronicthesisanddissertationsearch.
6. Woodley indicated several concerns with union catalog implementation, including interoperability issues, concerns that aggregated records were presented with inadequate local context, and variability in how service and data providers expose or add value to aggregated metadata. See Woodley, "Crosswalks, Metadata Harvesting, Federated Searching, Metasearching," 46–48. In the ETD context, institutions that host ETDs are data providers, while the NDLTD, as an aggregator of these documents' metadata, functions as a service provider. As Arms noted more than 15 years ago, "Full text and fielded searching are both powerful tools, and modern methods of information retrieval often use the techniques in combination." William Y. Arms, *Digital Libraries*, digital edition (Cambridge: MIT Press, 2000), www.cs.cornell.edu/way/diglib/MS1999/index.html.
7. PREMIS Editorial Committee, *PREMIS Data Dictionary for Preservation Metadata, version 3* (Library of Congress, November 2015), www.loc.gov/standards/premis/v3/premis-3-0-final.pdf.
8. Richard Pearce-Moses, "Authenticity," in *Glossary of Archival and Records Terminology* (Society of American Archivists, 2005), accessed December 16, 2015, www2.archivists.org/glossary/terms/a/authenticity; Richard Pearce-Moses, "Reliability," in *Glossary of Archival and Records Terminology,* accessed December 16, 2015, www2.archivists.org/glossary/terms/r/reliability.
9. Edward A. Fox, "Preface," in *Electronic Theses and Dissertations: A Sourcebook for Educators, Students, and Librarians*,

ed. Edward A. Fox et al. (New York: Marcel Dekker, 2004), iii–viii.

10. Communication of drafts and training in preparing electronic documents were also cited as advantages. See Gail McMillan, "Electronic Theses and Dissertations: Merging Perspectives," co-published simultaneously in *Cataloging & Classification Quarterly* 22, no. 3–4 (1996): 105–25 and in *Electronic Resources: Selection and Bibliographic Control*, edited by W. Pattie Ling-yuh and Bonnie Jean Cox (Philadelphia: Haworth, 1996), 105–25.
11. McMillan recalled that when the Associate Dean of the Graduate School at Virginia Tech approached the library about access to ETDs, they "presented several reasons for providing students with the opportunity to prepare electronic dissertations and over time several more reasons became clear." First on the list: "Greater freedom for authors to demonstrate creatively the result of their independent research" (106). McMillan, "Electronic Theses and Dissertations."
12. Early in the ETD movement, scholars of the movement assessed this expressive potential and predicted the rise of enhanced ETDs as a new genre, with the potential, for authors of truly technologically advanced works, to secure a hiring advantage. See John L. Eaton, "Enhancing Graduate Education Through Electronic Theses and Dissertations" and Seth Katz, "Innovative Hypermedia ETDs and Employment in the Humanities," both in *Electronic Theses and Dissertations: A Sourcebook for Educators, Students, and Librarians*, edited by Edward A. Fox et al. (New York: Marcel Dekker, 2004), 1–7. Matthew Kirschenbaum advocated for expression by differentiating between ETDs that presented no expressive advantage over print (the "plain vanilla" ETD—which "need avail itself of no method or presentation or organization that could not be duplicated on paper") and a "multigraphic" thesis and dissertation—which, "*self-conscious of its medium* . . . uses the electronic environment to support scholarship that could not be undertaken in print." See Matthew G. Kirschenbaum, "From Monograph to Multigraph: Next Generation Electronic Theses and Dissertations," in *Electronic Theses and Dissertations: A Sourcebook for Educators, Students, and Librarians*, edited by Edward A. Fox et al. (New York: Marcel Dekker, 2004), 19–32 (italics original).
13. McMillan, "Electronic Theses and Dissertations," 124.
14. McMillan related the influences of humanities computing and electronic text efforts in the humanities and pointed to work done at the University of Virginia's Electronic Text Center, the Text Encoding Initiative, and Annelies Hoogcarspel's *Guidelines for Cataloging Monographic Electronic Texts at the Center for Electronic Texts in the Humanities*, Technical Report No. 1, Center for Electronic Texts in the Humanities, Rutgers and Princeton Universities, 1994, as sources and influences. McMillan, "Electronic Theses and Dissertations," 108.
15. For a comprehensive overview of this literature, see Rebecca L. Lubas, "Defining Best Practices in Electronic Thesis and Dissertation Metadata," *Journal of Library Metadata* 9, no. 3–4 (2009): 252–63, http://dx.doi.org/10.1080/19386380903405165; "Chapter 6 Theses/Dissertations and ETD Cataloging: An Annotated Bibliography," *Technical Services Quarterly* (2008): 95–135, http://dx.doi.org/10.1080/07317130802127934. See also Sevim McCutcheon, "Basic, Fuller, Fullest: Treatment Options for Electronic Theses and Dissertations," *Library Collections, Acquisitions, & Technical Services* 35, no. 2–3 (2011): 64–68, http://dx.doi.org/10.1016/j.lcats.2011.03.019; Margaret Beecher Maurer, Sevim McCutcheon, and Theda Schwing, "Who's Doing What? Findability and Author-Supplied ETD Metadata in the Library Catalog," *Cataloging & Classification Quarterly* 49, no. 4 (2011): 277–310, http://dx.doi.org/10.1080/01639374.2011.573440; Michael Boock and Sue Kunda, "Electronic Thesis and Dissertation Metadata Workflow at Oregon State University Libraries," *Cataloging & Classification Quarterly* 47, no. 3–4 (2009): 297–308; McMillan, describing pioneering work done at Virginia Tech to develop bibliographic standards for ETDs, notes that the library "[pressed] the Graduate School to require authors of theses (in all formats) to provide keywords for use in the bibliographic record. Cataloging had been impressed for years with how labor-intensive was the task of assigning LC subject headings, so have the authors assign (even uncontrolled) subject headings was an appropriate alternative" (110).
16. Lubas, "Defining Best Practices in Electronic Thesis and Dissertation Metadata," 255.
17. As a press release explained, the NDLTD Global ETD Search, launched in July 2015, "allows researchers to find ETDs based on keyword, date, institution, language and subject." Networked Digital Library of Theses and Dissertations, "NDLTD Announces Global Electronic Thesis and Dissertation Search," July 6, 2015, www.ndltd.org/events/news/ndltdannouncesglobalelectronicthesisanddissertationsearch.
18. Joan M. Repp and Cliff Glaviano, "Dissertations: A Study of the Scholar's Approach," *College & Research Libraries* 48, no. 2 (March 1987): 148–59.
19. McMillan, "Electronic Theses and Dissertations," 110.
20. Repp and Glaviano, "Dissertations,"149. Indexes such as the *Dissertation Abstracts International* program, American Doctoral Dissertations, University Microfilm International played a significant role in attempting to collate and provide intellectual access to dissertations for these "extramural" users.
21. Jane Greenberg, "Metadata and the World Wide Web," in *Encyclopedia of Library and Information Science*, 2nd ed., edited by Miriam A. Drake. See also Jane Greenberg, "Theoretical Considerations of Lifecycle Modeling: An Analysis of the Dryad Repository Demonstrating Automatic Metadata Propagation, Inheritance, and Value System Adoption," *Cataloging & Classification Quarterly* 47, no. 3–4 (2009): 380–402.
22. McMillan, "Electronic Theses and Dissertations," 110.

23. Elizabeth Shepherd and Geoffrey Yeo, *Managing Records: A Handbook of Principles and Practice* (London: Facet, 2002), 5.
24. Greenberg, "Theoretical Considerations of Lifecycle Modeling," 385.
25. Ibid., 398.
26. While the Digital Curation Center's Curation Lifecycle Model is one of the more popular frameworks referenced in the professional literature, it is not the only data lifecycle model available to researchers. For more information on other types and examples of data lifecycle models, see: CEOS Working Group on Information Systems and Services, Data Stewardship Interest Group, "Data Life Cycle Models and Conceps, Version 1.0," 2011, accessed August 7, 2015, http://wgiss.ceos.org/dsig/whitepapers/Data%20Lifecycle%20Models%20and%20Concepts%20v8.docx; Bill LeFurgy, "Life Cycle Models for Digital Stewardship," *The Signal: Digital Preservation* (Washington, DC: Library of Congress, 2012), accessed August 7, 2015, http://blogs.loc.gov/digitalpreservation/2012/02/life-cycle-models-for-digital-stewardship/. Daniel Alemneh et al., *Guidance Documents for Lifecycle Management of ETDs*, edited by Matt Schultz, Nick Krabbenhoeft, and Katherine Skinner (Atlanta: Educopia Institute, 2014), vii. Completed in 2014, the *Guidance Documents* was an IMLS-funded project to frame the roles that various stakeholders play, including students, faculty, administrators, technologists, commercial vendors, and librarians, in confronting the "administrative, legal, and technical challenges presented by ETDs–from submission to long-term preservation" (i). As such, the document extends beyond metadata standards.
27. Alemneh et. al., *Guidance Documents*, viii.
28. Hussein Suleman et al., "Networked Digital Library of Theses and Dissertations," in *Electronic Theses and Dissertations: A Sourcebook for Educators, Students, and Librarians*, edited by Edward A. Fox et al. (New York: Marcel Dekker, 2004), 59.
29. Networked Digital Library of Theses and Dissertations, "Metadata, ETD-MS v1.1: an Interoperability Metadata Standard for Electronic Theses and Dissertations," edited by Thom Hickey, Ana Pavani, and Hussein Suleman, accessed August 7, 2015, www.ndltd.org/standards/metadata.
30. British Library, "The EThOS UKETD_DC application profile," accessed August 5, 2015, http://ethostoolkit.cranfield.ac.uk/tiki-index.php?page=The%20EThOS%20UKETD_DC%20application%20profile. EThOS documentation acknowledges the complexities associated with capturing dates: "Now your repositories might record a submission date, an award date, a digitisation date and/or a publication date and everyone may use different data and fields to record the information. EThOS currently records just one date: the date the thesis was awarded. In future there may be a case for introducing a second date field, to distinguish award date from publication or submission date for example" ("EThOS").
31. Alemneh et al., *Guidance Documents*, 122–23, 130, 133.
32. Database Management and Standards Committee, OhioLINK, "Standards for Cataloging Electronic Theses and Dissertations—Remote Electronic Version (non-Reproduction)," 2014, accessed August 7, 2015, https://platinum.ohiolink.edu/dms/catstandards/ETD-RDA.pdf.
33. Texas Digital Library, "Report for Texas Digital Library Description Metadata for Electronic Theses and Dissertations," v. 2, in *Texas Digital Library Descriptive Metadata Guidelines for Electronic Theses and Dissertations*, version 2.0 (September 2015), http://hdl.handle.net/2249.1/68437.
34. Thèses Électroniques Françaises, "Les métadonnées des thèses électroniques françaises," second ed., March 2006, www.abes.fr/abes/documents/tef/recommandation/index.html. The authors used Google Translate to translate TEF documentation into English.
35. The literature on metadata quality intersects with a diverse array of topics, including quality control and assessment (see Daniel Gelaw Alemneh, "Metadata Quality Assessment: A Phased Approach to Ensuring Long-term Access to Digital Resources," *Proceedings of the American Society for Information Science and Technology* 46, no. 1 (2009): 1–8, http://dx.doi.org/10.1002/meet.2009.1450460380; David Bade, "The Perfect Bibliographic Record: Platonic Ideal, Rhetorical Strategy or Nonsense?," *Cataloging & Classification Quarterly* 46, no. 1 (2008): 109–33; Yen Bui and Jung-ran Park, "An Assessment of Metadata Quality: A Case Study of the National Science Digital Library Metadata Repository," iDEA: Drexel Libraries E-Repository and Archives, 2006, https://idea.library.drexel.edu/islandora/object/idea%3A1600; Diane Hillmann, "Metadata Quality: From Evaluation to Augmentation," *Cataloging & Classification Quarterly* 46, no. 1 (2008): 65–80), continuing education (see Jung-ran Park, Yuji Tosaka, Susan Maszaros, and Caimei Lu, "From Metadata Creation to Metadata Quality Control: Continuing Education Needs Among Cataloging and Metadata Professionals," *Journal of Education for Library and Information Science* 51, no. 3 (2010): 158–76), revising legacy metadata (see R. Niccole Westbrook et al., "Metadata Clean Sweep: A Digital Library Audit Project," *D-Lib Magazine* 18, no. 5–6, http://dx.doi.org/10.1045/may2012-westbrook; Santi Thompson and Annie Wu, "Metadata Overhaul: Upgrading Metadata in the University of Houston Digital Library," *Journal of Digital Media Management* 2, no. 2 (2013): 137–47), and using automation to evaluate metadata quality (see Dongwon Lee, "Practical Maintenance of Evolving Metadata for Digital Preservation: Algorithmic Solution and System Support," *International Journal on Digital Libraries* 6, no. 4 (2007): 313–26, http://dx.doi.org/10.1007/s00799-007-0014-9; Mark Phillips, "Metadata Analysis at the Command-Line," *code4lib Journal* 19 (2013), http://journal.code4lib.org/articles/7818).
36. Jung-Ran Park, "Metadata Quality in Digital Repositories: A Survey of the Current State of the Art," *Cataloging &*

Classification Quarterly 47 (2009): 213-228, http://dx.doi.org/10.1080/01639370902737240.
37. Ibid., 220.
38. Ibid., 219. Park notes that several factors impact an object's completeness. For example, local metadata rules and requirements that dictate whether elements are required or optional frequently determine the extent to which record creators utilize certain fields to describe objects.
39. Ibid., 220.
40. Ibid., 221.
41. Ibid.
42. DuraSpace, "Metadata Recommendations," *DSpace 4.X Documentation*, 2014, accessed August 7, 2015, https://wiki.duraspace.org/display/DSDOC4x/Metadata+Recommendations.
43. Timothy W. Cole and Myung-Ja K. Han, *XML for Catalogers and Metadata Librarians* (Santa Barbara, CA: ABC-CLIO, 2013).
44. Kyle Rimkus and Thomas Having, "Medusa at the University of Illinois at Urbana-Champaign: A Digital Preservation Service Based on PREMIS," (paper, Joint Conference on Digital Libraries, July 22–26, 2013, Indianapolis, Indiana), www.ideals.illinois.edu/bitstream/handle/2142/45232/p49.pdf?sequence=2.
45. DSpace is not the only system for ETD management. Other popular alternatives include ETD-MS, EPrints, Fedora, and homegrown platforms.
46. Martin Dillon, "Prefatory Commentary from the Editor," in Rebecca Guenther, "MODS: The Metadata Object Description Schema," *portal: Libraries and the Academy* 3, no. 1 (2003): 137–50.
47. Lubas, "Defining Best Practices in Electronic Thesis and Dissertation Metadata," 255–57.
48. Ibid., 257.
49. DSpace JIRA card, "'dc.date.issued' is often incorrectly set (reported from Google)," created February 8, 2013, last modified May 8, 2015, DSpace website, https://jira.duraspace.org/browse/DS-1481.
50. Richard Rodgers, comment on DSpace JIRA card, "'dc.date.issued' is often incorrectly set (reported from Google)," February 8, 2013, accessed August 7, 2015, https://jira.duraspace.org/browse/DS-1481.
51. DuraSpace, "Metadata Recommendations."
52. "DS-1822: Find a way to report on existing, possibly inaccurate 'dc.date.issued' values," Duraspace website, accessed August 16, 2015, https://jira.duraspace.org/browse/DS-1822.
53. While this discussion focuses on DSpace, Vireo offered a compounding factor in the confusion over the date field. Vireo had been altered to include the "date of approval" as dc.date.issued, shifting the meaning of the values in the element without clear explanation.
54. The 2008 TDL guidelines specify that mods:dateIssued / dc.date.issued should be filled with the publication date, defined as "the date the ETD is released to the public." The guidelines note: "This date is automatically generated by DSpace upon ingest and does not need to be encoded prior to ingest." Texas Digital Library, "Descriptive Metadata Guidelines for Electronic Theses and Dissertations."
55. Brian E. Surratt, "MODS Meets Manakin: Innovations in the Texas Digital Library's Thesis and Dissertation Collection," (paper, 9th International Symposium on Electronic Theses and Dissertations, Quebec City, Canada, June 7–10, 2006), accessed April 22, 2015, http://docs.ndltd.org/dspace/bitstream/2340/668/1/SP6_Brian_SURRATT.pdf.
56. Texas Digital Library Metadata Working Group, "MODS Application Profile for Electronic Theses and Dissertations," Version 1 (December 2005), accessed November 4, 2015, www.tdl.org/wp-content/uploads/2009/04/etd_mods_profile.pdf; Alisha Little (University of Texas at Austin), Anne Mitchell (University of Houston), and Jason Thomale (Texas Tech University) were members of the 2005 working group. Brian E. Surratt (Texas A&M University) chaired the working group.
57. Surratt, "MODS Meets Manakin," 1–3.
58. Adam Mikeal et al., "Developing a Common Submission System for ETDs in the Texas Digital Library," July 2007, http://hdl.handle.net/1969.1/5679.
59. Texas Digital Library, "Descriptive Metadata Guidelines for Electronic Theses and Dissertations," Version 1.0, June 2008, www.tdl.org/wp-content/uploads/2009/04/tdl-descriptive-metadata-guidelines-for-etd-v1.pdf; Jay Koenig (Texas A&M University), Anne Mitchell (University of Houston), William Moen (University of North Texas), Tim Strawn (University of Texas at Austin), and Jason Thomale (Texas Tech University) were members of the 2008 working group. Amy Rushing (University of Texas at Austin) chaired the working group.
60. Ibid.
61. Included in the 2005 and 2008 guidelines were dates associated with the creation of the document, derived from both the "date of creation" and the "date of publication," and encoded in MODS as origin information. The 2008 standard mapped both of these dates to ETD-MS as dc.date. Additional date fields included birth dates of authors and thesis committee members. The guidelines accounted for ETD lifecycle events by capturing date information associated with the creation and revision of ETD bibliographic records. "MODS Application Profile for Electronic Theses and Dissertations"; "Descriptive Metadata Guidelines for Electronic Theses and Dissertations."
62. Texas Digital Library, "Dictionary of Texas Digital Library Description Metadata for Electronic Theses and Dissertations," v. 2, in *Texas Digital Library Descriptive Metadata Guidelines for Electronic Theses and Dissertations*, Version 2.0, September 2015, http://hdl.handle.net/2249.1/68437.
63. Texas Digital Library, *Texas Digital Library Descriptive Metadata Guidelines for Electronic Theses and Dissertations*, version 2.0, September 2015, http://hdl.handle

.net/2249.1/68437; Santi Thompson (University of Houston), Monica Rivero (Rice University), Kara Long (Baylor University), Colleen Lyon (University of Texas), and Kristi Park (Texas Digital Library) were members of the 2014–15 working group. Sarah Potvin (Texas A&M University) chaired the working group.

64. The *Guidelines* noted, "The authors of these recommendations . . . worked towards an ideal of repository-neutral guidelines. But . . . the constraints of DSpace, and its dominance in the TDL and Vireo User communities, provided an argument for tailoring some recommendations to the known constraints and behavior of DSpace repositories. As Vireo and TDL diversify to incorporate Fedora repositories, greater awareness should be paid to the aspects of the guidelines that are not repository-neutral, and to considering the need to tailor recommendations to Fedora and other repository systems." Texas Digital Library, "Report for Texas Digital Library Description Metadata for Electronic Theses and Dissertations," 18.
65. See "A Note on MODS, "Dictionary of Texas Digital Library Description Metadata for Electronic Theses and Dissertations," 4–5.
66. Suleman et al., "Networked Digital Library of Theses and Dissertations," 59.

Notes on Operations

Using Perceptions and Preferences from Public Services Staff to Improve Error Reporting and Workflows

Dawn McKinnon

The ongoing transition from purchasing mostly print materials to electronic resources (e-resources) continues to pose workload challenges in libraries. In response, many libraries have focused on improving workflows to increase efficiency, which provides better service. This paper discusses a project undertaken to tackle one aspect of these challenges, in which data was gathered on how front-line library staff report errors found in the library catalog and discovery layer, and their preferences and perceptions for reporting errors to Collection Services staff. It also identifies improvements that can be made to error reporting, workflows and communication between Collection Services and front-line staff, to create a more service-oriented and efficient working environment in the library.

Over the past decade most academic libraries have transitioned from purchasing mostly print materials to electronic resources (e-resources). This transition has been well documented in library literature, particularly the workload struggle that libraries face during this transition and the challenges because of the ad hoc fashion in which e-resources are often managed.[1] As libraries attempt to work more efficiently to improve service, a few studies have analyzed staff reporting of access issues and catalog errors but these tend to focus more on improving workflows from a technical services perspective or factors that lead staff to report problems.[2] McGill University Library is a large research library in North America. Within its Collection Services department, the e-resources division handles cataloging, access and troubleshooting related to e-resources such as electronic journals (e-journals) and databases. Like many libraries, this division has undergone considerable change during the transition from print to electronic. To address some of these challenges and to help fill a gap in the literature, a research project was undertaken to gather data on how the library's front-line staff reports errors they find in the discovery layer and catalog, their preferences for reporting, and perceptions of the response times and quality of the responses provided by the Collection Services staff. As many libraries face friction between front-line staff and those who work "behind the scenes," the author aims to share lessons learned from this project and continue the discussion on the need for best practices in this area. This paper discusses the project, analyzes the results, and identifies where improvements can be made to error reporting, workflows and communication between Collection Services staff and front-line staff to create a more service-oriented and efficient working environment. It should be noted that the word "errors" in this paper refers to questions asked and errors reported to Collection Services, including, but not limited to, e-resource access problems, questions about subscriptions and renewals, and cataloging errors. "Front-line staff" refers to librarians and nonlibrarian staff who work with patrons in public

Dawn McKinnon (dawn.mckinnon@mcgill.ca) is Electronic Resources & Serials Librarian at McGill University in Montreal.

Submitted July 7, 2015; returned to author September 21, 2015 for revisions; revised manuscript submitted November 17, 2015; returned to author for minor revisions December 10, 2015; accepted for publication January 24, 2016.

services, including subject librarians, library assistants and supervisors who work at the service desks, and the interlibrary loans staff.

Literature Review

A recurring theme in the literature pertaining to e-resource management is the fluctuating roles and responsibilities of e-resources librarians.[3] As roles remain in flux, frustrations with workflow inefficiencies are often highlighted. Waterhouse discusses the challenges of having "many systems involved in managing and delivering e-resources" at the University of Illinois Springfield (UIS), including SFX, Serials Solutions 360 and WorldCat Local.[4] In addition to many systems, inefficiencies also occurred because the "acquisitions, processing, and cataloging workflows were quite separate from those of e-resource management" and the staff supporting e-resources was unfamiliar with each workflow.[5] Mackinder refers to workflows as the "seemingly endless challenge" because "the staff time and effort involved in crafting, implementing, and revising process documentation can be overwhelming" because the workflows are not linear.[6] Four years after creating her "ER lifecycle," she is still "workflow brainstorming" because "change is the status quo" in this field.[7]

As librarians describe their unique challenges with workflows, software is often evaluated in the literature as a possible solution. Duke University Libraries turned to IBM's BlueWorks Live and Business Process Manager to improve e-resources workflows following what they called a "fallout" from cumulative errors made over several years.[8] At Ohio State University Libraries, Feather examined tools to improve e-resources communication workflows, productivity and efficiency. This makes sense, as her study and others found email to be a main tool used for reporting access issues and troubleshooting e-resources.[9] Electronic Resource Management (ERM) systems are one of the latest tools discussed to help libraries with their workflows. Although ERMs are good at "issuing renewal reminders . . . they are less successful with more complex workflow issues."[10] At UIS, the ERM is but one piece of the workflow and they also rely on the library's intranet and face-to-face meetings.[11] In 2008, Emery reported that "in theory, ERMs are a winner . . . yet, in practice, we have discovered that ERMs do not immediately solve all the problems as we expected." Grogg examined ERMs in 2008 and again 2011 with Collins, but still found unfavorable reviews where workflow was concerned, calling it "one of the biggest deficiencies (and disappointments) of ERMS functionality."[12] ERMs are continually improving, but their pros and cons are still being discussed, as evidenced by a 2014 ALA Midwinter Meeting panel on this topic.[13] Nearly all panelists expressed how ERMs helped overcome some workflow problems but they are only one tool for e-resource management.

Moving beyond software, the "core competencies" for e-resources librarians is another approach found in the literature.[14] Proponents cite that "cross functional, cross-trained" teams skilled in communication, problem-solving, and licensing models, who are flexible, persistent, and understand the organizational structure, will have a "high rate of problem resolution and user satisfaction."[15] In addition, the phrase "best practices" is often used but has yet to be fully fleshed out. For example, Samples and Healy describe a "need for libraries to develop best practices for troubleshooting electronic resources."[16] Pomerantz surveyed more than two hundred librarians and concluded that there is "a great deal of variation in practices and inconsistency in training experiences" and that a "set of best practices" is needed.[17] Although Sample and Healy were referring mostly to proactive troubleshooting and Pomerantz was referring to the role acquisitions librarians play in e-resource management, the sentiment applies to the broader picture, as shown by the development of Techniques for Electronic Resource Management (TERMS).[18] TERMS began in 2008 following a discussion about "what was lacking both in current practice and with the systems available" and has grown to be a reference point for managing e-resources.[19] The academic literature is sparse on systematic implementation of TERMS, and TERMS workshops have started occurring at conferences such as Electronic Resources and Libraries (ER&L).[20] TERMS provides "feedback from those in the field who are actively managing electronic resources" with what Mackinder calls invaluable "real-world data" that creates a "shared understanding" that can help with e-resources management.[21]

All these efforts are necessary because e-resources teams "must be responsive to the high expectations of users and other library staff."[22] Samples and Healy identified that "initiating a troubleshooting workflow can come from two main avenues—library staff and patrons."[23] Library literature includes an abundance of papers on patron perceptions and opinions, particularly as LibQUAL assessment moves into its second decade, yet little has been written about front-line staff expectations and preferences regarding e-resource error reporting.[24] Foster and Williams' 2010 article is one of the recent few that includes library staff in their study, which focuses on factors that lead staff to report errors and how to encourage more reporting. They refer to front-line staff as a "vital group in identifying problems" as they are "best positioned to discover problems with resources" that may not be revealed through other work done by e-resources staff.[25]

Several papers discuss using error reporting to improve e-resources workflows, but many of the data sets are now nearly a decade old.[26] Samples and Healy's 2013 survey polled libraries about error reporting forms and showed that just over half of respondents (57 percent) had a single form designed for both staff and patrons, and they echoed Dowdy and Raeford's sentiment that "effective communication across units is hampered by inefficient and largely non-automated techniques."[27]

Given that front-line staff are well situated to discover problems and that e-resources workflows constantly need improvement, the project discussed in this paper focuses on one gap in the literature: front-line staff's preferences and perceptions around reporting errors found in the library's discovery layer and catalog.

Background

McGill University is a research university with approximately 22,000 undergraduate students and 10,000 students in masters, doctoral, and postdoctoral programs. McGill Library is an Association of Research Libraries (ARL) member, with 174 employees—63 librarians and 111 full-time library staff, located in ten urban branches and one suburban branch. Its Collection Services department manages tasks related to cataloging, metadata, acquisitions, processing and most aspects of maintaining the discovery layer. In 2012–13, through attrition and austerity measures, the number of Collection Services staff decreased from 55 to 36 and the department was restructured to rebalance workloads. The ten-person "Serials, E-resources and Acquisitions" division became the "E-resources and Serials" division with two librarians and four staff, managing cataloging and access related to print and e-journals, databases, and streaming media, with primary responsibility for maintaining the discovery layer. This division also triages questions from patrons and staff sent to the Collection Services email account, as the majority of the questions are related to e-resources. Other types of questions, such as print cataloging or processing questions, are forwarded to the appropriate division.

During this period, the library administration moved to an "e-preferred" collection policy because of a concern about the lack of shelf space for print material and the ability to have more purchasing power for buying e-books in bulk packages. The e-book collection continued to grow as faculty and students provided positive feedback about e-books from publishers with unlimited simultaneous users. As the number of e-book acquisitions grew faster than they could be cataloged, the backlog swelled to over one million e-books. A new "E-books Cataloging" division was created, and the e-resources staff member who worked on e-books moved to this new division along with two others who had previously worked with print material. Formerly, at least one staff member in each division was dedicated to acquisitions. For example, one staff member in the e-resources division worked primarily on acquisitions tasks for e-resources; several people in (print) Cataloging completed acquisitions tasks for print material. These disparate acquisitions staff were merged into a new "Collection Development" division. Through attrition, Collection Development decreased from eight people to five during this period, and staff were not replaced because of financial constraints. As staffing numbers were reduced through attrition, the library decided to outsource most cataloging of current physical material (i.e., shelf ready monographs). The remaining "Processing and Cataloging" division handles rush monograph cataloguing and related end processing. A cataloging backlog of rare material became a priority for the library administration, who wanted to highlight unique items in the collection. Several people who had been doing a variety of cataloging and processing tasks were moved into the Rare and Special Collections Cataloging division to address this priority, bringing the number up to seven.

During these organizational changes, in 2012–13 the e-resources team completed a soft implementation of the discovery layer while maintaining the traditional catalog. The number of questions directed to Collection Services increased during and after the discovery layer implementation for three broad reasons: public services staff did not know who to contact for help, the discovery layer came with a learning curve, and it exposed more e-resources and access issues than the library's traditional catalog.

A year after the restructuring, in November 2013, the library prioritized the need to "improve mechanisms for reporting and responding to problems" with the discovery systems and the catalog during a strategic planning session.[28] Before the planning session, the process for reporting problems to Collection Services consisted of a mix of phone calls, email, web forms, and in-person visits. Front-line staff had difficulty remembering which form or email address to use to report problems. Foster and Williams reported a similar issue at Milner Library, where the "reason most often given for why someone was not likely to report a problem was being unsure of how or to whom to report" it.[29] The planning session also revealed that front-line staff felt Collection Services responses were often delayed or nonexistent. Many e-resources staff were frustrated and overwhelmed by the organizational and workload changes, and the lack of clear workflows.

To address some of these issues, email service accounts were created to relieve front-line staff from remembering who performed each task, and it allowed for workload sharing. However, so many service accounts were created that front-line staff then had difficulty remembering which account to use for each problem. In 2014, a single Collection Services email account was created and staff were encouraged to use it for all questions, from purchasing to cataloging to access. This mailbox is triaged by the E-resources and Serials division.

In June 2014, the library officially launched its discovery layer and a new link resolver. Noticeable changes included a having the new discovery layer as the default search from the library's website, updates to the look and behavior of the link resolver, and the removal of all e-resources from the legacy catalog. Although it is not prominently displayed on the website, the legacy catalog is available and can be used to locate circulation information for nonelectronic materials such as print books and journals.

The staff restructuring in tandem with the migration to the new discovery layer and link resolver were the catalysts for this research project. The number of issues reported dramatically increased but there were fewer staff to respond, which emphasized inefficiencies and gaps in existing workflows. The library's strategic goal of improving mechanisms for reporting problems became paramount for the e-resources staff as they searched for new ways to manage it.

Method

Data was collected in three ways: statistics on errors reported were collected, an online survey was conducted, and personal interviews were conducted. Statistics were collected for errors reported to Collection Services during a one-month period. This provided a sample that could be analyzed and compared against data and comments collected through the online survey and interviews. Data were compiled and analyzed in Microsoft Excel spreadsheets. Comments were summarized to ensure anonymity, which was important for gaining trust from the staff and helped to provide a higher response rate.

Statistics on Reported Errors

During October 2014, errors reported to the Collection Services and e-resources mailboxes were monitored as were errors reported through the "Report a Problem" form from the legacy catalog typically used to correct print holdings or other errors. Errors reported directly to e-resources staff by phone, email or in person were also included. Although errors had never been systematically tracked, October was selected for the project because anecdotally it seemed to be the month in which the most errors were reported every year. As with many academic libraries, students seem to start using the library's resources more heavily in October because of mid-term exams and papers and the beginning of group project work that is due at the end of the term.

Errors are normally triaged by several Collection Services staff. As a pilot method for this project, the author triaged the majority of the errors. Responses were provided by the author and other Collection Services staff. To mimic normal working conditions outside the project, work was done only during regular business hours and the staff was not encouraged to work faster than normal. Even with these parameters, the pilot method of one person triaging the errors may have created artificial response times and is discussed later in this paper. To provide more conclusive results in this area beyond a pilot, the triage method would need to be assessed further and addressed.

Since Collection Services does not use an automatic mechanism for tracking errors, during the project the following was entered into an Excel spreadsheet for each error:

- date and time the issue was submitted (using the email timestamp or time the person phoned/visited)
- date and time the issue was first viewed/heard by Collection Services staff
- name and division of the person reporting the issue (e.g., front-line staff, ILL, etc.)
- method of how the issue was delivered (e.g., email, phone, in-person)
- division responsible for responding (e.g., e-resources, e-books, Collection Development)
- if/how and when an acknowledgement was provided to the sender (e.g., verbally or by email)
- description of the issue

Noting the time differences between when errors were submitted and when they were first viewed by Collection Services staff served two purposes:

1. to detect delays between when errors are submitted and when they are viewed by Collection Services staff;
2. to provide possible explanations for longer response times when errors are submitted after business hours.

The staff who triage errors sent by email use a schedule so that at a given time, typically only one person is managing the inbox. This prevents multiple people from accidentally working on the same problem at the same time. Anecdotally, it was common practice for staff to begin resolving issues immediately upon first viewing of the report, and thus "viewed by Collection Services" captured the start of the process to resolve the error.

Tracking how and when acknowledgements were provided was in response to concerns from front-line staff who felt that reported errors were never addressed. As common practice, the e-resources division sends email acknowledgements for errors that they expect will take longer than a day to be resolved and when errors are forwarded outside of the division. Acknowledgements are not sent automatically, and occasionally staff forget to send them as it is not an explicit policy. This practice was not altered during the project.

For resolved errors, the following was added to the Excel spreadsheet:

- resolution date and time
- "response time from issue sent": the time difference between when the issue was sent and when it was resolved
- "response time from issue viewed": the difference in time between when the issue was viewed by Collections Services staff and when it was resolved

Online Survey

The author searched for an existing survey tool to evaluate staff preferences and perceptions, particularly related to error reporting. Foster and Williams also designed an online survey tool to gather feedback from library employees, but as this survey is longer and more detailed than desired, it was not used.[13] Thus, the author drafted a survey and collaborated with colleagues to establish validity (see the appendix).

The survey was created using LimeSurvey, open-source survey software (https://www.limesurvey.org). One advantage of using this program is that the raw data can be received in a variety of formats and there are settings to ensure anonymity. The survey was designed to be completed in ten minutes or less to elicit a high response rate. Visually, it is a single online page of ten numbered questions; some questions have multiple, related parts, so participants are actually asked fifteen questions.

The final two questions are open-ended and the remaining questions are a mixture of multiple choice and forced choice. All questions are optional, and participants can exit at any time. Controls were not in place to prevent individuals from responding multiple times since this appeared to be a low risk. It was assumed that staff would take the survey seriously and want to improve service.

The questions were grouped into these categories:

- current behavior when reporting errors (questions 1–3)
- expectations and preferences for reporting errors (questions 4–7a)
- perceptions of response times and quality of responses (questions 7b–9)
- comments (question 10a and 10b)

In addition to general comments, respondents were asked to describe a time when they were not satisfied with Collection Services and to describe what could have been done differently for a more satisfactory result. The objective of this question was to gather commonalities between the historical "worst case scenarios."

The questions, methods for distribution and dissemination of results were approved through the university's Institutional Review Board (IRB). Because of the potentially sensitive results and the possibility for employees to provide unfavorable feedback on their colleagues' work or to be accidentally identified, raw results were only viewed by the author and were anonymized and summarized before they were shared.

The survey was distributed by email to all (174) full-time library employees. The email instructed employees who were not front-line staff not to respond. A second email was sent to each division that was not considered front-line, reminding them again not to respond. Collection Services staff were reminded verbally as well, as responding would mean that they would be reporting on their own work and would invalidate results. Although there is no guarantee that other (non-front-line) staff did not respond, this risk is assumed to be low, not only due to the number of strong reminders but also because staff were interested in the results and were on board with the survey and improving service. The survey email was sent from the library's communication officer as she does not supervise any employees, and this minimized potential pressure to respond or answer favorably, as stipulated by the IRB. This resulted in 103 front-line staff or potential respondents; 56 people responded, yielding a 54 percent response rate.

Personal Interviews

At the end of the survey, participants were invited to email the author if they were interested in completing an in-depth interview. Six weeks after the survey closed, an additional email request for volunteers from the front-line staff was sent to all full-time employees using the same email distribution method as was used for the survey. This resulted in eight volunteers who completed the interviews in January and February 2015. A separate consent form was used for this part of the project, as the data was confidential but no longer anonymous. As stipulated by the IRB, only questions from the online survey could be asked during the interview; however, they could be asked in a different order. To facilitate an easy flow of conversation, all interviews started with the final question from the survey, asking interviewees if they would like to provide general feedback. The author then asked the questions from the online survey in their original order, skipping questions if they had already been answered through the normal course of conversation.

Results

Methods for Reporting Errors

During the project, 296 errors were reported in a variety of ways as shown in table 1. As nearly three-quarters of the errors were sent through the Collection Services email account, it is clear that using this single email account was the preferred reporting method during the project period.

Survey questions 1–3 asked which methods were used the last time respondents reported different types of errors, including problems with missing information in the record, unable to find known items using the discovery layer, and subscription or access problems. Respondents could select multiple responses, but for all error types, there was a clear preference for emailing service accounts, as shown in table 2. One interviewee stated, "In the past it wasn't always clear who we were supposed to report to . . . it's much clearer now with one service account." Most interviewees echoed this sentiment, specifying that not having to find email addresses or names of people responsible for each division is faster and less frustrating when reporting errors. Responses in the survey's "Other" comment box and some interviewees cited time constraints as a reason why an issue might not be reported.

Some survey comments and three interviewees mentioned a newly created pilot web form. It was not included as an option in the survey, as it was still being tested and not yet available to all staff. Due to pressure to re-create the old "Report a problem" form that was available in the legacy catalog, the e-resources division designed a new web form that includes fields for the title, URL, format (e.g., e-book, database, etc.), type of problem (e.g., broken link, missing print holdings, etc.) and a comment box for additional information. Upon clicking "Submit," an email is sent to Collection Services. Many said completing the new pilot web form is faster than writing an email. One interviewee said, "Once the form was created, I stopped using the service account to report basic errors."

Types of Errors Reported

Nearly 78 percent of the errors were related to e-books (29.4 percent) and e-resources (48.3 percent), as shown in in table 3. As e-book errors are resolved by a separate division, for this project they were considered separately from errors related to e-resources (databases, e-journals, etc.), which are resolved by the e-resources and Serials division.

Resolution Rates and Response Quality

Of the 296 errors reported during the project, 59 percent (175) were resolved by e-resources staff (see table 4). Although 10 percent (29) were assigned but unresolved at the end of the data collection period, the majority of these were resolved in the immediate weeks after the project closed.

A quarter of the errors were coded as "forwarded internally" and no longer deemed e-resources' responsibility. These errors were passed to other Collection Services divisions, and were typically subscription problems sent to Collection Development (5 percent) and access errors sent to the e-books Cataloging division (18 percent). It should be noted that while eighty-seven errors were coded as "e-book" errors (table 3), only seventy-five

Table 1. Actual Methods for Reporting Errors

Method	% of Total Reported	No. Reported
E-mail to Collection Services mailbox	73.3	217
E-mail to e-resources staff member	13.5	40
E-mail to e-resources mailbox	8.1	24
Phone	3.0	9
Feedback form on the Library's website	1.0	3
"Report a problem" form within the catalog	0.7	2
In person	0.3	1
Total		296

Table 2. Online Survey Responses on Methods used for Reporting Errors

Responses	Q1. Unable to Find Item Record	Q2. Catalog Record Missing Information	Q3. Suspected Subscription Problem
This has never happened to me / I can't remember.	6	2	4
Verbally told my colleague who works in Collection Services.	5	1	1
Verbally told a colleague who does not work in Collection Services.	5	0	1
E-mailed one of the Collection Services general mailboxes.	26	30	29
E-mailed a Collection Services staff member directly.	7	10	10
E-mailed a colleague outside of Collection Services (e.g., another librarian or a supervisor).	2	1	2
Used the "Report a problem" form within the traditional catalog.	3	4	1
Used the "Chat with a librarian" to report it.	0	0	0
I did not report it.	4	2	1
Other	7	3	4

were forwarded to the e-books Cataloging division (table 4). The remaining twelve e-book errors were handled by the e-resources staff member triaging the mailbox because the questions were short, customer-service questions such as printing from an e-books platform, rather than errors that required cataloging expertise. Partially because of the outcome of this project, all e-books questions are now forwarded to the e-books Cataloging team. The remainder were unique, "one-off" questions related to database maintenance, interlibrary loans, and patron reporting. It was outside the scope of the project to track response times for other divisions, although anecdotally, all e-book errors were resolved within two days. Errors forwarded to Collection Development are discussed in more detail later.

Five percent of the errors were sent to OCLC, the vendor responsible for the library's discovery layer and copy cataloging records. None of these errors had been resolved by January 31, 2015, three months after the study had closed, but e-resources staff continue to track and follow-up on these errors. OCLC has various responses, including that some features are not yet available, and they were working on system updates that would include resolutions.

Of the survey respondents who answered questions about quality (question 7b), about two-thirds were "satisfied" with the response they received the last time they reported an issue to Collection Services. Question 9 asked respondents to select the statement that best represents them regarding reporting errors; half of those who answered the question indicated that they felt their errors were answered to the best of the staff's abilities, as shown in table 5. All of the interviewees indicated this sentiment as a typical experience, excepting subscription problems and those forwarded to OCLC.

Question 10 asked respondents to describe a time when they were not satisfied and to comment on what Collection Services could have done differently. Fourteen people provided examples and four suggested improvements without specific examples. These comments can be grouped into the following themes:

- frustration with errors that cannot be resolved by Collection Services, in particular errors forwarded to OCLC
- frustration with little or no follow-up communication on outstanding errors
- poor treatment by Collection Services staff
- too much reliance on front-line staff to report errors

The remaining participants did not respond to this question or wrote that they had never had a bad experience.

Table 3. Types of Errors Reported to Collection Services

Type of Error	% of Total Reported	No. Reported
e-resources (databases, e-journals, etc.)	48.3	143
e-books	29.4	87
Print books	6.4	19
Print serials	6.4	19
Acquisitions and subscriptions	5.1	15
Database maintenance	1.7	5
Films	0.7	2
VPN	0.7	2
Using the library's website	0.7	2
Inter-library loans	0.3	1
Reports from the library's ILS	0.3	1
Total		296

Table 4. Errors Resolved, Unresolved and Forwarded

	% of Total Reported	No. Reported
Resolved by e-resources in October	59.1	175
Assigned to e-resources but unresolved as of October 31	9.8	29
Forwarded internally	25.3	75
Forwarded externally to OCLC (unresolved)	5.7	17
Forwarded externally to another vendor	0.3	1
Total		296

All interviewees said that *most* of the time they are generally happy with response quality. Three interviewees said they had never had a negative experience.

Response Times

Over half of the online survey respondents indicated that they preferred resolutions within the same day or next day, as shown in table 6.

When respondents were asked to indicate the response time for the last error reported to Collection Services (question 4), nearly the same number of respondents indicated that it had occurred within the same day or by the next day.

This is consistent with the response times collected during the project, where 156 errors were resolved within 24 hours of being submitted, representing 53 percent of all errors reported, or 89 percent of errors resolved by the e-resources division. Of those resolved within the

Table 5. Online Survey Responses on Perception of Response Quality

"Select the statement that best describes you. Typically, when I report errors regarding items in the Classic Catalog and/or WorldCat Local to Collection Services, I:"

Statement	No. Who Selected This Response	% Who Answered This Question
Feel like the problems are resolved to the best of the staff's abilities.	22	50
Feel like Collection Services is aware of the problem but they do not or cannot resolve it.	2	5
Feel like my particular case has been noted but it is part of a larger problem that has not yet been resolved.	5	11
Feel like my problems are eventually resolved but they are not a priority.	0	0
Depends—sometimes a, b, c, or d.	15	34
Total	44	

Table 6. Preferences for Response Times

"When you report an error found in the Classic Catalog or WorldCat Local, what is the preferable time frame for a response to be communicated?"

Response Time	No. Who Selected This Response	% Who Answered This Question
A few minutes	1	2
Within the same day	12	27
By the next working day	19	42
Within a week	5	11
Within a month	0	0
Depends on the problem	4	9
I don't have expectations for response times	1	2
Unsure	0	0
Other	3	7
Total	45	

twenty-four-hour period, more than half (84) were resolved within sixty minutes of submission (see table 8).

There were differences in response times when calculating from the time the issue was *viewed* by Collection Services staff, rather than when it was sent, particularly within the twenty-four-hour window. For example, the number of errors resolved within ten minutes increased from twenty-six (counting from time sent) to eighty-seven when counting from time viewed. However, at the twenty-four-hour turnaround time, both resolution times are equal (156/296 resolutions, from time sent and from time viewed). Of the errors resolved by the e-resources staff, all but one were resolved within five days of viewing; the outlier was resolved in fourteen days because it required the help of two different vendors.

Every interviewee indicated that most of the time they felt that errors were resolved in a timely manner. Four people indicated that "occasionally" their reported errors were never answered but that this happened less frequently now than in years past. Some survey respondents and interviewees perceived problems with response times for errors related to subscriptions and renewals. Several people stated that even though many of the errors forwarded internally were resolved quickly, the poor response time and lack of follow-up on the few outstanding subscription errors was so significant that it overshadowed the positive resolutions. As with all data collected, these comments were summarized for anonymity and then shared first with Collection Services division coordinators with recommendations for moving forward. The proposed solutions to these challenges are discussed in the Recommendations section.

Communication

During the project, email was the most commonly used method of communication, with 94.9 percent of errors reported via one of three email options (sum of 73.3 percent to Collection Services, 13.5 percent to individuals and 8.1 percent to the e-resources service account). Survey results indicated that email acknowledgements are preferred, with thirty-two participants preferring them and only seven indicating that they did not (in response to question 5). However, many respondents commented that acknowledgements are preferred only when resolutions cannot be provided within the same day. During the project, acknowledgements were delivered 65 percent of the time. They were not sent for errors that were expected to be resolved within a few hours, as an email confirming resolution was sent instead; this is common practice for the e-resources division. The interviewees were split evenly on the usefulness of acknowledgements: two said it helped them track outstanding errors; two people only wanted resolutions, not acknowledgements; the remainder had neutral opinions on this topic.

Most survey respondents and interviewees indicated neutral or positive encounters regarding communication with Collection Services staff, but as indicated earlier, several respondents mentioned that they felt that Collection Services staff was sometimes unresponsive or rude, although they all said that this feeling was happening less often than in previous years. They wished updates were sent proactively by Collection Services staff, particularly when there are delays. Several others mentioned feeling "lost" or not knowing the procedures to follow up on their errors themselves.

Discussion

It is indicative of the times that the majority of the errors reported during the project were for e-resources rather than print material. This notion has been documented by many in recent years, including Henderson and Bosch who predicted in 2010 that a "shift from print to digital is likely to accelerate greatly."[30] This is certainly true for McGill Library, further adding to the evidence found in the literature for a need to focus on e-resources workflows and management. Like Dowdy and Raeford at Duke University, taking stock of the existing environment and workflow helped staff at McGill University to determine a course of action for improvement.[31] In part, this project aimed to establish commonalities in the types of errors reported to better understand the situation. The following themes emerged from the interviews and the online survey comments:

- frustration with data errors that cannot be fixed in-house and must be forwarded to OCLC
- cases where front-line staff felt the communication from Collection Services was unpleasant
- difficulty receiving answers for subscription questions
- front-line staff feel they are relied upon too heavily to report errors found in the discovery layer, that this is beyond their responsibilities, and that Collection Services should be doing more to proactively fix these errors

Examples of errors that are forwarded to OCLC typically involve incorrect metadata. Some metadata can be corrected locally while others can only be fixed by OCLC. For example, when the discovery layer provides a link to an e-book or e-journal that has the same title as the item in the record but is actually a *different* item (with a different author, or ISSN, etc.), the correction can only be done by OCLC. It is often several months before these errors are resolved.

Difficulty with subscription issues is partly because of silos of information, where updates are not shared across Collection Services departments, which can cause delays when changes are made to subscriptions. McGill Library is not alone in this struggle; Samples and Healy also found silos to be one of the main points of workflow failure reported by the ARL e-resources librarians.[32]

The final theme revealed through survey comments was that front-line staff feel indicated that they feel they are relied upon too heavily to report errors found in the discovery layer, that this was beyond their responsibilities, and that Collection Services should be doing more to proactively fix these errors. This fits with Samples and Healy's research on the need for proactive troubleshooting best practices and Dowdy and Raeford's recommendation for proactive quality control in e-resources. For example, using a wiki or other mechanism for informing public services staff so that they would know about planned database

Table 7. Perceptions of Response Times

"Thinking of the last time you reported an issue, how long did it take for you to receive an answer or resolution regarding the reported problem?"

Response Time	No. Who Selected This Response	% Who Answered This Question
A few minutes	3	6.7
Within the same day	21	46.7
The next working day	7	15.6
Within a week	7	15.6
Within a month	1	2.2
I never heard back about the problem	0	0.0
Unsure	2	4.4
Other	4	8.9
Total	45	

Table 8. Response Times from when Errors were sent to Collection Services

Response Time	No. of Errors (accumulative count)	% of All Errors Reported	% of Resolved Errors
10 min or less	26	9	15
30 min or less	51	17	29
60 min or less	84	28	48
Within a half day (4h)	133	45	76
Within 24h	156	53	89
Within 5 days	173	58	99

down times, and doing subscription inventories to ensure that the databases, e-journals and e-book collections are all activated properly.[33] Ensuring that current subscriptions have the correct links, and that obsolete subscriptions are removed takes the onus of reporting access issues away from the patrons and staff.

One area that appears to have improved since the strategic planning session in 2013 is the confusion regarding who in Collection Services to contact for help. During the project, 74 percent of errors were reported through the Collection Services email account, and the majority of survey respondents indicated that they used this method the last time they reported errors. Other libraries also found email to be a one of the most popular ways report access errors, and also web forms.[34] As previously mentioned, the e-resources division had created a new web form for reporting errors to automate a step in the workflow to improve efficiency. During the research project, it was testing by a several staff members who provided positive feedback. Several interviewees mentioned that the form was a "huge improvement" and one interviewee said, "I'm reporting more lately because I love the form."

This feedback suggests that streamlining the reporting process has made it easier for front-line staff to report errors. Triaging through the one service account is also easier for Collection Services staff as several people can monitor the account, staff can share the workload, and scheduling is less of a concern. As Feather notes, there is a danger of using personal accounts as "if one person is absent and receives a message, no one else will be able to respond to it in a timely manner."[35]

Even though many front-line staff had previously said they were unsure of who to contact, it is interesting that emailing individuals in Collection Services is the second-highest survey response. This may be due to habit or because of a friendship between a front-line staff member and a Collection Services employee. It could also suggest that some people find they receive better service by emailing an individual that they know.

Not surprisingly, no one selected using the "chat with a librarian" service (QuestionPoint) to report errors. It was added to the survey to see if anyone preferred using this method. This service is only occasionally staffed by Collection Services librarians, and is not currently used to communicate between staff at McGill Library, so the result was expected.

Resolution Rates

As it was outside the scope of the project to analyze errors forwarded to other Collection Services divisions, resolution rates were only included for errors answered by the e-resources division. At the end of the project, the division staff was surprised at the high number of resolutions: 83 percent of those assigned to e-resources (or 59 percent of all errors) were resolved within the project timeframe, and the remaining were resolved in the weeks after the project ended. It speaks to human nature that staff remember the errors that they were unable to resolve or that took longer than expected.

In contrast, all errors reported to OCLC remained unresolved during the project timeframe and for many months afterward. Some were never resolved and some were marked as "features" for the future. Although these represented only 5 percent of reported errors during the project, the volume of comments and level of frustration from survey respondents and interviewees far outweighed what the statistics demonstrate. If OCLC errors continue to be unresolved for long periods of time, front-line staff may stop reporting them, as shown at Milner library, where staff do not always report problems that they "figure can't be fixed."[36]

Response Times

Many survey respondents and interviewees noted that most of the time, responses arrived within the preferred timeframe of the same day or the next day, and this timeframe was consistent with data collected during the project. However, unlike the other data collected in the research project, the pilot method of one person triaging the errors is not true to the working environment and may have affected response times. Anecdotally, response times during the project appeared to resemble response times external to the project, but cannot be stated with certainty. It is notable that during the project, the differences in the response times counting from when the issue was *sent* versus when it was *viewed* by Collection Services occurred only within the twenty-four-hour timeframe. After that point, both resolution times are equal (156/296 resolutions, from time sent and from time viewed). Thus, for the duration of the project, about half of all reported errors are resolved within 24 hours regardless of when they were sent or viewed.

Response times for errors relating to subscriptions were highlighted in the survey and interviews as an area that required particular improvement. This perception underscored the need to break down information silos between the Collection Services divisions and create a better workflow for e-resources acquisitions. As Mackinder pointed out, e-resources workflows "are in a near-constant state of flux by forces that are mostly outside of our control" including "shifting staff dynamics."[37] This holds true at McGill Library, where Collection Development, the division responsible for acquisitions, has faced high turnover since 2012 and new employees face a steep learning curve as they try to cope with the volume of work.

Communication

Poor communication from Collection Services staff was emphasized throughout the survey responses and interviews. It is common practice for e-resources staff to send acknowledgements, and this occurred 65 percent of the time during the project. Beyond showing receipt of the issue, the acknowledgement also demonstrates to the front-line staff member that the question is understood and implies that the issue will be investigated.

Interviewees were split on the usefulness of acknowledgements but over half (57 percent) of survey respondents preferred acknowledgements, particularly for responses that would take "a long time," or "longer than a day." This suggests that e-resources should continue their common practice and send acknowledgements for errors when there are delays.

Front-line staff also asked for more follow-up communication with subscription problems. Unlike fixing broken links, subscription errors typically take extra time to resolve because resolutions require responses from vendors. Given the staffing changes mentioned earlier, Collection Development is particularly low on human resources—the division is strapped and has little time for providing updates. Like Dowdy and Raeford faced at Duke, "there was a lack of transparency with information" that is "time consuming to dig out" and it is difficult to know when something has dropped out of the process.[38]

Recommendations

The recommendations below are specific to McGill Library, but similar improvements could be made in many other libraries It is evident from the project data that the library should continue promoting the single Collection Services email account. Given the popularity of web forms at other academic libraries and the positive feedback received thus far, the new web form for reporting errors should be made available to all staff.

Many of the front-line staff preferences reported throughout the project point to implementing better workflows for reporting access errors and being proactive about managing e-resources. To facilitate workflows, several other divisions in the library use formal ticketing systems, which could be investigated by the e-resources division as a possible solution to showing front-line staff the status of outstanding errors, work assignments and who to contact for more information. This option is popular among libraries according to the Samples and Healy study, as 43 percent of libraries that responded to their survey use a ticketing system to manage errors.[39] Alternatively a simpler, informal approach may be more appropriate, such as a dedicated page on the library's intranet. Both solutions should increase transparency and help with proactive and reactive troubleshooting. Each would need to be evaluated for effectiveness and how much it increases the workload.

In addition to investigating a tracking mechanism, it is clear that all communication surrounding acquisitions needs improvement, both within the Collection Services divisions and with the front-line staff. Librarians in acquisitions may see the complexity of their "acquire" portion of the lifecycle, yet not have much sense of the "provide access" and "provide support" workflows that make what is acquired actually accessible."[40] Pomerantz's research noted the need for staff to collaborate and "to develop a set of best practices for the acquisition of electronic resources" to help cope with the changes in the acquisitions model from print to electronic.[41] As a direct result of this project, a monthly meeting with all Collection Services staff who work on acquisitions-related tasks, regardless of division, was recommended. While Collection Development handles the bulk of this work, the e-books and e-resources and serials staff provide access and troubleshoot problems with new subscriptions, and liaise with front-line staff and vendors. The meetings allow everyone to share information and to collaborate on additional improvements to the workflow. It also helps resolve subscription problems more quickly.

As there were several examples from the survey and interview data that indicated that Collection Services staff were sometimes rude or sarcastic, and that front-line staff sometimes felt that they were "bothering" them, a future look into the tone of responses is warranted. Investigation and resolution of this issue was outside the scope of the project but one possible approach could involve creating template or standard responses when troubleshooting with front-line staff.

The idea that e-resources units should collaborate closely with front-line staff to provide excellent service is repeated in many studies and was demonstrated through this project.[42] As recommended by some survey respondents, Collection Services information sessions on various topics could facilitate such collaboration. Topics could include an overview of each division's primary function and its employees, and open sessions where front-line staff can have their questions answered by a panel of Collection Services staff. It is also recommended that this type of information be added to the library's intranet. Similarly, supervisors from each Collection Services division are encouraged to visit each branch library annually (at a minimum), to facilitate knowledge sharing between front-line staff and Collection Services.

Limitations

As the online survey was sent to all staff and results were anonymous, there is a risk that employees who are not front-line staff responded. There is also a risk that the same person could have completed the survey multiple times. However, it is assumed that these risks are minimal.

Typically two or three people in the e-resources division triage errors. It is a limitation of the method that the same individual triaged the errors during the month of data collection, creating an artificial environment that may have affected response times. Other staff may triage at different rates. To determine whether response times during the pilot are representative of the real working environment, the study would need to be replicated using standard procedures (i.e., having the entire team triage errors). As much as possible, precautions were taken to remind staff to respond at a normal rate (i.e., not faster than usual), and work was done only during business hours to minimize the potential for misrepresentation.

Response times may also have been affected by the way the information was collected, as each method for reporting presents information in a different manner. A web form collects specific, sparse information compared with a phone call. As this is true for work outside of the project as well, there was no tabulation for differences in response rates based on the reporting method.

Another limitation of this study is that it was beyond the scope to track resolution times for errors forwarded to other divisions in Collection Services. Even when resolutions by other divisions were known, the affected 25 percent are listed as "forwarded internally" rather than "resolved." If the project is repeated, full data should be captured to provide a more comprehensive picture.

Conclusion

This project focused on error reporting by front-line staff, identifying how errors are reported and preferences for reporting them. It also shed light on many other areas where Collection Services can improve, including workflows and communication. It demonstrated that in all aspects, from receiving to tracking to resolving errors, that efficiency will improve when Collection Services divisions can successfully communicate and collaborate with each other, and with front-line staff.

References

1. A. S. Chandel and Mukesh Saikia, "Challenges and Opportunities of E-resources," *Annals of Library & Information Studies* 59, no. 3 (2012): 148–54; Beverly Dowdy and Rosalyn Raeford, "Electronic Resources Workflow: Design, Analysis and Technologies for an Overdue Solution," *Serials Review* 40, no. 3 (2014): 175–87, http://dx.doi.org/10.1080/00987913.2014.950040; Jill Emery, "Beginning to See the Light: Developing a Discourse for Electronic Resource Management," *Serials Librarian* 47, no. 4 (2005): 137–47, http://dx.doi.org/10.1300/J123v47n04_13; Jill E. Grogg, "Electronic Resource Management Systems in Practice," *Journal of Electronic Resources Librarianship* 20, no. 2 (2008): 86–89; Richard P. Jasper, "Collaborative Roles in Managing Electronic Publications," *Library Collections, Acquisitions, & Technical Service* 26, no. 4 (2002): 355–61.
2. Jeannette Ho, "Enhancing Access to Resources through the Online Catalog and the Library Web Site: A Collaboration Between Public and Technical Services at Texas A&M University Libraries," *Technical Services Quarterly* 22, no. 4 (2005): 19–37, http://dx.doi.org/10.1300/J124v22n04_02; Taryn Resnick et al., "E-resources: Transforming Access Services for the Digital Age," *Library Hi Tech* 26, no. 1 (2008): 141–54, http://dx.doi.org/10.1108/07378830810857861; Jacquie Samples and Ciara Healy, "Making it Look Easy: Maintaining the Magic of Access," *Serials Review* 40, no. 2 (2014): 105–17, http://dx.doi.org/10.1080/00987913.2014.929483.
3. Chandel and Saikia, "Challenges and Opportunities of E-resources"; Jasper, "Collaborative Roles in Managing Electronic Publications"; Lisa Mackinder, "The Seemingly Endless Challenge: Workflows," *Serials Librarian* 67, no. 2 (2014): 158–65, http://dx.doi.org/10.1080/0361526X.2014.940481; Sarah B. Pomerantz, "The Role of the Acquisitions Librarian in Electronic Resources Management," *Journal of Electronic Resources Librarianship* 22, no. 1–2 (2010): 40–48, http://dx.doi.org/10.1080/1941126X.2010.486726.
4. Janetta Waterhouse, "Managing eResource Processes and Projects," *Serials Librarian* 67, no. 1 (2014): 62, http://dx.doi.org/10.1080/0361526X.2014.899292.
5. Ibid.
6. Mackinder, "The Seemingly Endless Challenge,"158–62.
7. Ibid., 162–65.
8. Dowdy and Raeford, "Electronic Resources Workflow," 176.
9. Feather, "Electronic Resources Communications Management," 208; Resnick et al., "E-resources," 154; Samples and Healy, "Making it Look Easy."
10. Grogg, "Electronic Resource Management Systems," 89.
11. Waterhouse, "Managing eResource Processes and Projects," 64–67.
12. Maria Collins and Jill E. Grogg, "Building a Better ERMS," *Library Journal* 136, no. 4 (2011): 23; Grogg, "Electronic Resource Management Systems," 89.
13. Caitlyn Lam, "Technical Services Report: Report of the ALCTS/LITA Electronic Resources Management Interest Group Meeting, American Library Association Midwinter Meeting, Philadelphia, January 2014," *Technical Services Quarterly* 31, no. 3 (2014): 271–77, http://dx.doi.org/10.1080/07317131.2014.908597.
14. Taryn Resnick, "Core Competencies for Electronic Resource Access Services," *Journal of Electronic Resources in Medical Libraries* 6, no. 2 (2009): 101–22.

15. Ibid., 115–16.
16. Samples and Healy, "Making it Look Easy," 111.
17. Pomerantz, "The Role of the Acquisitions Librarian in Electronic Resources Management," 46.
18. *TERMS: Techniques for Electronic Resources Management* blog, accessed March 1, 2015, http://6terms.tumblr.com.
19. Jill Emery and Graham Stone, "Introduction and Literature Review," *Library Technology Reports* 49, no. 2 (2013): 5.
20. Jill Emery and Graham Stone, "Developing Workflow from TERMS: Techniques for Electronic Resource Management" (workshop presented at the Electronic Resources & Libraries Conference, Austin, TX, March 19, 2014).
21. Mackinder, "The Seemingly Endless Challenge," 160.
22. Feather, "Electronic Resources Communications Management," 228.
23. Samples and Healy, "Making it Look Easy," 110.
24. Colleen Cook and Michael Maciel, "A Decade of Assessment at a Research-Extensive University Library using LibQUAL+®," *Research Library Issues*, 271 (2010): 4–12; Bruce Thompson, Martha Kyrillidou and Colleen Cook, "Library Users' Service Desires: A LibQUAL+ Study," *Library Quarterly*, 78, no. 1 (2008): 1–18, http://dx.doi.org/10.1086/523907.
25. Anita Foster and Sarah C. Williams, "We're All in this Together: Library Faculty and Staff and their Reporting of Electronic Resources Problems," *Journal of Electronic Resources Librarianship*, 22, no. 3–4 (2010): 126, http://dx.doi.org/10.1080/1941126X.2010.535738.
26. Ho, "Enhancing Access to Resources"; Resnick et al., "E-resources."
27. Samples and Healy, "Making it Look Easy," 110; Dowdy and Raeford, "Electronic Resources Workflow," 177.
28. McGill University Library, "Immediate Priorities," accessed July 7, 2015, https://www.mcgill.ca/library/about/planning/immediate-priorities.
29. Foster and Williams, "We're All in this Together," 130.
30. Kittie S. Henderson and Stephen Bosch, "Seeking the New Normal: Periodicals Price Survey 2010," *Library Journal* 135, no. 7 (2010): 36.
31. Dowdy and Raeford, "Electronic Resources Workflow," 176.
32. Samples and Healy, "Making it Look Easy," 111.
33. Dowdy and Raeford, "Electronic Resources Workflow," 187; Samples and Healy, "Making it Look Easy," 114.
34. Feather, "Electronic Resources Communications Management," 208; Resnick et al., "E-resources," 154; Samples and Healy, "Making it Look Easy," 112.
35. Feather, "Electronic Resources Communications Management."
36. Foster and Williams, "We're All in this Together," 130.
37. Mackinder, "The Seemingly Endless Challenge,"158.
38. Dowdy and Raeford, "Electronic Resources Workflow," 176.
39. Samples and Healy, "Making it Look Easy," 110.
40. Ibid., 106.
41. Pomerantz, "The Role of the Acquisitions Librarian," 46.
42. Foster and Williams, "We're All in this Together"; Ho, "Enhancing Access to Resources"; Jasper, "Collaborative Roles in Managing Electronic Publications."

Appendix

The following appendix includes the questions for the online survey and the first page of the survey indicating the participants' consent. The same questions were used in the personal interviews.

1. Think about the most recent occasion when you were unable to find an item record (print or electronic) in WorldCat Local but you were certain that the Library owned or subscribed to that item. What did you do? *Check any that apply*.
   - ○ This has never happened to me / I can't remember.
   - ○ Verbally told my colleague who works in Collection Services.
   - ○ Verbally told a colleague who does not work in Collection Services.
   - ○ Emailed one of the Collection Services general mailboxes.
   - ○ Emailed a Collection Services staff member directly.
   - ○ Emailed a colleague outside of Collection Services (e.g., another librarian or a supervisor).
   - ○ Used the "Catalog Correct" function to report it.
   - ○ Used the "Chat with a librarian" function to report it.
   - ○ I did not report it.
   - ○ Other
2. Think about the most recent occasion when you noticed that a WorldCat Local record was missing some information (such as an e-book record missing the link or a print book missing a call number). What did you do? *Check any that apply*.
   - ○ This has never happened to me / I can't remember.
   - ○ Verbally told my colleague who works in Collection Services.
   - ○ Verbally told a colleague who does not work in Collection Services.
   - ○ Emailed one of the Collection Services general mailboxes.
   - ○ Emailed a Collection Services staff member directly.

- ○ Emailed a colleague outside of Collection Services (e.g., another librarian or a supervisor).
- ○ Used the "Catalog Correct" function to report it.
- ○ Used the "Chat with a librarian" to report it.
- ○ I did not report it.
- ○ Other

3. Think about the most recent occasion when you noticed or suspected a problem with a subscription to a resource (e.g., hitting a paywall when looking for articles in e-journals, unable to access an electronic resource that we subscribe to). What did you do? *Check any that apply*.
   - ○ This has never happened to me / I can't remember.
   - ○ Verbally told my colleague who works in Collection Services.
   - ○ Verbally told a colleague who does not work in Collection Services.
   - ○ Emailed one of the Collection Services general mailboxes.
   - ○ Emailed a Collection Services staff member directly.
   - ○ Emailed a colleague outside of Collection Services (e.g., another librarian or a supervisor).
   - ○ Used the "Catalog Correct" function to report it.
   - ○ Used the "Chat with a librarian" function to report it.
   - ○ I did not report it.
   - ○ Other
4. Think about the most recent occasion when you reported a problem regarding items in the Classic Catalog and/or WorldCat Local.
   a. Did you receive a verbal or email acknowledgement that someone in Collection Services has received your error report? *Choose one of the following answers*.
      - ■ Yes
      - ■ No
      - ■ Unsure

   b. How long did it take for you to receive an answer or resolution regarding the reported problem? (Please choose the closest response, even if you received an answer but were not satisfied with the resolution.)
      - ■ A few minutes
      - ■ Within the same day
      - ■ The next working day
      - ■ Within a week
      - ■ Within a month
      - ■ I never heard back about the problem
      - ■ Unsure
      - ■ Other / Comments:

   c. Still thinking of this same occasion, would you consider this to be a typical response time for receiving answers/resolutions to errors/problems reported to Collection Services? *Choose one of the following answers*.
      - ■ No, it took less time than usual to receive a response.
      - ■ No, it took longer than usual to receive a response.
      - ■ Unsure
5. After you report an error found in the Classic Catalog or WorldCat Local, would you prefer to receive an acknowledgement that someone in Collection Services has received your error report, even if an answer or resolution cannot be provided right away? *Choose one of the following answers*.
   - ○ Yes, I prefer an email acknowledgement.
   - ○ Yes, I prefer a verbal acknowledgement.
   - ○ Yes, I prefer either an email or verbal acknowledgement.
   - ○ No, I prefer not to receive an acknowledgement. I prefer only to be informed when the issue has been resolved.
   - ○ It doesn't matter to me.
   - ○ Unsure
   - ○ Other
6. When you report an error found in the Classic Catalog or WorldCat Local, what is the preferable time frame for a response to be communicated? (Response in this case means that your question has been addressed, the error has been fixed, your question has been referred to someone else, or a tentative course of action has been presented; it does not necessarily mean you have received a satisfying resolution.) *Choose one of the following answers*.
   - ○ A few minutes
   - ○ Within the same day
   - ○ By the next working day
   - ○ Within a week
   - ○ Within a month
   - ○ Depends on the problem
   - ○ I don't have expectations for response times
   - ○ Unsure
   - ○ Other
7. Think about an occasion you reported an error found in the Classic Catalog or WorldCat Local and received a response from someone who works in Collection Services. *Choose one of the following answers*.
   a. Did you receive a response that answered your question?
      - ■ Yes
      - ■ Somewhat
      - ■ No
      - ■ Unsure
      - ■ No answer

b. Still thinking of the same occasion, were you satisfied with the response you received? Choose one of the following answers
   - Yes
   - Somewhat
   - No
   - Unsure
   - No answer

8. a) Select the statement that best describes you. Typically, when I report errors regarding items in the Classic Catalog to Collection Services, I: *(Choose one of the following answers.)*
   a. Feel like my problems are addressed in a timely manner.
   b. Feel like my problems are addressed eventually but they are not a priority.
   c. Feel like my problems are rarely addressed or not looked into at all.
   d. Feel like my problems are noted but are part of a larger problem that has not yet been resolved.
   e. Depends—sometimes a, b, c or d.

   b) Select the statement that best describes you. Typically, when I report errors regarding items in the WorldCat Local to Collection Services, I: *(Choose one of the following answers)*
   f. Feel like my problems are addressed in a timely manner.
   g. Feel like my problems are addressed eventually but they are not a priority.
   h. Feel like my problems are rarely addressed or not looked into at all.
   i. Feel like my problems are noted but are part of a larger problem that has not yet been resolved.
   j. Depends—sometimes a, b, c or d.

9. Select the statement that best describes you. Typically, when I report errors regarding items in the Classic Catalog and/or WorldCat Local to Collection Services, I: *(Choose one of the following answers.)*
   a. Feel like the problems are resolved to the best of the staff's abilities.
   b. Feel like Collection Services is aware of the problem but they do not or cannot resolve it.
   c. Feel like my particular case has been noted but it is part of a larger problem that has not yet been resolved.
   d. Feel like my problems are eventually resolved but they are not a priority.
   e. Depends—sometimes a, b, c or d.

10. a) Think about a time when you were not satisfied with a response that you received for reported error or problem. What could Collection Services staff have done differently?

    b) Do you have any other comments you wish to include, relating to errors and questions sent to Collection Services?

Book Reviews

Elyssa M. Gould

Digitizing Audiovisual and Nonprint Materials: The Innovative Librarian's Guide. By Scott Piepenburg. Santa Barbara, California: Libraries Unlimited, 2015. 94 p. $50.00 softcover (ISBN 978-1-4408-3780-7).

Anyone who works in a library knows that audiovisual materials can disintegrate and their playback equipment can quickly become obsolete. Does anyone remember the Betamax, or how about the laser disc? Digitization can be the solution to this problem. There are a handful of companies that will take care of this process for you, but if you are a do-it-yourselfer, then Piepenburg's new book, *Digitizing Audiovisual and Nonprint Materials*, is for you.

The best part is you do not have to be a rocket scientist to understand what Piepenburg has written. Anyone with a minimum understanding of technology can learn and follow the instructions in this book. The easy-to-read, conversational style book is a no-nonsense, step-by-step instruction manual. The author takes you through the entire process, starting with what to consider before taking on a project of this nature, the space requirements needed, the hardware and software required, and then focusing on both audio files and sound recordings as well as more complex video files. The book considers some of the more common audiovisual materials libraries have collected over the last half century, including "photographs, slides, records, cassettes, videotapes, and laserdiscs" (ix).

An entire chapter is devoted to hardware requirements and subsequent chapters provide greater detail about the hardware and software needed to capture a particular format, such as slides or sound recordings. In addition to the obvious hardware needs—computers, monitors, speakers, and scanners—the book covers other items most people probably have not considered, such as disc-labeling software. The author also discusses minute details such as how to name your files and where to save them (either on the computer's hard drive or backing them up to a larger separate storage device).

The book is divided into six chapters. The first two chapters cover such basics as things to consider before undertaking a digitization project, including some basic issues like space, lighting, and furniture. For example, if you are digitizing audiovisual materials, is there a secluded space where the noise and the music will not disturb staff and patrons? Is the electrical service adequate and does it have proper ventilation? Heat can wreak havoc on electrical equipment. Subsequent chapters are devoted to digitizing photographs and slides, capturing and editing sound recordings, and working with various video formats.

Each chapter ends with a checklist reiterating the important points. The book also has an eleven-page glossary. The book is very graphic intensive, with lots of pictures and charts explaining the various technologies and software needed for these types of projects. Piepenburg goes so far as to highlight various pieces of the hardware (inputs and outputs) in the photographs, making it easy to follow his directions. The charts, as well, make it simple to decipher the technologies.

Piepenburg gets a little technical at times, but it is nothing most librarians would not understand. In the two chapters on digitizing photographs and slides, he discusses the do's and don'ts of various formats. For example, he advises "save the image as a .TIFF format as it is lossless, albeit more space intensive. Don't use .JPEG as it is a 'lossy' format and will not serve well if the image is later enlarged electronically" (22). The book ends with a chapter called "Finishing Up," which not only discusses storage of the digitized material, but provides helpful hints on how to store the originals, such as LP records, 8-track tapes, and CDs.

Most chapters are short, easy reads, and thirty-two sophisticated pages are devoted to audio recordings, where Piepenburg goes into particular detail about how to capture audio from records (LPs), cassette and tape decks, and other sources. He discusses the recording, editing, and exporting process for audio sources, providing details on how to use the freeware Audacity. In the chapter on capturing video, Piepenburg not only makes software suggestions, but also shares a particular video capturing and editing package that he has used.

Piepenburg is a cataloger by profession and obviously thinks like one. He suggests scanning and saving everything. Even if you do not plan to use the digitized master copies, the cataloging staff may need these items later as they create the metadata for the catalog. At one point in the book he suggests organizing images by topic first (for example churches), then geographic location. He reasons that "the thought process behind Library of Congress Subject Headings (LCSH) is that 'place' is often the first subdivision in the cataloging subject heading string since that is most often how people are looking for information on a specific topic" (16). For example, a patron looking for pictures of old churches (main heading) in the United States (location being the first subdivision). In all formats he suggests saving the item "as is" and editing later for either image or sound correction. Researchers will want the master copy, while the public will most likely want the cleaned up version.

This small book is packed with information and librarians of any caliber will find it easy to follow Piepenburg's instructions to begin a digitization project of their own. The low-barrier technical threshold should not deter anyone. The book ends with the advice to "have fun." Librarians and archivists will enjoy reading this fast-paced book and most likely learn a thing or two in the process.—*Brian F. Clark (bf-clark@wiu.edu), Western Illinois University, Macomb, Illinois*

Fundamentals of Technical Services. By John Sandstrom and Liz Miller. Chicago: ALA Neal-Schuman, 2015. 213 p. $64.00 softcover (ISBN 978-1-55570-966-2). ALA Fundamentals Series.

This latest monograph in the ALA Fundamentals Series continues the series' mission of providing a broad overview of an area of library science. Written by a Cataloging Librarian and an Acquisitions Librarian from New Mexico State University Library, *Fundamentals of Technical Services* communicates the conceptual practices clearly and succinctly. The tone of writing clearly conveys the authors' enthusiasm and passion for technical services and emphasizes the crucial role that technical services staff play in providing access to resources through purchasing, cataloging, physical processing, and authority control.

This book begins with a chapter describing the management of technical service departments, followed by a chapter on library systems. The subsequent six chapters follow the general workflow of technical services: "Collection Development," "Acquisitions," "Cataloging," "Physical Processing," "Authority and Catalog Maintenance," and "Collection Management." Each chapter provides basic foundational knowledge; lists of key terminology with clear definitions, trends and issues related to each chapter's subject; and recommended print resources for further reading. As is stated multiple times thorough out the text, this is an introductory text, and as such, does not contain vast quantities of historical information. Though the book can certainly be read through in its entirety, each chapter could be consulted distinctly as a surface introduction to that area of technical services, supplemental to more substantial works.

The introduction explains that it was intended for use by library science students and as a resource for staff or faculty whose positions have been reassigned to technical services departments. The latter audience is strongly emphasized. The initial section in each chapter is titled "Before you Begin" and instructs readers to answer questions about their library's current practices, to collect institutional policies, or to identify staff attitudes about a subject before reading the chapter. It is unlikely that a graduate student would have access to this information.

Each chapter contains reading aids that display concepts graphically or elaborate on associated topics. Sidebar texts are included for related concepts such as library security systems in the chapter on Physical Processing, or listing tips for holding effective meetings in the Managing Technical Services chapter. Diagrams of basic workflows included in both the Acquisitions chapter for monograph and serial acquisitions and in the Cataloging chapter for the cataloging workflows for physical and electronic materials display concepts that would have been tedious to explain solely within the text. The book also includes four well-written yet brief appendixes about specific cataloging-related topics: "Content Standards" outlines RDA and its differences from AACR2; "Classification Systems and Call Numbers" depicts the Dewey Decimal and Library of Congress classification systems; "Subject Term Lists" describes the utility of providing subject access, and Library of Congress Subject Headings specifically; and finally "MARC Digital Format" describes MARC bibliographic and MARC holdings formats, with descriptions of selected common fields. Finally, the terminology and definitions found in every chapter are compiled into a glossary.

Each chapter concludes with a brief description of trends, and though these trends are those that are most current, they are handled with excessively broad strokes. Another idiosyncrasy of this book is the inclusion of incongruous statements: in a discussion of budget meetings the reader is cautioned to "keep your facial expressions neutral at all times. . . . Believe it or not, a poorly timed arched eyebrow can change the entire tone of a discussion" (47). At random, an individual Milwaukee Public librarian's searches being redirected in her catalog are cited as examples of successful authority control (129–30), though there is no explanation of why she was mentioned or included. While neither of these comments is inappropriate, they lend a certain chattiness that seems inconsistent with the tone of the remaining text. One final small criticism is that the acronym OCLC is only explained using its initial usage and not the current fuller form (86).

Despite the uneven tone, all chapters were well-organized, accessible, and enjoyable to read; the chapters on Acquisitions and Cataloging were particularly well-conceived. I was very pleasantly surprised to find chapters that included discussions of both cataloging maintenance and collection maintenance, since similar texts frequently only discuss these areas in a cursory manner. The concluding "Collection Maintenance" chapter includes descriptions of activities that could involve staff from multiple areas of the library (i.e., not only technical services staff) such as the review of gift materials for possible addition to the library collection, deselection or weeding, the replacement of lost or missing items, disaster preparedness, and the repair of damaged library resources.

Finally, this book does not describe best practices or precise workflows because these could vary widely by library. While the components of technical services that are included in this text are certainly necessary in any technical services department, they may not apply in all situations or in all libraries, particularly those with smaller staff. This well-written, enthusiastic text provides a great introduction to the many aspects of Technical Services.—*Julene L. Jones (julene.jones@uky.edu), University of Kentucky Libraries, Lexington, Kentucky*

The Critical Component: Standards in the Information Exchange Environment. Edited by Todd A. Carpenter. Chicago: Association for Library Collections & Technical Services, 2015. 298 p. $71.00 softcover (ISBN: 978-0-8389-8744-5).

People may not realize it, but standards are all around us. The seat belt in cars that keeps passengers safe, the power outlets found in homes to plug an electronic device, or the JPEG file that is downloaded to view a friend's picture all adhere to standards. If it were not for these standards, many of the things people do would not be possible. Information professionals in particular need standards to communicate, access, retrieve, and display information. *The Critical Component: Standards in the Information Exchange Environment* provides the reader with an understanding and appreciation for the standard creation process in the information environment. It highlights some of the most important standards in the library profession, from the description standards that librarians and publishers use for resources, to the various types of identifiers found in the information supply chain.

The book is divided into eleven chapters and nine case studies. Each chapter highlights a segment of the standard creation process. Chapters 2–10 include case studies that describe an information standard which in turn illustrates the context found in the chapter. For example, chapter 5 deals with the role of identifiers in content management and distribution. Several standards have been created to fulfill this part, such as the international standard book number (ISBN) and digital object identifier (DOI). The transactions that occur between organizations in the information supply chain (e.g., libraries, publishers, distributors, and content providers) can be difficult to manage without the use of identifiers. Identifiers help these organizations differentiate one resource from another. The case study that supplements chapter 5 is on the international standard name identifier (ISNI). As the book explains, ISNI strives to make people and organizations discoverable; this role helps stewards of information identify entities involved with an intellectual work. ISNI also supports effective management in rights payments and rights clearance, which is a vital task for publishers. Without ISNI these communities would have to identify authors or organizations through their own means rather than using one identifier to resolve their individual needs; hence a standard that created cost-effective solutions.

The book starts by discussing the importance of standards; how standards are developed and the organizations that contributed to their development; the information standards landscape; and the basic concepts that are undertaken by information standards. Chapters 5–8 address type-specific standards that affect the information community such as identifiers, descriptive metadata schemas, and digital preservation. Chapter 7 at first glance seem out of place in these chapters since it discusses discoverability as a goal, but after reading about identifiers, metadata, and preservation standards one realizes that they are all created with the goal of increasing discoverability. Chapters 9–11 return to the general standard process (similar to Chapters 1–4), explaining how standards are marketed, how to get involved, and the future of information standards. The concluding chapter is especially revealing because the editor features some of the future standards that are being worked on or that will eventually be developed. Examples of these upcoming standards are in the area of rights management and open access.

A definite highlight for readers will be to learn about all the organizations involved with the creation of standards. Similar to the saying "it takes a village to raise a child," so is the case when creating a standard. As consumers of information, it is reassuring to know that organizations like the American National Standards Institute (ANSI) and the National Information Standards Organization (NISO) exist to support standards and to guarantee that the process meets the needs of its constituents. Along with these organizations, the individuals that are involved in making a standard a reality put in a great effort to move the professional agenda forward. Without the dedication of these volunteers, standards would not be feasible. As stated by the editor, "Standards provide the most effective strategy for addressing the issues in a way that provides a framework for all community members to contribute to the process" (291). Standards help the professional community come together to solve a common problem and to accomplish a goal.

When looking for other books that discuss information standards, it is difficult to find something that is as all-encompassing as this publication. Now, a reader would have to search for individual resources to learn about the various standards mentioned in the book. This book serves as a convenient resource for readers by providing a basic overview of several important information standards in the field. The book is also valuable in that each chapter and case study is authored by individuals that were involved with making a particular standard or have extensive experience in the standard creation process. Many of the contributors are well known and respected authorities in the field of information science and technology.

Despite the content of the book being introductory, it can be very technical at times, especially when it discusses a particular standard. If a reader does not have experience working with information schemas or systems it can be difficult to understand certain case studies. The language and acronyms used throughout the book can be vast for a first time reader in information protocols. An index would have been helpful in allowing a reader to reference back on a particular term.

In conclusion, the information environment can always seem to be in flux. Although this may be the case, standards have withstood this fluctuation and provided much needed stability in the library field. The book showcases how information standards are a critical component to the world of information. Its distinctive content makes it a unique resource in the literature of information science, where standards, for the most part, have not been reconciled in this manner, with the standard creation process and real world examples. With its emphasis on the importance of standards and how they play a major role in the work of information professionals, the book leaves readers thinking that standards should take a higher priority in our professional lives. As practitioners, it is vital to consider being a part of the standards creation process whether by volunteering, providing feedback when standards are under review, or just being aware of and using them!—*Heylicken Moreno (hmoreno3@central.uh.edu), University of Houston, Houston, Texas*

The Complete Guide to Acquisitions Management. 2nd ed. By Frances C. Wilkinson, Linda K. Lewis, and Rebecca L. Lubas. Westport, CT: Libraries Unlimited, 2015. 208 p. $60.00 softcover (ISBN 978-1-61069-713-2).

The second edition of *The Complete Guide to Acquisitions Management* is an update to the first edition, which was published in 2003. Since then, the market in which an acquisition librarian operates has seen many changes. New technological developments affecting business models, products, and services have had an impact in the management and operation of today's library acquisitions department. The same holds true today as it did in 2003: "Acquisitions librarians must continue to learn and remain flexible in order to meet the needs of their libraries and customers" (xiv).

This book is organized in two ways. First, it is a guide book that any library school student or practicing librarian interested in acquisitions would find useful. It is interspersed with practical knowledge and contains an abundance of definitions and resources in the glossary and the appendixes. Second, as an update to its first edition, it describes how new technology has had an impact on each of the areas of acquisitions.

The purpose and organization of acquisitions departments are presented in the context of the role that acquisitions plays within the organization. "The mission and goals of the acquisitions operation should align with those of the library" (1). The role of acquisitions is blending more with the roles of the other units in the library. This blending is caused by technological developments and their applications. "Acquisitions has morphed from meaning exclusively the purchase and/or receiving of physical materials to including the licensing and obtaining access to electronic content" (1).

Acquisitions librarians are encouraged to understand the impacts that new technologies and that recent economic pressures have had on the publishing industry, as these impacts will affect their policies and workflows. Wilkinson, Lewis, and Lubas explain that economic pressures caused by the recent recession and changes in the marketplace have eroded profit margins for publishers. According to publishers, the effort to maximize profit is for sustainability. However, there is a feeling in the market that the only interest for publishers is in making as much money as fast as possible. Wilkinson, Lewis, and Lubas navigate through these issues and explore the wide variety of pricing models as they affect acquisitions, library funding and budgets, and library collections and access.

Acquisitions librarians also need to understand the physical acquisition systems in which they operate. These systems can range from a stand-alone system to a part of a more encompassing integrated library system. It is important to understand the features and functions in an acquisitions system to utilize them in the most efficient way. Through the explanation of these aspects of an integrated acquisitions system, Wilkinson, Lewis, and Lubas walk us through how it can be a tool to perform the daily tasks of acquisitions, provide information to other library units, address the technological developments of electronic resources, and to accomplish the fiscal tracking and reporting that the institution requires.

Understanding the acquisition of monographic content sets a solid foundation for acquisition librarians and students looking to enter the field. "In some ways, all other purchasing models flow from book buying. The book is a basic unit of content" (51). Buying a book may sound like a simple task, but the acquisition process of purchasing monographic content is driven by many decisions. These decisions lay in the different purchasing models offered by the many vendors who compete for the libraries' business. From firm ordering to approval plans, from print books to streaming media, Wilkinson, Lewis, and Lubas discuss what goes into these decisions by clearly defining the steps in the acquisitions processes and even some of the legal issues involved in acquiring different media formats.

After understanding the acquisition of monographic resources, Wilkinson, Lewis, and Lubas take us into the complex world of acquiring continuing and electronic resources. "The Internet and availability of electronic content have changed the course of library collections" (79). As a majority of the acquisitions budget is spent on these recurring purchases, we are offered a significant presentation of the different avenues of continuing resources acquisition and the impacts of each one. This includes the maintenance involved in the print and electronic serials acquisition, the knowledge required to work with the systems that can track and manage continuing resources, the economics of a world in transition to more digitized content but where print is still produced, and a discussion of different pricing models.

Wilkinson, Lewis, and Lubas give a sound overview of the complicated world of licensing for electronic resources. Institutional lawyers, purchasing officers, and librarians are all involved in the reviewing and negotiating process. It is increasingly important to understand the clauses that appear in licenses, to have an understanding of fair use and copyright law, and to know your library's goals to ensure that you obtain licenses that meet your needs.

"Acquisitions librarians are required both legally and professionally to safeguard the budgets over which they have control and to assure themselves that dollars are being spent wisely and efficiently."[1] This requires librarians to take a more formal approach in acquisitions and lead a more competitive process to obtain vendor products and services, a formalized process known as the Request for Proposal (RFP). According to the authors, the RFP provides for an objective evaluation of different solutions where service is just as much a consideration as price. The planning, participants, timeline, document, evaluation, and awarding of a contract are all clearly described. This is vital knowledge for anyone looking to enter or move forward in the profession.

Aside from the mechanics of acquisitions, Wilkinson, Lewis, and Lubas appropriately include a chapter on professional ethics. "Our need for ethical awareness has grown as the practice of librarianship has become more professional, our roles and services more complex, and information technologies faster and more pervasive."[2] Acquisitions librarians have control over a large portion of the budget, input into purchases with a large dollar value as well as input into vendor selection, and increasing complexity in negotiating licenses and contracts. So, the chapter moves forward with the discussion of values and ethics as it pertains to the decisions that affect our personal and professional decisions. Wilkinson, Lewis, and Lubus state that "such discussions may help teach those new to the field about the core philosophy of the profession and may illuminate values for those who are more experienced."[3] Notably, the chapter includes a rigorous analysis of the ALCTS Acquisitions section's Statement on Principles and Standards of Acquisitions Practice offering greater specificity for ethical guidelines for an acquisitions librarian.

Lastly, Wilkinson, Lewis, and Lubas look at how acquisitions has the opportunity to play a major role in improving the user experience in the library. The technological changes that have been presented throughout the book emphasize that the work of acquisitions is connected with other library operations to improve the discovery, delivery, and preservation of our collections. "Articulating how you are part of the plan and how you are incorporating new directions and efforts to complement and further the part will help make your department a recognized, vital part of the library" (167). This book can stand alone as a complete guide to library acquisitions management or each chapter can stand alone addressing a specific topic. As such, this book is recommended for students in LIS masters programs, as well as current and future acquisitions librarians looking to develop or enhance skills in acquisitions management.—*Lee Sochay (sochayle@msu.edu), Michigan State University, East Lansing, Michigan*

References

1. Karen E. Cargill, "Vendor Evaluation," in *Understanding the Business of Library Acquisitions*, 2nd ed., edited by Karen A. Schmidt. (Chicago: American Library Association, 1999), 94.
2. Jean Preer, *Library Ethics* (Westport, CT: Libraries Unlimited, 2008), xiii.
3. Rob Richards, "Biz of Acq-Stewardship, Partnership, Self-Understanding: An Exploration of Values in Acquisitions Work," *Against the Grain* 12, no. 3 (2000): 88.

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