



Making Archival and Special Collections More Accessible

Making Archival and Special Collections More Accessible

Foreword by James Michalko



© 2015 OCLC Online Computer Library Center, Inc.

This work is licensed under a Creative Commons Attribution 3.0 Unported License.
<http://creativecommons.org/licenses/by/3.0/>



April 2015

OCLC Research

Dublin, Ohio 43017 USA

www.oclc.org/research

ISBN: 1-55653-493-0 (978-1-55653-493-5)

OCLC Control Number: 906730124

Please direct correspondence to:

James Michalko

Vice President, OCLC Research Library Partnership

michalkj@oclc.org

Suggested citation:

Making Archival and Special Collections More Accessible. 2015. Dublin, Ohio: OCLC Research.

<http://www.oclc.org/content/dam/research/publications/2015/>

[oclcresearch-making-special-collections-accessible-2015.pdf](http://www.oclc.org/content/dam/research/publications/2015/oclcresearch-making-special-collections-accessible-2015.pdf)

Acknowledgments

Special thanks to the following colleagues for their help with compiling and publishing this report: Melissa Renspie, Linda Shepard and Eric Childress.

Contents

1	<i>Foreword, by James Michalko</i>	1
2	<i>Taking Our Pulse: The OCLC Research Survey of Special Collections and Archives – Executive Summary, by Jackie M. Dooley and Katherine Luce.....</i>	5
3	<i>Survey of Special Collections and Archives in the United Kingdom and Ireland – Summary and Recommendations, by Jackie M. Dooley, Rachel Beckett, Alison Cullingford, Katie Sambrook, Chris Sheppard, and Sue Worrall</i>	11
4	<i>Taking Stock and Making Hay: Archival Collections Assessment, by Martha O'Hara Conway and Merrilee Proffitt.....</i>	17
5	<i>Over, Under, Around and Through: Getting Around Barriers to EAD Implementation, by Michele Combs, Mark A. Matienzo, Merrilee Proffitt, and Lisa Spiro.....</i>	39
6	<i>Thresholds for Discovery: EAD Tag Analysis in ArchiveGrid, and Implications for Discovery Systems, by Marc Bron, Merrilee Proffitt and Bruce Washburn</i>	63
7	<i>The Metadata IS the Interface: Better Description for Better Discovery of Archives and Special Collections, by Jennifer Schaffner</i>	85
8	<i>“Capture and Release”: Digital Cameras in the Reading Room, by Lisa Miller, Steven K. Galbraith, and the RLG Partnership Working Group on Streamlining Photography and Scanning</i>	99
9	<i>Tiers for Fears: Sensible, Streamlined Sharing of Special Collections, by Dennis Massie</i>	113

1

Foreword

James Michalko
Vice President,
OCLC Research Library Partnership

.....

"Libraries, archives, and cultural institutions hold millions of items that have never been adequately described. These items are all but unknown to, and unused by, the scholars those organizations aim to serve. ...Nationally, this represents a staggering volume of items of potentially substantive intellectual value that are unknown and inaccessible to scholars."

—from the CLIR announcement of the Cataloging Hidden Special Collections and Archives program, 17 March 2008.

.....

.....

"What is important about books and serials is that moving digital surrogates and newly produced works to the network level generates aggregations operating at a scale that advances existing lines of inquiry and opens new ones and makes scholars and students more productive, even when using individual works. These same criteria must form the heart of the value proposition for special collections."

—Donald J. Waters in “The Changing Role of Special Collections in Scholarly Communications.” *Research Library Issues* (A bimonthly report from ARL, CNI, and SPARC), no. 267 (Dec. 2009): 35-36.

.....

OCLC Research supports research institutions in collaboratively designing their future. This leads us to work in areas that will reduce redundant efforts, change community economics, transform processes and respond to known and emerging needs. Revealing the hidden assets stewarded by research institutions so they can be made available for research is a prime opportunity for creating and delivering new value. Over the last seven years we have worked to support change in the end-to-end process that results in archival and special collections materials being delivered to interested users.

The overarching goal of this work continues to be the achievement of economies and efficiencies that permit these materials to be effectively described, properly disclosed, successfully discovered and appropriately delivered. Achieving control over these collections in an economic fashion will mean that current resources can have a broader impact or be invested elsewhere in other activities.

Special collections and archives are part of almost every library collection no matter the size or type of library. They are ubiquitous at research institutions and present a similar schizophrenic management challenge across the community. They are treasures. They are burdens. They are valued. They are costly. They are important. They are hidden.

I have frequently framed the goal of the concerted attention that OCLC Research has given to archives and special collections as a change in the cost/benefit equation associated with these types of materials. There are two ways to increase the ratio. You can affect the denominator by reducing the ongoing investment necessary to steward these materials, or you can change the numerator by increasing the utility of the materials for teaching and research both locally and globally. Our work has tried do both. We captured that intent in the phrase “Mobilizing Unique Materials.”

As is often the case with OCLC Research we began our work in this area with a system-wide view that described the scope of the opportunity and the problem. Our two surveys—*Taking Our Pulse: The OCLC Research Survey of Special Collections and Archives in the US and Canada* and the *Survey of Special Collections and Archives in the UK and Ireland*—were large multi-year efforts that brought the community an updated understanding of the

state of these types of materials. The executive summaries of each are included in this volume and include recommendations for action that are notable for their good sense and moderated scope, and will have significant impact. The counts, charts and associated analysis that were the heart of this effort are available from the OCLC Research website. They continue to be among the most referenced of our work in this area.

Understanding the scope and range of unprocessed and therefore hidden collections made it clear that the community needed renewed encouragement to take on the daunting task of describing these collections. In *Taking Stock and Making Hay: Archival Collections Assessment*, we urged institutions to undertake an accurate census of their archival collections as a foundation for acting strategically in meeting user needs, allocating available resources, and securing additional funding. This kind of data about collections informs important decisions regarding collection management, processing priorities, and selection and other activities associated with digitization and exhibit preparation. Understanding that resources for this type of work are scarce and the scale can be daunting, this report provides a variety of practical approaches that encourage progress by avoiding prescriptions.

Of course, understanding the scope of local collections does not mean that they are properly described, which is a necessary condition for them to be discovered. For more than a decade the target for describing archival collections has been the Encoded Archival Description (EAD) standard, which has long been considered a high value but costly descriptive mechanism. Consequently it has been implemented by a bare majority of archives, and many institutions have been daunted by its related political, logistical and technical issues. In *Over, Under, Around and Through: Getting Around Barriers to EAD Implementation* we offered tools—information, persuasion and technology—to help practitioners surmount these roadblocks. This enormously practical document encourages by providing lots of examples and offering simplification of the unnecessarily complex.

This complexity grows over time as standards expand, practice changes, and local choices drive the descriptive effort. How well the resulting descriptions actually serve to advance the discovery of materials has not been much studied in the literature. Certainly it has not been rigorously demonstrated or seriously challenged. In *Thresholds for Discovery: EAD Tag Analysis in ArchiveGrid, and Implications for Discovery Systems* we add to the evidence base by examining the large (120,000+) corpus of EAD documents harvested and made available through the ArchiveGrid aggregation maintained by OCLC Research. Our analysis looks to determine how well the documents support various aspects of online discovery. And we conclude that the picture for archival discovery and EAD is decidedly mixed. We also offer advice about where it would be worthwhile to invest local descriptive effort in elements particularly crucial to discovery.

Understanding the way tags are used in systems that support search, browse, results displays, sorting and limiting is one very important way to evaluate the investment in description. It proceeds, however, from a system vantage. Guiding this investment on the basis of what users need and value is the intent of the synthesis of user studies that we provided in *The Metadata IS the Interface: Better Description for Better Discovery of Archives and Special Collections*. This well organized, definitive survey concludes that we best respond to users by putting the right descriptive metadata in the right places. Those places are network discovery environments not local portals.

Finally the volume concludes with two very influential pieces that urge practitioners to go beyond traditional bounded practices in order to satisfy their users who have discovered special collections and archival materials. In “*Capture and Release*”: *Digital Cameras in the Reading Room* we acknowledge the ubiquity of digital cameras and other mobile capture devices which has led researchers to expect to use cameras in reading rooms. We argue that embracing and supporting the use of these devices provides benefits to researchers, repositories, and collection materials. We provide advice to support changes in practice that will satisfy on-site researcher expectations and are consistent

with institutional practice. Special collections cannot release their value if they are camera-shy.

Nor can they be valued if they only stay home. In *Tiers for Fears: Sensible, Streamlined Sharing of Special Collections* we argue that interlending of actual physical items from special collections for research purposes should be supported. Special collections have long done this for exhibitions but providing the physical item to the distant scholar is rare and elicits a fear response in many special collection managers. This exhortatory piece acknowledges that trust is essential to establishing new lending practices and helps practitioners evaluate the tiers of effort required to lend certain materials and the trustworthiness of the other parties to the transaction. I’m pleased to say that this piece has changed practice at an impressive cadre of world-class special collections.

This volume stands as evidence of a body of effort devoted to areas that we think hold enormous promise for enhancing the library’s value proposition. The unique materials stewarded by our institutions need to release their value to a global audience of researchers in ways that will enhance the reputation of the steward. That will happen only when we devote structured effort to the full range of selection, description, discovery and delivery. With the right attention as signaled by the pieces in this volume we can mobilize our unique materials for maximum value in the networked environment.

2

Taking Our Pulse: The OCLC Research Survey of Special Collections and Archives

Executive Summary

Jackie M. Dooley
Program Officer

Katherine Luce
Research Intern

OCLC Research

This executive summary was originally published in October 2010 by OCLC Research in *Taking Our Pulse, The OCLC Research Survey of Special Collections and Archives* at <http://www.oclc.org/content/dam/research/publications/library/2010/2010-11.pdf>.

Executive Summary

Special collections and archives are increasingly seen as elements of distinction that serve to differentiate an academic or research library from its peers. In recognition of this, the Association of Research Libraries conducted a survey in 1998 that was transformative and led directly to many high-profile initiatives to “expose hidden collections.”

As this OCLC Research report reveals, however, much rare and unique material remains undiscoverable, and monetary resources are shrinking at the same time that user demand is growing. The balance sheet is both encouraging and sobering:

- The size of ARL collections has grown dramatically, up to 300% for some formats
- Use of all types of material has increased across the board
- Half of archival collections have no online presence
- While many backlogs have decreased, almost as many continue to grow
- User demand for digitized collections remains insatiable
- Management of born-digital archival materials is still in its infancy
- Staffing is generally stable, but has grown for digital services
- 75% of general library budgets have been reduced
- The current tough economy renders “business as usual” impossible

The top three “most challenging issues” in managing special collections were space (105 respondents), born-digital materials, and digitization.

- We updated ARL’s survey instrument and extended the subject population to encompass the 275 libraries in the following five overlapping membership organizations:
 - Association of Research Libraries (124 universities and others)
 - Canadian Academic and Research Libraries (30 universities and others)
 - Independent Research Libraries Association (19 private research libraries)
 - Oberlin Group (80 liberal arts colleges)
 - RLG Partnership, U.S. and Canadian members (85 research institutions)

The rate of response was 61% (169 responses).

Key Findings

A core goal of this research is to incite change to transform special collections, and we have threaded recommended actions throughout this section. We focused on issues that warrant shared action, but individual institutions could take immediate steps locally. Regardless, responsibility for accomplishing change must necessarily be distributed. All concerned must take ownership.

Assessment

A lack of established metrics limits collecting, analyzing, and comparing statistics across the special collections community. Norms for tracking and assessing user services, metadata creation, archival processing, digital production, and other activities are necessary for measuring institutions against community norms and for demonstrating locally that primary constituencies are being well served.

- ACTION:** Develop and promulgate metrics that enable standardized measurement of key aspects of special collections use and management.

Collections

ARL collections have grown dramatically since 1998, ranging from a 50% increase in the mean for printed volumes and archival collections to 300% for visual and moving-image materials. Two thirds of respondents have special collections in secondary storage. As general print collections stabilize, such as through shared print initiatives and digital publication, a need for more stacks space for special collections will become all the more conspicuous. The arguments to justify it will have to be powerful.

The amount of born-digital archival material reported by respondents is minuscule relative to the extant content of permanent value: the mean collection size is 1.5 terabytes, the median a mere 90 gigabytes. It is striking that only two institutions hold half of the material reported, and only thirteen hold 93% of it.

Receipt of a gift is the most frequently stated impetus for undertaking a new collecting emphasis. Some respondents noted, however, that they do not plan to acquire other materials to strengthen the new area, which may signal that the gift was outside the library's areas of strength or need. Such gifts sometimes become a liability over time. Deaccessioning of unwanted materials, some of which have languished unprocessed for years, occurs for appropriate reasons but is not widely practiced. Informal collaborative collecting is fairly widespread on a regional basis, but formal arrangements of any kind are rare.

- ACTION:** Identify barriers that limit *collaborative collection development*. Define key characteristics and desired outcomes of effective collaboration.

The preservation needs of audiovisual collections (both audio and moving image) are well known to be staggering, and our data confirm that these materials have by far the most serious problems.

- ACTION:** Take collective action to share resources for cost-effective *preservation of at-risk audiovisual materials*.

User Services

More than 60% of respondents stated that use by faculty, undergraduates, and visiting researchers has increased over the past decade. Nearly half, however, were unable to categorize their users by type, even those in their primary user population.

User services policies are evolving in positive ways: most institutions permit use of digital cameras and 90% allow access to materials in backlogs. More than one third send original printed volumes on interlibrary loan, while nearly half supply reproductions. Conservative vetting of requests may, however, result in unwarranted denial of all three types of access.

- ACTION:** Develop and liberally implement exemplary policies to *facilitate* rather than inhibit *access to and interlibrary loan* of rare and unique materials.

Cataloging and Metadata

The extent to which materials appear in online catalogs varies widely by format: 85% of printed volumes, 50% of archival materials, 42% of maps, and 25% of visual materials are accessible online. Relative to ARL's 1998 data, 12% more printed volumes have an online record, as do 15% more archival materials and 6% more maps. This limited progress may be attributable in part to lack of sustainable, widely replicable methodologies to improve efficiencies.

- **ACTION:** Compile, disseminate, and adopt a slate of *replicable, sustainable methodologies* for cataloging and processing to facilitate exposure of materials that remain hidden and *stop the growth of backlogs*.
-

- **ACTION:** Develop *shared capacities to create metadata* for published materials such as maps and printed graphics for which cataloging resources appear to be scarce.
-

On the other hand, great strides have been made with archival finding aids: 52% of ARL collection guides are now accessible online, up from 16% in 1998. Across the entire population the figure is 44%, which would increase to 74% if all extant finding aids available locally were converted. The other 26% reveals the archival processing backlogs that remain.

- **ACTION:** Convert *legacy finding aids* using affordable methodologies to enable Internet access. Resist the urge to upgrade or expand the data. Develop tools to facilitate conversion from local databases.
-

Backlogs of printed volumes have decreased at more than half of institutions, while one fourth have increased. For materials in other formats, increases and decreases are roughly equal.

Archival Collections Management

The progress made in backlog reduction for archival materials is aided by the fact that 75% of respondents are using minimal-level processing techniques, either some or all of the time. Tools for creation of finding aids have not, however, been standardized; some institutions use four or more.

The institutional archives reports to the library in 87% of institutions, while two thirds have responsibility for records management (of active business records). The challenges specific to these materials should therefore be core concerns of most libraries—and it is in this context that the impact of born-digital content is currently the most pervasive.

Digitization

Nearly all respondents have completed at least one special collections digitization project and/or have an active digitization program for special collections. One fourth have no active program, and the same number can undertake projects only with special funding. More than one third state that they have done large-scale digitization of special collections, which we defined as a systematic effort to digitize complete collections—rather than being selective at the item level, as has been the norm—using production methods that are as streamlined as possible. Subsequent follow-up with respondents has revealed, however, that the quantities of material digitized and/or production levels achieved generally were not impressive or scalable.

- **ACTION:** Develop models for *large-scale digitization* of special collections, including methodologies for selection of appropriate collections, security, safe handling, sustainable metadata creation, and ambitious productivity levels.
-

One quarter of responding institutions have licensing contracts with commercial vendors to digitize materials and sell access. It would be useful to learn more about the existing corpus of digitized materials, particularly rare books, some important collections of which are not available via open-access repositories.

- **ACTION:** Determine the scope of the existing corpus of *digitized rare books*, differentiating those available as open access from those that are licensed. Identify the most important gaps and implement collaborative projects to complete the corpus.
-

Born-Digital Archival Materials

The data clearly reveal a widespread lack of basic infrastructure for collecting and managing born-digital materials: more than two thirds cited lack of funding as an impediment, while more than half noted lack of both expertise and time for planning. As a result, many institutions do not even know what they have, access and metadata are limited, only half of institutions have assigned responsibility for managing this content, few have collected more than a handful of formats, and virtually none have collected at scale. Clearly, this activity has yet to receive priority attention due to its cost and complexity. Community action could help break the logjam in several ways.

- **ACTION:** Define the characteristics of born-digital materials that warrant their management as “*special collections*.”

- **ACTION:** Define a reasonable set of *basic steps* for initiating an institutional program for responsibly managing born-digital archival materials.

- **ACTION:** Develop *use cases and cost models* for selection, management, and preservation of born-digital archival materials.

Staffing

The norm is no change in staff size except for in technology and digital services, which increased at nearly half of institutions. Even though more than 60% of respondents reported increased use of collections, staffing decreased in public services more frequently (23%) than any other area. Across the population, 9% of permanent special collections staff are likely to retire within the next five years.

The areas most often mentioned in which education or training are needed to fulfill the institution’s needs were born-digital materials (83%), information technology (65%), intellectual property (56%), and cataloging and metadata (51%).

- **ACTION:** Confirm high-priority areas in which *education and training* opportunities are not adequate for particular segments of the professional community. Exert pressure on appropriate organizations to fill the gaps.

The gradual trend in recent decades toward integration of once-separate special collections continues; 20% of respondents have done this within the past decade. Multiple units continue to exist at one in four institutions.

3

Survey of Special Collections and Archives in the United Kingdom and Ireland

Summary and Recommendations

Jackie M. Dooley
OCLC Research

Katie Sambrook
King's College London

Rachel Beckett
University of Manchester

Chris Sheppard
University of Leeds

Alison Cullingford
Bradford University

Sue Worrall
University of Birmingham

A co-publication of OCLC Research and RLUK

This summary and recommendations was originally published in February 2013 by OCLC Research at
<http://www.oclc.org/content/dam/research/publications/library/2013/2013-01-sumrecs.pdf>.

Read the complete *Survey of Special Collections and Archives in the United Kingdom and Ireland* report at
<http://www.oclc.org/content/dam/research/publications/library/2013/2013-01.pdf>.

Executive Summary

It has become widely recognised across the academic and research libraries sector that special collections and archives play a key role in differentiating each institution from its peers. In recognition of this, Research Libraries UK (RLUK) established the workstrand ‘Unique and Distinctive Collections’ in support of its strategic aims for 2011-2014. The UDC workstrand will identify ways in which special collections can ‘make the most of their potential for research, teaching and community engagement.’ This survey forms part of the overall project by gathering data to enable better understanding of the sector. It was conducted as a collaboration between RLUK and OCLC Research.

As this report reveals, we face numerous challenges if we are to maximise potential and bring special collections to the attention of those whose research or learning would benefit from their use.

A few of the most salient issues that emerged from the data:

- Alignment of special collections with institutional missions and priorities is an ongoing challenge.
- The special collections sector is undergoing a major culture shift that mandates significant retraining and careful examination of priorities.
- Philanthropic support is limited, as are librarians’ fundraising skills.
- Use of all types of material has increased across the board.
- Users expect everything in libraries and archives to be digitised; national strategies for digitisation of rare and unique materials are therefore needed.
- Many cataloguing backlogs have decreased, while some continue to grow.
- One-third of archival collections are not discoverable in online catalogues.
- Management of born-digital archival materials remains in its infancy; upper management must actively support this important work to ensure progress.

We asked respondents to name their three ‘most challenging issues.’ The following were the most frequently cited:

- Outreach (broadly defined)
- Space and facilities (particularly for collections)
- Born-digital materials
- Collection care
- Cataloguing and archival processing

One hundred twenty-two academic and research libraries with significant special collections received invitations to participate in the survey. The rate of response was 67% (82 responses), including 100% of RLUK members.

This report presents a summary and analysis of the data for all respondents, for RLUK members, and for non-RLUK respondents, with a complete set of data figures and tables for each. Also included is a comparison of the RLUK data with that of the Association of Research Libraries (US) members who responded to an OCLC Research survey of the United States and Canada (Dooley and Luce, 2010).

Key Findings

Outreach and User Services

More than half of respondents stated that use of special collections by all types of users has increased over the past decade. Few, however, were able to categorise their users by type, even those in their primary user population: 90% of users were reported as ‘other’ (i.e., type of user not identified). This could be problematic if it results in an inability to demonstrate the extent to which the primary audience is being served.

User services policies are evolving in productive ways: three-quarters of institutions permit use of digital cameras, and up to 80% allow access to printed volumes and archival materials in backlogs. On the other hand, 81% do not permit interlibrary loan, even of reproductions, which could be considered a disservice to distant researchers.

Despite these very promising data, many respondents indicated that the need to embrace new modes of outreach and service presents enormous challenges. This appears to stem from two principal factors: staff skills are being stretched by the need to undertake new duties, and, as a result, fulfillment of ‘traditional’ responsibilities is thereby rendered more difficult.

Staffing

As mentioned above, the need to undertake new duties is proving to be a major challenge. The areas most often mentioned in which education or training are needed to fulfill the institution’s needs were born-digital materials, fundraising, intellectual property, and outreach.

The data show that the mean number of permanent special collections staff across the entire population is 16.6 FTE. The median is only six, which reveals wide variation across institutions. This comparison is very different when the data are analyzed by type of institution. Forty percent of respondents have experienced an increase in the number of professionally qualified staff in recent years, while 29% had an increase in support staff. Across the population, 7% of special collections staff are likely to retire within the next five years.

A trend exists toward integration of once-separate special collections departments—more than half of respondents have done so within the past decade.

Collections

Insufficient space for collections, or inadequate space needing renovation to satisfy current needs, ranked very high among the ‘challenging issues.’ More than one-third of respondents have special collections in secondary storage. Deaccessioning of unwanted materials, some of which have not been processed many years after they were acquired, occurs for appropriate reasons but is practiced by only a few. Review of unprocessed collections for retention could be one way to contend with insufficient space.

As the size of general print collections stabilise, such as through shared print initiatives and digital publication, a need to add more

storage space for special collections would become all the more conspicuous.

One-third of respondents have undertaken one or more new collecting emphases in recent years. Although informal collaborative collecting is fairly widespread on a regional basis, formal arrangements of any kind are rare.

Born-digital Materials

The data clearly reveal a widespread lack of basic infrastructure for collecting and managing born-digital materials. Sixty percent cited lack of funding as an impediment, while only slightly fewer noted lack of both expertise and time for planning. As a result, many institutions do not even know what they have, access and metadata are limited, half of institutions have not yet assigned responsibility for managing this content, few have collected more than a handful of digital formats, and virtually none have collected at the level that is warranted.

The amount of born-digital archival material reported is minuscule relative to the extant content that warrants being preserved in archives: the mean collection size is only 2,800 gigabytes, and the median is zero. It is striking that only two institutions hold 80% of the material reported, while five hold 99%. Clearly, academic and research libraries have barely scratched the surface of the born-digital challenge.

Digitisation

Perceived pressure to digitise collections comprehensively seems to be ubiquitous. Ninety-seven percent of respondents have completed at least one special collections digitisation project and/or have an active digitisation programme that includes special collections. Progress is impeded, however, by the fact that less than half can undertake projects without special funding, while one-third have a recurring budget for digitisation.

One-third stated that they have done large-scale digitisation of special collections (defined as a systematic effort to digitise complete collections and employing production methods that are as streamlined as possible) rather than selecting and interpreting particular items.

More than 40% have licensing contracts with commercial vendors to digitise materials and sell access.

Archival Collections

While shared archival online catalogues have proven to be successful discovery hubs, only one half of archival finding aids are accessible online. This percentage would increase to 82% if all extant finding aids available only at the host institution were converted. The remaining 18% (no finding aid exists) reveals the archival processing backlogs that remain. The progress made in backlog reduction may be due, at least in part, to the use of minimal-level processing techniques by 70% of respondents.

The institutional archives reports to the library in two-thirds of institutions, while nearly half have responsibility for records management (of active business records). The challenges specific to these materials should therefore be core concerns of most libraries—and it is in this context that the impact of born-digital content is currently the most pervasive.

Cataloguing and Metadata

Backlogs of printed volumes have decreased at nearly half of institutions, while somewhat fewer backlogs have increased. For materials in other formats, increases and decreases are roughly equal. The continuing existence of backlogs may be attributable in part to the lack of sustainable, widely replicable methodologies to improve efficiencies.

The extent to which materials appear in online catalogues varies widely by format: 78% of printed volumes, 64% of archival materials, half of maps, and one-third of visual materials are accessible online.

Collection care

The preservation needs of both audiovisual and born-digital materials are well known to be huge, and our data confirm this.

The most widespread collection care problems are conservation repair of materials to enable their use and rehousing into improved boxes and other housings. Issues related to quality of storage facilities were cited by about 40%.

Metrics

A lack of established metrics placed some constraints on the data that respondents could contribute and our ability to analyse it closely. Norms for tracking and assessing user services, metadata creation, archival processing, digital production, and other activities would make it more feasible to establish reliable community norms against which to measure individual institutions.

We did not explore the particular purposes that would be served by deployment of a set of uniform metrics; it would be important to do so before undertaking such work.

Recommendations

These recommendations were formulated by the authors of this report and are wholly based in analysis of the survey data. Participants in the RLUK Unique and Distinctive Collections symposium held at the University of Aberdeen on 29 March 2012 very usefully vetted an early version, which the authors then significantly revised.

Note: This is not a set of recommendations officially endorsed by RLUK or intended for RLUK action; a forthcoming report on the UDC workstrand will fulfill that need.

In general, under each category we consider the first recommendation a higher priority than the other(s) in that group (e.g., we feel that 1.1 would potentially have a higher impact than 1.2).

1. Staffing

- 1.1. Analyse the array of duties performed by special collections staff and identify the new skills and expertise needed to move the profession forward in alignment with institutional missions.
- 1.2. Develop a plan to provide educational and development opportunities in areas, both traditional and emergent, in which skills need enhancement across the sector.

2. External Funding

- 2.1. Develop a set of arguments to assist institutions with development of external sources of funding in support of special collections.

3. User Services

- 3.1. Develop an outreach toolkit, including case studies illustrating best practices, to build skills for presentation, promotion, and engagement with special collections.
- 3.2. Develop pricing models, templates, and shared policies for user-initiated digital scanning to encourage consistency across the sector.

4. Born-digital Materials

- 4.1. Define the basic steps involved in initiating a program for managing born-digital archival materials to assist libraries that have not yet begun this work.
- 4.2. Investigate the feasibility of extending broadly across the sector the adoption of successful technical environments for managing born-digital materials that have been developed by a small number of UK institutions.

5. Digitisation

- 5.1. Develop both a national strategy for continued digitisation of special collections and a national gateway for discovery of digitised content. As part of the strategy, identify sustainable funding strategies and international partners with which to collaborate.
- 5.2. Develop cost-effective models for large-scale digitisation of special collections that take into account the special needs of these materials while also achieving high productivity.

6. Archival Collections

- 6.1. Convert print archival catalogues using affordable methodologies to enable Internet access. Develop approaches to modifying existing descriptions that strike a balance between incurring overheads and being effective for discovery. Develop tools to facilitate conversion from local databases.
- 6.2. Develop a shared understanding of the goals, characteristics, and benefits of 'simplified archival processing.'
- 6.3. Establish a methodology to assess unprocessed archival collections and develop a plan to make the national collection more fully accessible.

7. Metrics

- 7.1. Determine the potential value and uses of metrics for reporting core statistics (e.g., collection size, users, outreach efforts, catalogue records) across the sector. If warranted, define categories and methodologies and encourage their use across the sector.

8. Collection Development

- 8.1. Define key characteristics and desired outcomes of meaningful collaborative collection development, and encourage collaborations in areas of national significance.
- 8.2. Scrutinise local collecting policies to determine how well they reflect the institutional mission and can feasibly be implemented.

9. Cataloguing and Metadata

- 9.1. Collaborate to share expertise and create metadata for cartographic materials to enable improved discovery of the national collection.
- 9.2. Build on the findings of RLUK's 'hidden collections' survey of print materials to identify national cataloguing priorities.

10. Collection Care

- 10.1. Further inflect the COPAC collection management tool to meet the requirements of special collections. Investigate its potential for determining priorities for preservation and other management activities across the national print collection.
- 10.2. Take collective action to share resources for cost-effective preservation of at-risk audiovisual and born-digital archival materials.

11. Building community

- 11.1. Identify beneficial ways in which to build productive relationships across the diverse community of special collections libraries that participated in this survey.

4

Taking Stock and Making Hay: Archival Collections Assessment

Martha O'Hara Conway
Associate Director,
Special Collections Library,
University of Michigan

Merrilee Proffitt
Senior Program Officer
OCLC Research

Acknowledgments

This report would not have been possible without the support of a working group from the OCLC Research Library Partnership. Members of that group included: Martha O'Hara Conway, University of Michigan; David de Lorenzo, University of California, Berkeley; Christine Di Bella, Institute of Advanced Study; Sarah Stauderman, Smithsonian Institution. Members represented a broad range of knowledge and perspectives, and their real world, hands on experiences with collections assessment was invaluable in writing this report.

Many thanks to colleagues who read an early draft and provided helpful suggestions and recommendations: Jackie Dooley, Ricky Erway, Jennifer Schaffner (OCLC Research), Rachel Onuf (consultant) and Mark Allen Greene (American Heritage Center). Thanks are also due to John Fleckner (Smithsonian Institution), Barclay Ogden (University of California, Berkeley), Mary Ide (WGBH), Nancy Elkington (OCLC Research), and David Molke-Hansen (retired, Historical Society of Pennsylvania) for giving background interviews on their involvement with early collections assessment projects.

Introduction

Archival collections assessment is an important component of a successful collections management program. In most institutions, however, conducting an assessment is feasible only with additional resources. For this and a number of other reasons, collections assessment has not been a regular part of collections management practice. In recent years, however, a number of institutions have created or adapted collections assessment tools, employed them successfully, and made them available for use by others. The wheel has been invented.

What is Archival Collections Assessment?

In this report, the term “archival collections assessment” is used to refer to the systematic, purposeful gathering of information about archival collections. It includes collection surveys of all kinds, including those undertaken for purposes of appraisal, setting processing and other priorities, conservation decision-making, and collection management.

An accurate census of its archival collections enables the institution to act strategically in meeting user needs, allocating available resources, and securing additional funding. The systematic gathering of quantitative and qualitative data about collections makes possible the creation of adequate, consistent, collection-level descriptions; affords a better understanding of unmet preservation needs; and informs important decisions regarding collection management, processing priorities, and selection and other activities associated with digitization and exhibit preparation.

A Common Approach?

Although a number of institutions have undertaken collections assessments, a single, commonly-understood approach neither exists nor is practical. Rather than recommending a single strategy or advocating a particular approach, this report identifies and characterizes existing surveys that can be used as-is by, or serve as models for, librarians, archivists, and others who are considering collections assessment to meet one or several institutional needs. It describes the many possible components of collections assessment; emphasizes the importance of approaching collections assessment with an informed understanding of its purpose and desired outcome(s); and provides pointers to existing methodologies and tools that have been used by various institutions.

Our goal in assembling this report and making it available to the widest possible audience is to encourage the archival and special collections communities to use existing approaches in order to leverage good work, foster a growing community of practice, and encourage efficiencies for institutions both individually and collectively.

How to Use This Report

This report provides both food for thought and fuel for activity. It presents a rationale for conducting a collections assessment; describes the components of archival collections assessment; and encourages readers to consider their own needs and capacities. Additionally, we hope this report will serve to inspire and empower those who are considering collections assessment by suggesting an array of possibilities that can be readily applied to meet immediate and/or long-term needs. **Appendix A** contains pointers to a variety of exemplar projects, many of which have tools and more information available online. **Appendices B and C** are links to project documentation, which contain useful instructions and definitions.

Current Context: Tackling the Backlog Problem

It is no longer a “dirty little secret” (Tabb 2004, 123) that libraries, archives, and cultural institutions hold significant amounts of special collections material that have not been adequately described and therefore are not known, cannot be discovered, and will not be used. These uncataloged and unprocessed (i.e., “hidden”) collections have become the focus of considerable attention in recent years and efforts to address the problems they represent are numerous, varied and well documented (ARL 2008; CLIR 2011; Hewitt and Panitch 2003; Pritchard 2009; Schreyer 2007; Steele 2008).

More Product, Less Process

Cataloging and processing backlogs have long been the bane of the cultural heritage institution, and calls for addressing them have been around for almost as long as the backlogs themselves. One of the most recent of these was put forth in “More Product, Less Process: Revamping Traditional Archival Processing,” an article by Mark Greene and Dennis Meissner that gave voice to the small but growing number of archivists who have quietly abandoned traditional approaches to archival processing in favor of those that expedite user access to archival collections (2005).

In their article, Greene and Meissner issue a call for change that specifically references the successful reduction of cataloging backlogs in large research libraries through various procedural and technical innovations and by redefining quality. In redefining quality as it applies to processing, they assert that,

it must be our aim to provide sufficient physical and intellectual access to collections for research to be possible, without the necessity of processing each collection to an ideal or arbitrary standard. We should be paying more attention to achieving basic physical and intellectual control over, and thus affording research access to, all our holdings, rather than being content to process a few of them to perfection. What this means is that all collections should have collection-level intellectual control before any collection receives folder-level intellectual control. More importantly, researchers cannot come to do research if at least minimal information about the collections is not available to them. (237)

In other words, “Describe everything in general before describing anything in detail.” And make those descriptions available to the widest possible audience.

Exposing Hidden Collections

It is worth noting that the same principle was endorsed by those participating in the “Exposing Hidden Collections” conference that took place in September 2003 at the Library of Congress. A working conference planned by the Association of Research Libraries (ARL) Task Force on Special Collections, “Exposing Hidden Collections” served as a forum for interaction between various communities of professionals and set the stage for the collaborative development of an action plan aimed at surfacing “hidden” collections. One of the overriding themes of the conference was “Some access to all is preferable to no access to some.” In fact it was proposed at the outset of the two-day event that one of the outcomes of the conference should be “a pledge by participants to return to their institutions committed to providing a web-accessible collection-level record for all unprocessed materials (ARL 2009).”

This apparently has proven to be either more difficult or more problematic than the conference participants imagined. For some institutions, providing a web-accessible collection-level record for an unprocessed collection is difficult because an appropriate record neither exists nor is easily created until the collection is processed. Some institutions might be reluctant to provide a web-accessible collection-level record for an unprocessed collection because it would suggest that the collection is available for use when, for any of a variety of reasons, it might not be. Although few of the conference participants followed through on that pledge, some institutions have made descriptions of their unprocessed collections available via the web.

Cataloging Hidden Collections

A 1998 survey of ARL member institutions revealed that “significant portions” of special collections material have not been processed or cataloged and therefore are not known, cannot be discovered, and will not be used (Panitch 2001, 8). The survey results suggest that 15% of printed volumes, 27% of manuscripts, and 35% of the audio and video collections held by the 100 ARL respondents were unprocessed or uncataloged at that time. By comparison, a 2009 survey of a broader population

of North American research libraries revealed that the situation has improved only marginally over the last decade, in spite of widespread focus on the hidden collections problem. Fifteen percent of printed volumes still are not cataloged online, while for other formats, the situation may even have worsened. The survey suggests that a large percentage of materials lack online access, including 44% of archives and manuscripts, 58% of cartographic materials, and almost 25% of video and audiovisual materials. Perhaps the most sobering statistic: 71% of born digital materials held in special collections are not represented in online catalogs. (Dooley and Luce 2010, 46). This represents a staggering amount of material that is neither known by nor available to the research community.

Help Is on the Way

Heightened awareness of both the scope and the implications of the hidden collections “problem” is, fortunately, matched by a number of new and existing initiatives aimed at addressing it.

The most recent of these—launched in 2008—is the Cataloging Hidden Special Collections and Archives Program. With generous funding from The Andrew W. Mellon Foundation, the Council on Library and Information Resources (CLIR) administers this national program that awards grants in support of “innovative, efficient description of large volumes of material of high value to scholars.” Projects are evaluated and selected for funding according to the following criteria:

- potential national impact on scholarship and teaching;
- use of innovative and/or highly efficient approaches to description that could serve as models for others;
- adoption of workflow and outreach practices that maximize connections to scholarly and other user communities; and
- application of descriptive and other standards that would provide interoperability and long-term sustainability of project data in the online environment.

Over the course of its three-year history, the Cataloging Hidden Special Collections and Archives Program has awarded more than \$11.9 million to a total of 46 projects selected from approximately 300 proposals. In coordination with this program, CLIR maintains a web-accessible registry of hidden collections, based upon information supplied by applicants and others (CLIR 2011).

The National Historical Publications and Records Commission (NHPRC) has a long history of providing funds for “fundamental archival activities” in the form of basic processing projects that “reveal collections that researchers cannot easily discover” (NARA 2011). Institutions are required to

- create and share collection-level information;
- develop or implement appraisal, processing, and other techniques that will eliminate existing backlogs and/or prevent future backlogs; and
- promote the use of processed collections.

The National Endowment for the Humanities’ (NEH) Humanities Collections and Reference Resources program (NEH 2011) supports efforts that “provide an essential foundation for scholarship, education, and public programming in the humanities” by funding projects that address one or more of the following activities:

- arranging and describing archival and manuscript collections;
- cataloging collections of printed works, photographs, recorded sound, moving images, art, and material culture;
- providing conservation treatment for collections;
- digitizing collections;
- preserving and improving access to born-digital sources; and
- developing databases, virtual collections, or other electronic resources to codify information on a subject or to provide integrated access to selected humanities materials.

Practice with Purpose: Why Collections Assessment?

The first step when considering a collections assessment is a careful articulation of the reason—or reasons—for which it is to be undertaken. Because even a small survey project is very likely to be a complex undertaking, and because resource allocators are more likely to support an effort that prescribes one or more concrete outcomes, it is important to design the project in such a way that its objectives are clear, its audience is apparent, and its benefits are maximized. Depending upon its intended purpose and the resources allocated to it, a collections assessment can range from a one-time-only inventory of some or all holdings to a comprehensive, ongoing, data-gathering activity.

Most survey projects are undertaken for one or more of the four purposes described below. Although none of these precludes another, it is difficult to put equal emphasis on all of them. Early in the project, therefore, it is essential to decide which goals or outcomes are considered primary, which are considered secondary, and which will not be addressed. The assessment projects described later in this report have gathered information with the goal of accomplishing at least one of the four aims described below.

Expose Hidden Collections

Many institutions have undertaken a collections assessment for the primary purpose of preparing and sharing consistent, comparable, summary descriptions of some or all of the collections in their care. If this is indeed the primary goal, the assessment activity may consist primarily of assembling, normalizing, and/or augmenting existing descriptive information at the collection level; indicating whether or not the collection is available for research; and making this information available—preferably online. More often than not, however, creating uniform collection-level descriptions necessitates the gathering of information that can only be obtained by physically inspecting some portion of the collections and collecting information about those collections with a systematic, well-documented approach.

Some of the institutions that have undertaken collections assessments of this type have done so with the explicit intention of exchanging collection-level information with other institutions and/or depositing collection-level descriptions in a consortially—or regionally—managed database (PACSCL 2009).

Establish Processing Priorities

Especially for institutions with large backlogs of un- and under-processed collections, a collections assessment serves as a very useful tool for planning, informing, and guiding priorities for collections processing. With this purpose as its primary goal, the collections assessment becomes a more complicated undertaking, as it requires collecting information and making judgments about various aspects of the collection, only some of which may already be known or are easily determined. A collections assessment aimed at establishing processing priorities includes but goes well beyond the gathering of basic information about the size, scope, and contents of the collection. It typically requires that surveyors assess the condition of the collection material as well as the containers in which it is housed; determine the ease with which material in the collection can be located; and evaluate the ease with which the collection can be discovered, identified as relevant, and used, based upon the existence and the accessibility of catalog records, finding aids, and other collection surrogates.

A collections assessment intended to inform the assignment of priorities for processing should also include for each collection some kind of estimation of its research value for present and future users. Techniques for determining research value are described in the “Collecting Qualitative Information” section of this report.

Assess Condition

Even if establishing preservation and/or conservation priorities isn’t the primary goal of a collections assessment, it is difficult to resist the opportunity to capture information about physical condition when a collections assessment is underway. This appears to be the

case for all types of institutions and across all categories of collections. The information typically gathered ranges from a basic assessment of the overall condition of collection material and of the containers in which it is housed to a more detailed, systematic evaluation that provides the institution with a better understanding of the prevalence of specific conservation issues as well as unmet preservation challenges.

Libraries and archives have a long history of using a variety of well-established, well-documented methods to capture essential information about the current state and the ongoing needs of the collections in their care. Preservation surveys focus primarily on diagnosing large-scale and/or pervasive problems at the collection level and assessing the overall storage and housing environment, usually to make the case for facilities improvements that will slow or prevent future damage. Conservation surveys tend to highlight the scope and the distribution of problems that plague particular media (such as acetate film, brittle paper, and deteriorating magnetic tape) and support the allocation of limited resources for treatment. Increasingly, however, the consideration of preservation challenges and/or conservation issues is but one component of a larger balancing act, the overarching goal of which is to make collections accessible.

Because collections that cannot be handled physically without causing additional damage cannot be used, information about physical condition is typically used to help answer basic questions such as “How is use of this collection hampered or limited?” and “Does the degree of damage or deterioration, or the value of the collection, justify reproduction or treatment?” For many institutions, however, laying the groundwork for the establishment of preservation and/or conservation priorities is the primary goal of the collections assessment, warranting greater emphasis on the comprehensive capture and systematic tracking of essential information about condition (see, for example, Columbia University 2011).

Manage Collections

Much of the information gathered during a collections assessment can be used almost immediately to address a number of collection management issues including optimizing storage efficiencies, identifying strengths and gaps in collecting areas, and validating de-accessioning decisions. A comprehensive inventory is the foundation of effective collection management, and, when coupled with the value judgments that usually accompany a collections assessment, provides a powerful tool for repositories with burgeoning backlogs of un- and under-processed collections, significant “information gaps” regarding the contents of collections, and/or pressing space concerns. Without exception, those institutions that have undertaken collections assessment for any of the first three primary purposes described above have reaped inevitable secondary benefits in the form of better informed, more active collection management.

Ready, Set, Go! Conducting the Assessment

Several important activities must be accomplished before the survey team can get to work. These include defining the scope of the project; determining the methodology and the resources that will be employed; and documenting the policies and procedures that will govern the assessment.

Scope

Guided largely by the purpose—or purposes—of the collections assessment, scope is a fundamental consideration that must be determined at the start of the project and carefully managed throughout. Other factors that should be taken into consideration when determining the scope of the assessment include the availability of resources (human and financial), time, and physical space. For many institutions, some or all of these may be limited, and the scope of the assessment undertaking should reflect that reality.

Will all collections be surveyed? Un- or under-processed collections only? Or will other criteria determine the scope of the collections assessment? With the exception of those surveys that have been undertaken primarily to amass information about conservation issues, the scope of a collections assessment is typically limited to un- and under-processed collections only. Often there are good reasons to limit the scope of an assessment undertaking to collections consisting of or containing certain types of material or special formats (such as artworks, audio-visual material, photographs, realia, etc.). Increasingly, however, institutions are using collections assessment for purposes that require a broader scope. Examples of purposes that necessitate surveying all collections include discerning changes to collection development policies and populating a collection management database (see, for example, UCB 2011).

Collecting Information

Collections assessment is essentially an information-gathering activity. It is centered on the systematic collection of quantitative and qualitative data about various characteristics of collections, including extent and contents, condition, accessibility (physical and intellectual), and research value. Its immediate result is an array of data that makes possible the provision of adequate, consistent, collection-level descriptions; affords a better understanding of unmet preservation needs; and informs important decisions regarding collection management, processing priorities, and selection and other activities associated with digitization and exhibit preparation.

Methodology

The collections that have been identified for assessment are likely to vary considerably in many respects, including size, complexity, condition, and type of material. For each, however, the basic approach is the same: open the boxes and look at the stuff. In keeping with the stated objectives governing the assessment, surveyors will do some or all of the following for each collection:

- count and assess the condition of the containers in which collection material is housed;
- identify and assess the condition of the material(s) of which the collection consists;
- evaluate its arrangement in terms of the ease with which material can be located;
- determine the existence and the accessibility of catalog records, finding aids, and other collection surrogates; and
- assess its research value.

Clear instructions—including definitions, illustrations and examples—for all of the above are essential. The survey tool, along with accompanying forms, checklists, etc., should be thoroughly tested before actual surveying begins. Although survey data may be recorded on paper worksheets, it is typically stored in a relational database, such as FileMaker Pro or Microsoft Access, where it can be accessed and manipulated as necessary.

Staffing

Who will do the surveying? What do they need to know? Although appropriately staffing the assessment is an important consideration, and in some cases the availability of human resources may have the effect of defining or limiting the scope of the assessment, successful assessment projects have been accomplished with a variety of staffing models, ranging from those that employ experienced archivists, curators, and conservators (experts/professionals) to those that rely on individuals who have some knowledge but no experience with the collections (generalists) to those that draw on a large body of students, volunteers, and/or others who have neither knowledge of nor experience with collections (novices). In all cases,

adequate training and good documentation are key factors to a successful undertaking.

Collecting Quantitative Information

In collections assessment, quantitative methods are used to collect basic information about the extent of each collection and the types of materials of which it consists. Collecting quantitative information should be relatively easy and require little or no judgment. “How many of what?” is the basic question; because it can be asked—and answered—in a number of ways, however, it is important to consider the following:

- Will every box be opened, or is some form of sampling OK?
- How will extent be measured (items, boxes, linear feet, shelf feet)?
- How will content (in terms of types of materials, special formats, etc.) be identified and categorized (checklists, guidelines, etc.)?

Anticipating with good planning and addressing with good documentation, these and similar questions are essential.

Collecting Qualitative Information

Qualitative methods are used to collect information about the condition, physical accessibility (arrangement), intellectual accessibility (description), and research value of the collection. Collecting qualitative information usually requires making some kind of judgment in order to assign a rating (value) along a numeric or descriptive continuum (scale). In a numeric continuum, 1 is usually the lowest or worst rating and 5 is the highest or best. In a descriptive continuum, values range from, for example, “poor” to “excellent” or from “negligible” or “none” to “significant” or “very high.” Ratings can (and should) be defined and documented in such a way that they can be assigned systematically and consistently. Other possibilities for measures include estimating the amount of collection material that meets a particular criterion, for example, “What percentage of the collection needs new housing?”

Condition

Assessing the physical condition of collection material and the quality of the housing in which it is contained is often one of the most important—and most difficult—activities in collections assessment. This is especially the case for collections that contain or consist primarily of material in formats other than paper. It is helpful to keep in mind that a collections assessment is neither a preservation planning survey nor a collection condition survey. As such, effort is not usually dedicated to noting the condition of particular items or to identifying groups of materials of particular concern, although these may be called out in some way. Assessing physical condition and housing quality as one component of a larger, more general collections survey is aimed at providing a better understanding of the overall condition of collection material, the overall quality of the boxes, folders, and other containers in which it is housed, and the degree to which one or both of these might or will hinder its use.

Arrangement

Also potentially hindering the use of a collection is its physical arrangement, which is one of the reasons why collections assessment typically includes an evaluation of the ease with which material in the collection can be located. That evaluation usually takes into account both the size and the complexity of the collection, does not assume that arrangement to the item level is necessary or desirable, and is focused on rating the collection in terms of how successfully it can be used for research. A small, relatively homogeneous collection in rough order, for example, is generally more physically accessible than a large, heterogeneous collection in rough order, and the ratings should reflect that.

Description

Before a collection can be used, however, it must be discovered and identified as relevant. Both of these depend upon the existence and the accessibility of catalog records, finding aids, and other collection surrogates. Rating the “intellectual accessibility” of a collection, then, typically requires determining if and how well the collection is described (in an accession report, catalog record, finding aid, etc.) and evaluating the accessibility—especially the

online availability—of any existing descriptions. While a catalog record and/or a simple inventory might provide adequate access to a small, relatively straightforward collection, neither is sufficient for a large or complex collection. Finding aids typically provide the best intellectual access to archival collections, especially large and/or complex collections, and the ratings that are assigned during this component of a collections assessment are governed by that assumption. The ratings also reflect the expectation that a collection is only truly accessible when a researcher can find information about it online. The best rating, therefore, is reserved for those collections that are described online; the worst is assigned to those that are not described at all or are described only in an accession record, donor/control file, or other document that is inaccessible to researchers.

Research Value

Assessing research value is probably the most troublesome component of collections assessment for a number of reasons, most of which can be attributed to the difficulty—real or perceived—attendant in defining and measuring “research value” in the first place. Assuming, however, that it can be defined and measured, a thoughtful assessment of research value usually provides compelling information that can be used to inform important decisions regarding collection management, processing priorities, and selection and other activities associated with digitization and exhibit preparation.

In the collections assessment context, the term “research value” usually refers to the value of the collection in terms of the extent to which it includes relatively rare, extensive, and/or detailed information about a topic that has received considerable prior attention, is gaining currency, and/or has apparent potential to attract significant interest. It is frequently expressed as a composite of two measures: documentation interest and documentation quality.

The documentation interest rating provides an indication of the value of the collection in terms of its topical significance, with values ranging from 1 (negligible or none) to 5 (very high). Similarly, the documentation quality rating provides an indication of the value of the collection in terms of

its topical richness, again with values ranging from 1 (slight) to 5 (very rich). Research value can then be expressed as the sum of the documentation interest rating and the documentation quality rating.

Here it must be pointed out that many collections have values in addition to, or even other than, research value, and that these can—and should—be measured if the overall purpose of the assessment warrants such an evaluation. Examples of these “other” values include intrinsic value and local or institutional value.

Collecting “Other” Information

In addition to collecting the above-described information about the collection, many

institutions also collect information about the assessment process itself, including who conducted the assessment, when it was accomplished, how long it took and if any activity (such as reboxing) was undertaken.

Putting it All Together

Of course, the reason for collecting assessment data is to put it to use. Here are some examples of collection assessment in action.

The following example (figures 1 and 2) from the University of Michigan Special Collections Library shows that while the collection rates fairly high in terms of research value, it is intellectually inaccessible and physically difficult to use.

<u>COLLECTION</u>	
Creator:	Ann Arbor Tenants' Union
Title:	Ann Arbor Tenants' Union Records
Type:	records
Dates (Inclusive):	1969-1995
Dates (Bulk):	1969-1980
Language (Primary):	English
Language (Secondary):	Spanish
Extent: manuscript boxes	01
+ records center boxes (filed 1980)	

Figure 1. University of Michigan Special Collections Library collection sample

<u>RATINGS</u>	
Housing Quality	3
Physical Condition	3
Physical Access	2
Intellectual access	1
Documentation Interest	4
Documentation Quality	3
Research Value (calculated)	7
Exhibition Value (yes/no)	yes
+ (Rating):	

Figure 2. University of Michigan Special Collections Library ratings sample

The Smithsonian Archives uses collections assessment to measure and demonstrate change over time in their preservation module. Represented in figure 4, the initial assessment of this collection in 2001 (Accession 000182 United States Civil Service Commission) shows an Overall Priority score of 2.

Title: Correspondence	Location: NUS	Priority Score: <input type="text" value="2"/>
CONDITION ASSESSMENT		
This collection has been assessed for condition		
HOLDINGS MAINTENANCE		
Housing Needs • 75% needs to be placed on shelves or in boxes, folders, or drawers	3	Assessment Key: 0 = 0% 1 = <25% 2 = 25-75% 3 = >75%
Housing Materials • 75% are in acidic boxes, or have acidic containers or folders	3	Holdings Maintenance Score: <input type="text" value="6"/>
Positioning of Records 0% are in boxes that are too full or too loose, or file folders that are messy or too full.	0	Holdings Maintenance Grade: <input type="text" value="B"/>
Difficult Formats Sizes 0% are the wrong size or format for its current housing	0	
Damaging Attachments 0% has inappropriate attachments like staples, paperclips, and poor quality adhesives	0	
Key to Holdings Maintenance Grades 10-15 = A 5-9 = B 0-4 = C		
PHYSICAL CONDITION		
Physical Damage • 75% has damage like dirt, staining, losses, embrittlement	3	Collection Grade: <input type="text" value="A"/>
Unstable Materials 25-75% of the collection is inherently unstable, like newsprint, magnetic media, colored photographs	2	
ALERTS		
Active Mold This collection has mold.		Comments <input type="text"/>
Treated Mold This collection has been treated for mold.		

Figure 3. Smithsonian Archives preservation model: initial assessment of United States Civil Service Commission 2001 collection (Accession 000182)

In 2003, the collection was re-assessed after preservation actions were taken (shown in figure 4). Some items were discovered to be rolled while reboxing the collection and a score of 2 was given to the Difficult Formats/Sizes category. While the overall Priority Score for this collection did not change at this time, the Holdings Maintenance Score went from a 6 to a 2.

Location:	NUS
CONDITION ASSESSMENT	
This collection has been assessed for condition	
HOLDINGS MAINTENANCE	
Housing Needs 0% needs to be placed on shelves or in boxes, folders, or drawers	0
Housing Materials 0% are in acidic boxes, or have acidic containers or folders	0
Positioning of Records 0% are in boxes that are too full or too loose, or file folders that are messy or too full.	0
Difficult Formats Sizes 25-75% are the wrong size or format for its current housing	2
Damaging Attachments 0% has inappropriate attachments like staples, paperclips, and poor quality adhesives	0
Assessment Key:	
0 = 0% 1 = <25% 2 = 25-75% 3 = >75%	
Holdings Maintenance Score:	
2	
Holdings Maintenance Grade:	
C	
Key to Holdings Maintenance Grades	
10-15 = A 5-9 = B 0-4 = C	
PHYSICAL CONDITION	
Physical Damage • 75% has damage like dirt, staining, losses, brittleness	3
Unstable Materials 25-75% of the collection is inherently unstable, like newsprint, magnetic media, colored photographs	2
Collection Grade:	
A	
ALERTS	
Active Mold This collection has mold.	
Treated Mold This collection has been treated for mold.	
Pest Infestation This collection has a pest infestation.	
Treated Pests This collection has been treated for pest infestation.	
Comments	

Figure 4. Smithsonian Archives preservation module: reassessment of United States Civil Service Commission 2001 collection (Accession 000182)

Finally, the entire collection was assessed before moving the collection offsite in 2006. This generated a new overall Priority Score of 5 (shown in figure 5).

Location:	NUS	Priority Score:	5
CONDITION ASSESSMENT			
This collection has been assessed for condition			
HOLDINGS MAINTENANCE			
Housing Needs 0% needs to be placed on shelves or in boxes, folders, or drawers	0	Assessment Key:	
Housing Materials 0% are in acidic boxes, or have acidic containers or folders	0	0 = 0% 1 = <25% 2 = 25-75% 3 = >75%	
Positioning of Records 0% are in boxes that are too full or too loose, or file folders that are messy or too full.	0	Holdings Maintenance Score:	0
Difficult Formats Sizes 25-75% are the wrong size or format for its current housing	0	Holdings Maintenance Grade:	C
Damaging Attachments 0% has inappropriate attachments like staples, paperclips, and poor quality adhesives	0	Key to Holdings Maintenance Grades	
10-15 = A 5-9 = B 0-4 = C			
PHYSICAL CONDITION			
Physical Damage -25% has damage like dirt, staining, losses, brittleness	1	Collection Grade:	B
Unstable Materials • 75% of the collection is inherently unstable, like newsprint, magnetic media, colored photographs	3	Comments	
ALERTS			
Active Mold This collection has mold.			
Treated Mold This collection has been treated for mold.			
Pest Infestation This collection has a pest infestation.			
Treated Pests This collection has been treated for pest infestation.			

Figure 5. Smithsonian Archives preservation module: final assessment (pre-offsite move) of United States Civil Service Commission 2001 collection (Accession 000182)

The Logjam Project (Northwest Archives Council, UK) provided an “audit toolkit” as a data collection tool (shown in figure 6). The toolkit provides a calculation that generates an estimate of Cataloging Resources for collections, taking into account such factors as the Extent of the collection, main Covering Dates, Level of Cataloging necessary and the potential Complexity of cataloging the collection*. Each data field is weighted and a Resources Score is generated, producing a cataloging estimate for each collection, with times given for both professionals and paraprofessionals. Calculations are based on regional norms, but could be adjusted for other circumstances.

Collection: Archer Family, Earls of Borsetshire **Covering Dates:** 18th-19th century
Cataloguing Complexity: Complex
Level of Cataloguing: High - uncatalogued
Extent: 40 linear metres
Resources Score: 290
Cataloguing Resources: 9-12 months archivist time and 2 months paraprofessional time

Figure 6. Logjam audit kit sample from Northwest Archives Council, UK

**Logjam gives definitions for values for “level of cataloguing” (High—Uncatalogued; Medium—Box-listed; Low—Listed to series level; and “cataloguing complexity” (Very Complex, Complex, Moderate, Moderate Straightforward, Straightforward).*

What's Missing from This Picture?

By providing both the opportunity and a process for documenting a wide range of characteristics about the collections in our care, archival collections assessment can be used to address a variety of important needs, including collection management issues and processing priority-setting. Existing practices, however, do not fully support other equally-pressing concerns.

Researcher Needs

Much of the focus of archival collections assessment is oriented to the needs of the collections themselves. Which require rehousing? Need basic conservation? Lack adequate description? The ever-increasing emphasis in libraries and archives on meeting the needs of researchers—for whom we have collections in the first place—will likely result in less support for assessment activities that do not include the identification of collections that are expected to be of high research interest.

Assigning a research value rating as a component of archival collections assessment is one way to estimate potential scholarly significance; another might be mining use and other data to determine how heavily used a particular collection, or group of collections, is, especially in relationship to other collections. Although recording “amount and type of use” is not typically integrated into collections assessment activities, and would in fact require data external to the survey process per se, it should feature more prominently in our user-centric world.

Collection Development Policies

The 2009 survey of special collections and archives in North American academic and research libraries paints a picture that is both encouraging and sobering. Among the key findings described in the report are the following:

- monetary resources are shrinking;
- collections, and user demand for access to them, are growing; and
- space for collections is inadequate.

This “trifecta” of sorts serves to remind us of the importance of acquiring and devoting resources to the needs of those collections that are most

valuable from a research perspective and that fit best within existing collection strengths. The fact that few institutions are likely to secure additional storage capacity, and even fewer are likely to stop collecting, underscores the importance of collecting policies.

Only half of the respondents in the 1998 survey of special collections in ARL libraries indicated that they have formal collection development policies. By revealing existing collection strengths, collections assessment can serve as a powerful motivator for those institutions that need to develop and/or refine meaningful collecting policies. Collections assessment data can also be used, when necessary, to make a case for deaccessioning “out of scope” and “not a good fit” collections.

Digitization Readiness

In a world that is increasingly shaped by the view that “if it isn’t online it doesn’t exist,” digitization of special collections material is—or should be—at or near the top of our priority list. Although some of the data gathered during the course of a “typical” collections assessment contributes significant value to the selection for digitization process, other important data is not usually collected. By anticipating the need to answer questions about copyrights, access and/or use restrictions, and the extent to which a particular collection (or related collections) has already been digitized, archival collections assessment can play a critical role in helping us move forward in this important arena.

Conclusion

The combination of almost limitless collecting opportunities and increasingly limited resources with which to get the job done requires that we identify, articulate and focus our attention on the priorities that are most central to our mission. Whether undertaken as a one-time, for-one-purpose-only project or integrated into an overall approach to managing collections, archival collections assessment can help us set those priorities by taking much of the guesswork out of the picture.

Appendix A: Project Descriptions

The Black Metropolis Research Consortium

Consortial Survey Initiative of African American Materials (January 2009-December 2011)

<http://lucian.uchicago.edu/bmrcsurvey/>

With funding provided by The Andrew W. Mellon Foundation, the Chicago-based Black Metropolis Research Consortium (BMRC) is undertaking a comprehensive survey of collections of materials that document African American and African diasporic culture, history, and politics held by its 14 member institutions and by 20 community-based African American organizations and creators. The goals of the survey are several and include making possible the creation of preliminary descriptions of collections that are inaccessible to researchers; informing the prioritization of preservation and access needs; and enabling collaboration, building partnerships, and sharing “best practices” between and among initiative participants. The project website includes links to survey documentation, status reports, and the Second Space Initiative, which facilitates access to relevant research material held outside the library, museum, and archival communities.

Canadian Museum of Nature

Assessing and Managing Risks to Your Collections

<http://nature.ca/en/research-collections/our-scientific-services/assessing-managing-risks-your-collections>

The Canadian Museum of Nature regularly offers workshops on identifying, ranking, and mitigating risks to collections of cultural property. Based upon the Cultural Property Risk Analysis Model developed by the Canadian Museum of Nature, the workshop provides participants with a methodological approach to identifying types of risk, calculating magnitudes of risk, determining methods for controlling risks, and evaluating mitigation

strategies. Participants receive a manual and a Risk Assessment Worksheet (in Excel) designed for use in a comprehensive collection risk assessment.

Chicago History Museum

Manuscripts Cataloging, Survey, and Processing Project (October 2009-March 2010)

For more information, e-mail Peter Alter (alter@chicagohistory.org)

With funding from the National Historical Publications and Records Commission (NHPRC), the Chicago History Museum conducted a cataloging and assessment survey of the Museum's archival and manuscripts holdings. Informed by projects at the Historical Society of Pennsylvania and the Philadelphia Area Consortium of Special Collections Libraries (PACSCL), the survey methodology was modified to support the Museum's MPLP-based, tiered processing approach. The purpose of the survey was to (1) create and/or verify and enrich catalog records for all collections of half a linear foot or larger; (2) assess holdings to determine each collection's ideal minimal processing level (i.e., collection, series, sub-series, or folder); (3) identify un- and under-processed collections (4); prioritize collections for processing and (5) flag “found in collection” material (unaccessioned collections and/or collections with inadequate accession documentation). Project staff verified and enhanced more than 1,000 catalog records describing the Museum's manuscript holdings, created approximately 30 new catalog records, and generated a non-public database to guide the planning and management of future preservation and processing activity. More than 100 collections (totaling nearly 1,300 linear feet) were processed to the series-level in a later phase of the project, and approximately 300 “found in collection” problems were resolved through legal and/or administrative measures.

Columbia University

Mellon Special Collections Materials Survey (2003-2004)

[http://www.columbia.edu/cu/lweb/
services/preservation/surveyTools.html](http://www.columbia.edu/cu/lweb/services/preservation/surveyTools.html)

Between October 2003 and July 2004, staff at Columbia University Libraries surveyed unprocessed collections held in the Rare Book and Manuscript Library, the Avery Architectural and Fine Art Library, and the C.V. Starr East Asian Library. In total, 1,588 survey hours were spent entering data on 569 collections and accounting for 26,299 units stretching 15,867 linear feet. These collections are composed of 8,703 feet of loose paper; 87,948 bound volumes of all types; 100,903 architectural drawings; 14,218 graphic works; 158,478 photographic materials; 136,457 negatives, slides, motion pictures and microfilm; 1,288 phonographs; 6,559 audiotapes, videotapes and computer media; 277 optical media items, and nearly 3,400 pieces of realia and memorabilia.

The project website includes a guide to the survey instrument/database and a description of the ratings, both of which are modeled on, but vary from those developed at the Historical Society of Pennsylvania. Staff in the Rare Book and Manuscript Library are using the database to track accessions and as a source for box lists and other forms of preliminary and intermediate access tools; Preservation Department staff rely on it for preservation and conservation planning and for establishing departmental goals and priorities.

Historical Society of Pennsylvania

Mellon Collections Preservation and Backlog Processing Planning Project (2000-2002)

[http://www2.hsp.org/collections/
manuscripts/Mellon/about.html](http://www2.hsp.org/collections/manuscripts/Mellon/about.html)

This comprehensive survey appears to be the first in a series of Mellon-funded projects aimed at collecting qualitative and quantitative data about unprocessed special collections material. The model developed in this project includes measures of the following for each collection surveyed: physical condition, quality of housing, physical access (arrangement), intellectual access (description), and research value (interest and quality of documentation). Between 2000 and 2002, project staff surveyed approximately 5,000 collections, including 3,000 manuscript collections; 300,000 maps, prints, drawings, broadsides, and photographs; and approximately 11,000 art objects and artifacts.

North West Regional Archive Council (UK)

Logjam: An Audit of Uncatalogued Collections in the North West

<http://www.northwestcultureobservatory.co.uk/>

[You must create a free account. Once you are logged in, search for “Logjam” in the databank to download documentation.]

Taking the form of a detailed audit, the Logjam project was designed to “scope the size and type of uncatalogued collections held in 30 of the region’s principle archive-holding institutions.” The work was undertaken by the North West Museums, Libraries and Archives Council (NWMLAC) on behalf of North West Regional Archive Council (NWRAC). The project represents one component of a strategy aimed at improving and expanding access to the region’s archives by making finding aids and collections more widely available and by developing a collaborative approach to cataloging backlogs. Specific goals of the project include (1) producing a detailed picture

of the uncataloged archival collections held in each repository and in the region as a whole; (2) describing the resources required to catalog these collections; (3) assigning priorities for cataloging these collections and (4) identifying priorities for future collaborative projects within the region.

Philadelphia Area Consortium of Special Collections Libraries (PACSCL)

Consortial Survey Initiative (2006-2008)

<http://www.pacsclsurvey.org/>

The Philadelphia Area Consortium of Special Collections Libraries (PACSCL) Consortial Survey Initiative is a 30-month project funded by The Andrew W. Mellon Foundation to assess unprocessed, underprocessed, and underdescribed archival collections in a range of physical formats held in 22 Philadelphia area institutions. Modeled on the Historical Society of Pennsylvania project, the purpose of the survey is to collect data that can be used to (1) inform planning for, and prioritization of, collections work within individual institutions and across the consortium and (2) improve intellectual access to unprocessed and underprocessed collections by making collection-level records available to the public. As of the end of October, 2,100 collections totaling over 19,400 linear feet in 22 institutions have been surveyed.

Survey data is recorded in a shared, publicly-accessible database developed specifically for the project. Because it includes fields that allow institutions to maintain internally significant data, such as location and provenance information, the database can be used as a basic accessions or collection management system by individual institutions. The website includes links to project documentation including a survey checklist, a description of the ratings, and a guide to the database.

Smithsonian Institution, National Museum of Natural History

Angels Project (1996)

<http://cool.conservation-us.org/coolaic/sg/bpg/annual/v15/bp15-18.html>

In conjunction with the annual meeting of the American Institute for Conservation, an Angels Project connects conservators with a collection that needs care. The project described in this report served as a pilot to develop and demonstrate “ideal” procedures for the processing, rehousing, and reformatting of an important collection of scientific illustrations.

Smithsonian Institution Archives

Preservation Assessment Component of Collection Management System

For more information, e-mail Sarah Stauderman (staudermans@si.edu)

The Preservation Assessment Component provides a mechanism for the capture and tracking of essential information about the condition of collections. Seven questions guide the assessment; answers (provided on a scale from 0 to 3) are used to automatically calculate and assign preservation priority. Assesses the percentage of the collection that needs housing; has inappropriate housing material (e.g., acidic folders, envelopes); is poorly positioned (e.g., messy, overstuffed); has format problems (e.g., crushed, folded, rolled); has damaging and/or inappropriate attachments (e.g., staples, paper clips, etc.); has physical damage (from dirt, adhesive, water, etc.); and has unstable materials (e.g., newspaper, thermo fax paper, color photographs, etc.). Also provides mechanisms for recording actions taken during the assessment, including rehousing and digitization, and for alerting preservation staff to immediate and/or long-term needs that cannot be addressed during the assessment.

University of California, Berkeley, The Bancroft Library

*Manuscripts Survey Project
(February 2008-January 2011)*

<http://blogs.lib.berkeley.edu/bancsurvey.php>

With funds provided by The Andrew W. Mellon Foundation and the Rosalinde and Arthur Gilbert Foundation, the Bancroft Library is undertaking a comprehensive survey of all manuscript holdings processed before 1996, including a backlog of some 595 collections representing 25,000 linear feet of archival and manuscript material that is currently unavailable for research. Project staff, working over a three-year period, will apply standard archival appraisal methodologies to each collection in order to determine its scope and content, identify preservation needs, make recommendations regarding arrangement and description, and estimate the resources required to make it fully accessible to researchers. The survey will yield updated, accurate, and detailed information that will be used to establish processing goals, develop funding priorities, and facilitate collection management decisions, including those involving the de-accessioning of out-of-scope materials.

University of Michigan

Unprocessed Collections Survey Project (2009)

For more information, e-mail Martha Conway (moconway@umich.edu)

This project engaged masters-degree students at the School of Information in two consecutive projects surveying un- and under-processed collections held by the Special Collections Library. Working in teams of two and three, using an assessment methodology derived from the Historical Society of Pennsylvania project, 55 students collected quantitative and qualitative information on a total of 40 unprocessed collections of archival and manuscript material. Their findings, and the reports documenting their effort and their observations, have been used to populate a web-accessible database that the Special Collections Library will employ to create and make available adequate and uniform collection-level descriptions; understand more fully the prevalence of unmet preservation challenges; inform collection management decisions; and establish and guide processing priorities. Project documentation includes a field-by-field description of the database tables and an illustrated procedure manual.

University of Virginia

Andrew W. Mellon Special Collections Assessment Project (2002-2004)

[https://www.lib.virginia.edu/
small/collections/mellon/](https://www.lib.virginia.edu/small/collections/mellon/)

Modeled on the Historical Society of Pennsylvania project, this survey of the archival and manuscript holdings in the Special Collections Library resulted in data that has been used to determine cataloging and processing priorities and to generate time and cost estimates for the work associated with collections that require additional processing. Project staff collected several types of use data to measure current interest in the holdings of the Special Collections Library, developed a methodology to identify current and future research trends that those collections might support, and evaluated the ease with which staff can locate and serve collection materials to patrons and the ability for patrons to identify relevant materials in those collections. The survey procedure manual and the data collection form are available at the project website.

State University Libraries assessed the physical condition of processed manuscript and photograph collections held by Manuscripts, Archives, and Special Collections.

Staff surveyed approximately 4,400 linear feet of manuscripts and 120 collections containing more than 500,000 photographic images. The project website includes links to survey forms, sample database records, and a photo gallery.

WGBH Media Library and Archives

Assessment for Scholarly Use

[http://openvault.wgbh.org/pdf/
WGBHMLAAssessment.pdf](http://openvault.wgbh.org/pdf/WGBHMLAAssessment.pdf)

The WGBH Media Library and Archives (MLA) Assessment for Scholarly Use project, funded by The Andrew W. Mellon Foundation, was designed to achieve two goals: to determine the educational value of WGBH's extensive archival collection for higher education research and instruction, and to accomplish this by designing an assessment instrument for surveying audio-visual collections that could be shared with other institutions. At the start of the project, the MLA housed approximately 29,000 programs with 570,000 related production elements and documents, numbers that suggest both the potential worth of this collection to the academic community and the complexity of evaluating its educational value. The study approached this challenge by (1) creating a framework and tool for collecting information about the archived programs (2) assembling a detailed composite portrait of the archival collection and (3) modeling potential approaches to analyzing and employing the data compiled through this work. The extensive project report includes the assessment tool and recommendations regarding its use by other institutions.

Washington State University Libraries

Comprehensive Preservation Survey of Manuscript and Historical Photograph Collections (2004-2005)

[http://www.wsulibs.wsu.edu/holland/
masc/preservationsurvey.html](http://www.wsulibs.wsu.edu/holland/masc/preservationsurvey.html)

With a grant from the Washington Preservation Initiative, a LSTA-funded program administered by the Washington State Library, the Washington

Appendix B: Procedure Manual (University of Michigan)

Available online: <http://www.oclc.org/research/activities/backlogtools/michiganmanual.pdf>

Appendix C: Ratings Descriptions (Columbia University)

Available online: <http://www.oclc.org/research/activities/backlogtools/columbiaratings.pdf>

References

- ARL (Association of Research Libraries). 2008. Special Collections Task Force Final Status Report, 2006. Last modified 4 April. <http://www.arl.org/rtl/speccoll/spcolltf/status0706.shtml>
- . 2009. "Transforming Research Libraries. The unique Role of Special Collections. Exposing Hidden Collections: 2003 Conference Summary. Hidden Collections Conferees Call for Local and Collective Actions" Last modified 6 April. http://www.arl.org/rtl/speccoll/EHC_conference_summary.shtml.
- CLIR (Council on Library and Information Resources). 2011. "Cataloging Hidden Special Collections and Archives." <http://www.clir.org/hiddencollections/index.html>.
- Columbia University. 2011. "Columbia University Libraries Preservation and Digital Conversion Division. Special Collections Materials Survey Instrument." http://library.columbia.edu/services/preservation/survey_tools.html
- Dooley, Jackie M., and Katherine Luce. 2010. "Taking Our Pulse: The OCLC Research Survey of Special Collections and Archives." Dublin, Ohio: OCLC Research. <http://www.oclc.org/research/publications/library/2010/2010-11.pdf>
- Greene, Mark A. and Dennis Meissner. 2005. "More Product, Less Process: Revamping Traditional Archival Processing." *The American Archivist*. 68 (2): 208–263. <http://archivists.metapress.com/content/c741823776k65863/fulltext.pdf>
- Hewitt, Joe A. and Judith M. Panitch. 2003. "The ARL Special Collections Initiative." *Library Trends* 52 (1):157–171. http://www.ideals.illinois.edu/bitstream/handle/2142/8497/librarytrends52i1p_opt.pdf?sequence=1.
- NARA (National Archives and Records Administration). 2011. "National Historical Publications and Records Commission (NHPRC)." FY 2011 Grant Announcement: (initial). Archives—Basic Projects. <http://www.archives.gov/nhprc/announcement/basic.html>.
- NEH (National Endowment for the Humanities). 2011. "Humanities Collections and Reference Resources". Posted 6 April. <http://www.neh.gov/grants/guidelines/HCRR.html>.
- PACSCL (Philadelphia Area Consortium of Special Collections Libraries). 2009. "PACSCL Consortial Survey Initiative." Last modified 15 March. <http://www.pacsclsurvey.org>.
- Panitch, Judith M. 2001. "Special Collections in ARL Libraries: Results of the 1998 Survey Sponsored by the ARL Research Collections Committee." Washington, DC: Association of Research Libraries. http://www.arl.org/bm~doc/spec_colls_in_arl.pdf
- Pritchard, Sarah M. 2009. "Special Collections Surge to the Fore." *Libraries and the Academy*. 9(2): 177–180.
- Schreyer, Alice. 2007. "University of Chicago Explores Library-Faculty Partnerships in Uncovering Hidden Collections." *ARL: A Bimonthly Report* 251 (April): 4. <http://www.arl.org/bm~doc/arlbr251.pdf>
- Steele, Victoria. 2008. "Exposing Hidden Collections: The UCLA Experience." *C&RL News* 69 (6). <http://www.ala.org/ala/mgrps/divs/acrl/publications/crlnews/2008/jun/hiddencollections.cfm>.
- Tabb, Winston. 2004. "'Wherfore Are These Things Hid?' A Report of a Survey Undertaken by the ARL Special Collections Task Force" *RBM: A Journal of Rare Books, Manuscripts, and Cultural Heritage* 5 (2): 123-126. <http://rbm.acrl.org/content/5/2/123.full.pdf>.
- UCB (University of California, Berkeley). 2011. "The Bancroft Library Manuscripts Survey Project" Last modified 18 February. http://bancroft.berkeley.edu/info/mss_survey/index.html.

5

Over, Under, Around, and Through: Getting Around Barriers to EAD Implementation

Michele Combs
Syracuse University

Mark A. Matienzo
Yale University

Merrilee Proffitt
OCLC Research

Lisa Spiro
Rice University

Introduction

This report frames some of the obstacles that archivists have experienced adopting Encoded Archival Description (EAD). It also suggests pathways to help you get out of the ruts, around the roadblocks, and on the road to success. This report is addressed to those who have a basic understanding of standard archival descriptive structures and modest acquaintance with EAD. Our objective is to help you communicate EAD's value as a key element of successful archival information systems and overcome potential barriers to its implementation. This paper does not contain an EAD primer, or cover the basics of document encoding. For those who are not familiar with EAD, we recommend the EAD Help Pages as an excellent starting place for more information.¹

Archivists have been encoding finding aids using EAD for over a decade. An impressive number of institutions have implemented EAD, but many have not. A 2008 survey revealed that nearly half of respondents (79 out of 168) had not yet implemented EAD.² A further analysis of the characteristics of those who had not yet adopted EAD reveals that all types of institutions are represented, including archives affiliated with large and small universities and those with a range of information technology (IT) services, from no professional IT staff to those with access to the services of a large IT department.

Our professional literature articulates obstacles ranging from political to technical, and much in between. Over the last ten years a growing body of relevant articles detail barriers: Jill Tatem's article "EAD: Obstacles to Implementation, Opportunities for Understanding"; James M. Roth's "Serving up EAD: An Exploratory Study on the Deployment and Utilization of Encoded Archival Description Finding Aids"; and Elizabeth H. Dow's "EAD and the Small Repository."³ These early works were followed by Katherine M. Wisser's EAD Tools Survey and Sonia Yaco's article, "It's Complicated: Barriers to EAD Implementation."⁴

Political or logistical issues may keep you from getting going; technical issues may get you bogged down along the way. Against this backdrop of challenges, there are a growing number of

tools that support EAD.⁵ Nevertheless, real and perceived barriers to EAD implementation still exist, all of them well documented. For every roadblock, as Sesame Street's Grover says, there is a way "over, under, around, and through."⁶ This paper presents useful tools—informational, persuasive, or technical—for overcoming barriers you may encounter in your journey towards EAD implementation.

Section I of this report addresses political and logistical issues. These include gaining buy-in from institutional decision makers, overcoming the urge to achieve perfection, finding ways to maximize scarce resources, and getting over the initial humps of dealing with a relatively complicated standard and what can be perceived as overwhelming logistical issues.

Section II navigates technical problems and solutions, such as thinking about lossless data streams in conversion and management, selecting software (and challenges around open source software in particular), publishing, and mitigating the complexity of the standard.

Members of this working group (under the auspices of the RLG Partnership and OCLC Research) authored this report jointly. We all have had experience with EAD and have struggled with the range of issues. Thus, the advice we offer comes from practical experience.

This paper addresses a wide range of needs because of the assortment of issues. We hope that you will dip directly into the sections that are most appropriate to your particular need. We present barriers as articulated in published literature. We then propose one or more solutions that may work for you. Our goal is to show you that implementing EAD is easier than you think. We hope these strategies will be helpful and will smooth the way to successful implementation.

Section I: Political and Logistical Issues

I'm preaching to the unconverted⁷

 **Solution:** Prepare effective arguments about EAD's significance

The following arguments may help you communicate that EAD is a good investment of institutional funds and staff resources. We begin with a brief “elevator speech” to introduce the nature and purpose of EAD, followed by more specific points.

The elevator speech—What is EAD, and why should my institution invest in it?

EAD is an international standard for encoding finding aids established to meet the needs of both end-users and archivists. EAD is represented in XML (Extensible Markup Language), a platform-neutral data format that ensures data longevity when migrated from one software environment to another. EAD ensures the long-term viability of your data by encoding intellectual rather than only presentational data (HTML, for example, only accomplishes the latter). EAD can be produced from (or mapped to) a variety of formats, including relational databases, MARC, Dublin Core, HTML and others, which makes it an excellent format for porting data. In addition researchers can have a more robust interaction with EAD finding aids because EAD enables better searching and subsequent delivery from a single source document.

...and more!

Pick and choose from among the following ideas that will be the most persuasive in your circumstances.

EAD is an internationally-used encoding standard

EAD complies with data content standards such as ISAD-G (the General International Standard Archival Description, developed by the International Council of Archives) and DACS (*Describing Archives: A Content Standard, developed by an international working group under the auspices of the Society of*

American Archivists).⁸ EAD is global; EAD has been implemented by a wide variety of institutions, not only in the US and Canada, but also throughout Europe, Australia, New Zealand and Asia.

EAD plays well with others

EAD has been mapped to and from other data encoding standards such as MARC and Dublin Core.⁹ Because EAD supports hierarchical description, you can map data from a relational database; many commonly-used EAD tools are, in fact, built on relational databases. EAD need not be the environment in which you produce, store and manage your description, but it works well as a global transfer syntax.

EAD encoding facilitates aggregation

It would be difficult, if not impossible, to create effective subject gateways like the American Institute of Physics’ Physics History Finding Aids Web site, or regional collection gateways such as the Online Archive of California, without the consistency imposed by EAD.

An abundance of tools support EAD implementation

Tools exist to facilitate every aspect of EAD use, from encoding to publication. So many tools exist that we’ve included only a selection in Appendix II. An even wider variety of tools are covered in Archival Management Software: A Report for the Council on Library and Information Resources (2009).¹⁰

EAD implementation is supported by significant opportunities for training and collaboration

Opportunities abound for formal and informal EAD training, advice and consultation to support the growing population of EAD implementers. Some examples include the EAD discussion list, courses offered by the Society of American Archivists and Rare Book School, and workshops at local, regional and national conferences.¹¹ Various state and regional consortia offer EAD training opportunities, tools, and guidelines.

EAD is good for researchers

...in a number of ways:

- 1) Researchers can discover collections in more places through wider availability. EAD's consistent coding and structure means it's easy to submit your finding aids to multiple access points (to the Online Archive of California, or to OCLC's ArchiveGrid, or to a subject-based portal such as the one maintained by the Niels Bohr Library & Archives at the American Institute of Physics, for example) so they're more likely to be found by researchers.
- 2) Inexperienced researchers can use finding aids more easily. Consistency of content and presentation eases the use of collection descriptions for inexperienced researchers. Finding aids that are exposed online are far more likely to be found by inexperienced researchers—an audience whose needs we must always bear in mind—than collection descriptions that are only available locally.¹² As user studies reveal better and more intuitive ways to present finding aid content, reformatting collection guides encoded in EAD is painless. If one presentation/display method proves problematic or confusing for researchers, you can change it with minimal time and effort and zero rekeying or editing.
- 3) Researchers can filter and refine searches. Some applications can utilize EAD's structured tags. This makes it possible to limit searches to scope and content notes or collection titles, for example.
- 4) Display and output can be tailored for research needs. One single EAD encoded file can provide multiple output versions for multiple researcher needs (online version, printer-friendly version, etc.). You can also easily create different display options for different audiences.
- 5) Researchers can explore old data in new ways. EAD enables archives to offer researchers new, interesting, powerful, and productive visual explorations of collections. There are some great new tools under development. Examples include: Jeanne Kramer-Smyth's ArchivesZ, an "elastic list"

prototype at Syracuse, and relationship mapping tools such as NNDB Mapper.¹³

EAD gets you money

Grant agencies and other funders look favorably on and encourage EAD implementation as part of their granting process. For example, the guidelines for the NEH Preservation and Access, Humanities Collections and Resources encourage the use of EAD.¹⁴ The NISO/IMLS A Framework of Guidance for Building Good Digital Collections includes EAD as an appropriate metadata scheme for archives.¹⁵ NHPRC similarly endorses EAD in their guidelines.¹⁶

Knowledge gained mastering EAD is applicable in other contexts

In learning EAD, you will also develop skills that extend beyond encoding finding aids by gaining a basic understanding of XML and XML tools. So much digital data—in the library and archival communities and beyond—is stored and/or exchanged in the form of XML. These skills for staff will allow them to work with other standards such as MARCXML, MODS, and METS.

EAD paves the path to the future

Although today's researchers find collection descriptions using keyword searching on search engines, the Web of the future will be no place for unstructured data. The future is the "semantic Web" or linked data. Implementing EAD will help to position your institution for the future of internet applications.

Everybody's doing it!

Recognition of EAD's significance has become increasingly widespread, both within the US and internationally, and its use has expanded accordingly. The EAD Help Pages include a comprehensive list of various types of institutions that are currently successfully implementing EAD.¹⁷ We encourage you to look for institutions with a profile similar to yours. Knowing that your peer institutions are implementing may help you persuade those that hold the purse strings that EAD implementation is worth the resources it will take.

Let me just tidy this up first...¹⁸

Solution: Encode the data you have to provide minimum access.

As Voltaire wisely observed, “the perfect is the enemy of the good.”¹⁹ The desire to achieve perfection can sometimes get in the way of small improvements that iteratively help us to reach larger goals. In the case of providing better access to our collections, the urge to rewrite finding aids (or reprocess collections and then rewrite finding aids) is a huge barrier to providing interim access to the collection descriptions as they are now.

You must make every effort to make existing collection descriptions as accessible as possible, regardless of your intentions for them in the future. Although technically these collections do not represent a processing backlog since they do have descriptions, if those descriptions are not accessible, they present the same problems as unprocessed collections. They are hidden from all but those inside the institution. A survey conducted in 2003-2004 by Dennis Meissner and Mark Greene as background for the “More Product, Less Process” report found that backlogs are a key concern for the majority of donors, researchers, and resource allocators.²⁰ SAA’s code of ethics reminds us that “Archivists strive to promote open and equitable access to their services and the records in their care without discrimination or preferential treatment.”²¹ Likewise, the SAA/ALA Joint Statement on Access includes the following statements: “A repository is committed to preserving manuscript and archival materials and to making them available for research as soon as possible” and “As the accessibility of material depends on knowing of its existence, it is the repositories responsibility to inform researchers of the collections in its custody.”²²

In an increasingly online world, making your collection descriptions as accessible as possible to the widest possible audience is of paramount importance. Access deferred is access denied.

Who will do the work, and when?²³

Solution: Find low-impact ways to tackle EAD implementation.

Your staff is already stretched to the limit, and adding EAD implementation to an already bulging workload is rarely feasible. Taking a creative approach may help your institution get a toe in the door with a modest investment of time or money.

Join the club

Numerous state-, regional- or subject-based consortia have pooled resources to benefit member institutions by lowering barriers to EAD implementation. They generally offer a range of services such as best practice guidelines, stylesheets, templates and other tools, training, and hosting of data. They often apply for grant funding to convert finding aids or provide other services at little or no cost. Some consortia may be able to handle all of your needs, including encoding. Even if you don’t belong to a consortium, many of these organizations make their tools and guidelines freely available on their Web sites, so that others need not reinvent the wheel. Many consortial projects are grant-funded; the more contributors and users they have, the more likely they are to continue being funded. Contributing your EAD thus helps not only you but many other institutions and patrons. A partial list of regional, national and international consortia is included in *Appendix I*.

Take the first step

If you have collection-level records in the MARC format, consider creating basic EAD records through export (easily done using a tool such as MARCEdit).²⁴ You will then have a set of collection-level EAD data, and some experience working with the standard. The resulting files are also suitable to contribute to an EAD consortium. You may then decide to expand the minimal records, or you may decide to live with the fact that your EAD descriptions will not include inventories. Something is better than nothing, particularly from the user perspective.

Take an iterative approach

There is no rule that EAD encoding must be done once and only once, nor that it cannot be done until arrangement and description are complete. Implement EAD with a “More Product, Less Process”-like approach! A collection-level finding aid with minimal information can be produced at the time of accession. Areas such as `bioghist` and `scopecontent` can be expanded later. If the collection comes with a simple box list, that can be included at the outset, to be replaced later by a fuller inventory when processing is complete. Some collection management tools will produce collection descriptions as a byproduct of the arrangement and description process. EAD finding aids and online publishing free us from the static paper finding aid, offering instead an evolving document that changes and grows though the life of the collection.

Use someone else's time and talent

Do you have access to library school students, or other interns? Do you have an internship program? Consider offering an internship (or a series of internships) that focus on researching options and implementing EAD for your institution.

You don't have the time, but you do have the money

Maybe you don't have staff time, but you can make a one-time or ongoing investment of funds. If wholly or partially outsourcing is of interest, there are more details below.

It's so complicated!²⁵

- Solution: There are many options that make EAD simple
-

EAD can intimidate even tech-savvy staff, given the number of tags and the seemingly endless variety of ways they can be implemented, but not all elements are required. Use collection-level descriptions and minimum-level description elements—as given by DACS—to simplify EAD adoption.

A variety of tools exist to help mask the complexity of EAD and smooth the encoding path. Tools that assist in migration can also aid in the tagging process. See the section on migrating to a database or content management system (page 49). Other tools that can help mask complexity are listed in Appendix II. Please note that some of these tools are local solutions or strategies—that is to say, they have been customized for use within a consortium, institution or repository and may not work in your setting. We've included these because they may provide inspiration for your own use. The section on “*Specialized Migration and Conversion Tools*” may be especially useful.

Templates

Templates are EAD documents pre-populated with text that never changes (repository name, address, etc.) and with boilerplate text guiding the encoder to fill in proper data. This removes much of the angst of choosing what tags to use and how to use them, making it more like filling out a form.

Templates are simple to create. Using commercial XML authoring tools, you can create an EAD file with as much information as possible—including both text and attributes—already filled in, and use it as the basis for all new EAD files. For example, in the `controlaccess` section include one each of the possible child elements (`subject`, `persname`, `corpname`, etc) with the appropriate `@source` and `@encodinganalog` attributes filled in. Or in the `bioghist` section, include the phrase “MARC 545: Insert brief bio or company history.” Some example of templates are included in *Appendix II*.

The EAD Schema

Using the schema rather than the DTD version of EAD during authoring allows you to enforce various content limitations, such as correct formatting of `@normal` attributes for date elements, thus reducing the chance for errors. While leveraging the full power of the EAD schema may require a more substantial investment of time, schema-based validation can be used in combination with templates to ensure strict internal compliance. An example of institution-wide best practice guidelines implemented using the XML Schema version of EAD is that of Yale University’s Finding Aid Coordinating

Committee. Yale finding aids validate against the W3C Schema version of EAD, and compliance with Yale's EAD best practices is monitored via external validation against a RelaxNG schema.²⁶

Forms

An EAD finding aid can be presented as a fill-in-the-blank HTML form so that the archivist never sees any EAD at all. Although it does require behind-the-scenes coding to add the EAD tags, several institutions have employed this approach and there is likely a Web-based form that your institution can use immediately, or with minimal adaptations. We've included examples of several Web-based forms along with two examples of stand-alone forms in Appendix II.

Style sheets

Style sheets can display or hide various attributes or other text to enhance readability and aid in completing a template. If using oXygen, XMetaL, or some other commercial XML authoring tool, a Cascading Style Sheet (CSS) file can be automatically created the first time you open an EAD document; this style sheet can then be modified, enhanced, etc., to assist editors visually. Some examples of stylesheets are included in *Appendix II*.

...and more!

We've included information about other means of simplifying EAD implementation including pointers to commercial XML tools, content management systems for archives, and a variety of papers, production guides and case studies. You will find lists in *Appendix II*.

I don't know where to start!²⁷

■ **Solution:** Take it one step at a time and create a plan. The process can be broken into logical steps.

If you are overwhelmed and don't know how to get started, it may be helpful to think about EAD implementation in terms of a number of small steps. Start with what you have and where you need to go. Steps usually include documenting inhouse standards (e.g., should extent be expressed as linear feet or cubic feet? Are there existing best practice guidelines you wish you adhere to? What controlled vocabulary will you use for the controlaccess terms), selecting an encoding method (e.g. template with XML editor, database, full content management tool), data entry, selecting or creating a style sheet (to format the XML), and putting the files online. Additional decision areas may include whether and how to provide search capability, whether to provide alternate formats (e.g., a printer-friendly version), whether to link finding aids to digitized content, and so on. For an overview of possible steps, we have included one such plan in Appendix III, and this example may get you started. You may also refer to the EAD Application Guidelines, specifically Chapter 2, for more details and additional food for thought.²⁸

Section II: Addressing Technical Problems

I want my data to be stored in a format that will give me flexibility going forward²⁹

Solution: EAD is a non proprietary component of a flexible framework.

In any conversion process, you should be concerned with maintaining flexibility, and not losing information. Converting paper files to electronic format increases the usability of your data; EAD encoding offers both flexibility and additional options since EAD can be “crosswalked” into other standards like MARC XML and Dublin Core. Going further, a well-chosen database or CMS (Content Management System) can provide additional output options. It’s all about choices along this continuum. We’ve presented three ways to think about moving forward—each one will advance you to greater flexibility.

Option 1: Migrating hard copy or word-processing files to EAD

Outsource

Outsourcing data conversion is the fastest and easiest solution for hard copy or word-processing file conversion. Depending on your staff costs, this may or may not be the cheapest option. Generally conversion houses quote a per-page (for hard copy) or a per-byte (for electronic files) rate. Outsourcing requires considerable up-front work in determining tagging specifications, since EAD is highly flexible. Adhering to DACS, ISAD(G), and/or RLG’s EAD Best Practice Guidelines³⁰ or other widely-accepted standards or guidelines is strongly advised. Standards and “best practices” save time and effort, and your end product will be more likely to work with widely-used tools and in aggregations of other EAD files. The EAD Listserv is an excellent resource in terms of getting up to the minute information regarding current vendors.

Outsourcing is most successful when both parties are very clear on the encoding specifications. A sample encoding specification, including where to get various pieces of data and how to handle particular situations is available on the Syracuse University Library Web site.³¹

In-house encoding

In-house conversion offers numerous benefits: fostering staff skills, flexibility in schedule and workflow, and direct control over process and inputs. Basic XML skills are not difficult to acquire, and having internal staff with XML knowledge may benefit other library processes and projects as well. Various tools can speed/ease creation of EAD. The two following options—authoring tools and scripts—assume starting with electronic files (e.g., word processing files, or text files obtained through converting hard copy using optical character recognition, or OCR).

As part of in-house conversion you can use commercial XML authoring tools (we’ve included a list of the most well-established in Appendix II). You may also be able to manipulate the text using scripts. Once you have gained confidence in understanding and defining the EAD output you expect, then any programmer with experience of scripting languages like Perl or VisualBasic and XML could write scripts to produce the desired output from your existing input. EAD produced this way can be validated and/or edited using commercial tools, or using the free online W3C validator³² or validation could be part of the scripted process(es).

Option 2: Exporting EAD from a content management system

Some content management systems (CMS), such as the Archivists’ Toolkit or Archon, are capable of producing EAD. If your CMS does not output EAD directly, two key questions are 1) whether required EAD elements are separate fields in the database, and 2) whether your database exports XML.

Table 1. Tips for producing EAD from managed content under various export scenarios.

Separate Fields?	Data Exports as XML?	Notes
Yes	Yes	Export the data as XML and then use XSLT to convert to EAD. (XSLT is a language for transforming XML documents into HTML documents or to other XML documents, in this case transforming data to EAD). This requires some knowledge of XSL, or the funds to contract out XSL development. Ideally it would be a one-time cost.
Yes	No	Export the data in some other structured form (comma-separated values, for example) and identify a scripting approach to process the data and convert to EAD. This requires some programming or scripting knowledge. Perl is an example of a scripting language that is useful in this context.
No	No	Determine whether scripts can be written to parse output from your database and generate EAD; whether, and if so how, the database needs to be modified to be able to export EAD; or whether perhaps the appropriate solution is migration to an EAD-capable database.

Option 3: Migrating to a database or content management system capable of producing EAD for permanent storage and maintenance

Using a database to create and store data elements of finding aids simplifies data entry, reduces the possibility of tagging errors, ensures consistency in output, and offers the possibility of exporting to formats other than EAD. However, some full-fledged archival management systems may be “overkill” for a legacy conversion project in terms of features, price, and learning curve.

If your data is in spreadsheet or word processor format, or in a database that will not map directly to EAD, migrating to an EAD-capable database may be a useful solution. The key question here is whether the data is easily mapped to the target database, and whether the time involved in migration will in the long run result in the best solution for your needs. A list of content management systems is included in Appendix II.

Doors are closed to open source³³

- **Solution:** Outline the upsides of open source software

Making a choice

The open source/commercial distinction is one of many factors that should play a role in your archival management system decision-making process. The most important part of selecting a system is to choose one that has the features you need. Resources such as *Archival Management Software: A Report for the Council on Library and Information Resources* will help you with the selection process.

Availability of open source software

At least two tools that produce EAD are distributed as open source software (OSS).³⁴ OSS is produced in a way that allows others to adapt, modify and redistribute the underlying code and is frequently associated with a “community” of developers.

Your institution is most likely using open source software already in some context and that may make it easier to bring in an open source EAD tool. You may be using the Apache Web server, database platforms such as MySQL, and desktop applications such as the Mozilla Firefox Web browser. Several open source integrated library systems are available, including Koha and Evergreen³⁵ Open source digital repository systems include Fedora, DSpace, EPrints, and Greenstone.³⁶

Lack of conflict with commercial software/commercial enterprises

Some institutions have a policy against implementing open source software, preferring instead to license or purchase software that includes support or is backed by a reputable company. Open source software does not preclude commercial support. Support contracts are available for many open source software packages, including the open source ILS system previously mentioned. Commercial support for OSS EAD tools is not currently available, but this is evolving.

Problems with publishing³⁷

- Solution:** Let the browser do the work, or use existing tools that incorporate publication functionality.
-

A major obstacle preventing wide-scale adoption of EAD is delivering EAD-encoded finding aids to users online. Creating EAD finding aids may require a different set of skills than publishing them, including authoring XSLT stylesheets, installing software, configuring a server, and so forth. There are few inexpensive, “out-of-the-box” solutions for publishing EAD online. However, archives have several options. From simplest to hardest, these include: contributing records to a shared finding aids repository; delivering EAD directly to the browser; converting records to HTML or PDF for Web display; using inexpensive tools to enable searching of HTML and XML files; using an archival management system; and using an XML publishing platform.

Contribute to a shared finding aids repository

Rather than developing their own technical infrastructure for delivering finding aids, some archives choose to deposit them in regional finding aid repositories. The finding aids are hosted centrally and provide a single point of access to finding aids from multiple institutions. We've included a partial list of finding aid repositories/regional consortia in Appendix I.

Some archives may want to contribute finding aids to a repository *and* make them available via their own Web sites.

Deliver EAD directly to the browser

This is by far the simplest and easiest approach. You can deliver XML directly to most recent Web browsers (e.g. IE 5+, Firefox .9+). To transform the EAD XML file to HTML within the Web browser (on the client side), include a processing instruction in the XML document pointing either to an XSLT stylesheet³⁸ (the preferred method) or CSS file.³⁹ However, some institutions may not want to provide access to their raw XML files, particularly if they include sensitive information in their finding aids that they don't want to display to the public. Moreover, browser support for XML is still uneven⁴⁰ (for instance, at the time of the writing of this report, Google's Chrome browser is reported to not provide full XML support).

Convert your EAD to HTML or PDFs for Web display

Instead of displaying the raw XML using a Web browser, convert EAD finding aids to a static files in a human-readable format. By applying XSLT stylesheets to XML finding aids, archives can generate multiple forms of output, including HTML and PDF. Such conversion can be accomplished in batch. HTML or PDF files can then be uploaded to a standard Web server to support research and discovery.

Developing XSLT stylesheets requires some technical knowledge, but several consortia and archives have made available XSLT stylesheets that archives can easily adapt for their own institutions. Some examples are listed in Appendix II.

Delivering HTML or PDF rather than EAD may be attractive to archives that lack technical staff

to support XML publishing, but these methods have several drawbacks. They do not take full advantage of having archival information marked up in EAD; searches cannot be restricted to particular EAD elements. Moreover, every time the finding aid is updated, the HTML must be regenerated and uploaded to the server. Some archives use a hybrid approach; indexes are created from EAD files to enable fine-grained searching, but the HTML file is delivered to the user when they want to view the finding aid. Syracuse University Libraries take this approach.⁴¹

Use inexpensive tools to enable searching of HTML and XML files

Even if an archive lacks a substantial budget or large technical staff, it can choose from several inexpensive, easy-to-implement tools that support indexing and searching EAD files. One example is Swish-e, “a fast, flexible, and free open source system for indexing collections of Web pages or other files.”⁴² Google Site Search also provides an inexpensive, customizable way of searching your Web pages.⁴³

Use an archival management system that supports publication

Many archival management systems enable publication via export of finding aids in EAD, HTML, or PDF. By using archival management systems, archivists can streamline workflows, avoid duplicating data in multiple places, find and share information more easily, manage collections, and generate reports and statistics.⁴⁴ A list of archival management systems that support Web publishing of finding aids are listed in *Appendix II*.

Archival management systems have some drawbacks: they may enforce a rigid workflow, it can be difficult to import data, and some are costly to implement. On the other hand, archival management systems can enable archives to create, manage, and share archival information more efficiently.

For the sake of interoperability, the selection criteria for a commercial archival management system must include the ability to import and export EAD files, ideally both one at a time and as a batch process. Commercial packages provided by Adlib, CALM, CuadraStar, and Eloquent Systems all provide batch and individual import and export of EAD finding aids. If your institution requires a hosted solution, many vendors offer such an option.

Use an XML publishing platform

XML publishing platforms enable documents to be searched, browsed and displayed. Implementing them requires fairly sophisticated systems administration and programming knowledge. Some XML publishing platforms are listed in Appendix II.

EAD can be complicated (but there's hope...)

EAD was designed to be flexible in order to accommodate a broad range of archival practice. In offering flexibility, the standard has succeeded almost too well.

Freedom of choice in implementation means, for example, that three people could encode the extent of a collection in three different ways. This flexibility in implementation can cause difficulties for aggregators who harvest EAD data from multiple institutions for indexing and searching. It also hinders tool development since tool builders must either allow for multiple encoding options or choose one “right way,” when there are multiple correct ways to encode the same thing. EAD’s inherent complexity makes it difficult for institutions to make decisions regarding implementation. Those who are choosing tools must evaluate the choices made by tool builders to ensure that outputs meet their own best practice guidelines.

So what to do? Make a decision. Document the decision. Apply it consistently. Until the flexibility inherent in EAD is reigning in, institutions can maximize the consistency of their data by:

- 1) Selecting a template in use at one or more institutions, or creating a template that adheres to a “best practice” document in use at one or more than one institution. Once you’ve established a template, deviate from it as little as possible.
- 2) Clearly document how dates, extent, etc., should be encoded. Follow your own documentation rigorously.
- 3) Refrain from excessively complex coding (for example, nesting duplicate `scopecontent` elements within each other).
- 4) Refrain from adding unnecessarily elements such as `list` elements within a control access simply to achieve a desired appearance in the output. EAD should be only be used to encode the structure and content of a document; appearance should be controlled by the stylesheet.

Remember, too, that the entire EAD tagset need not be used. As mentioned above, limiting yourself

to collection-level descriptions and the DACS minimum-level description elements can simplify EAD immensely.

Getting through it

Despite a more than a decade of practice, archivists still encounter significant barriers in EAD implementation. We hope this paper gives you options to get over hurdles, under obstacles, around complexity, and through difficulty.

We recognize that EAD can be challenging. Examples of EAD’s complexity can be found easily by looking through the EAD Tag Library. Many elements, including `accessrestrict`, `controlaccess`, `bioghist`, and `note`, may be repeated within an element with the same name to an arbitrary depth; for example, EAD allows one to encode nested `controlaccess` elements with no restrictions on how deep that nesting goes. In addition, EAD has seventeen linking elements; of those seventeen, twelve of those elements allow the `href` attribute, which allows linking to resources external to a given EAD file. Elements that allow “mixed content” (those that can contain both text and other elements in arbitrary order) can present problems when importing EAD to a database or porting to another data scheme. Some elements that can be full of mixed content and contain information that would be lost in migration to a database (or would require additional tagging after export) are `p`, `listitem`, `bibref`, and `head`.

EAD will be under active review in the near future. We recommend that the Technical Subcommittee for EAD (the soon-to-be charged successor to the EAD Working Group) and the archival community as a whole consider ways in which EAD can be simplified.

As reflected in the large number of EAD tools listed in this paper and its appendices, there are many choices for would-be EAD implementers. While diversity and choice is a good thing, the range and number of available tool choices provide additional complexity. By highlighting tools that are already available, we encourage institutions to utilize work that has been done elsewhere and not to invest what might be unnecessary development effort.

Acknowledgments

We are grateful to the reviewers who provided valuable and useful feedback on early drafts of this document:

- Doug Dunlop
Smithsonian Institution
- Glenn Gardner
Library of Congress
- Suzanne Pilsk
Smithsonian Institution
- Michael Rush
Yale University
- S. Diane Shaw
Smithsonian Institution
- Bradley D. Westbrook
University of California, San Diego
- Jackie Dooley
OCLC Research
- Ricky Erway
OCLC Research
- Jennifer Schaffner
OCLC Research

Appendix I. Consortia and EAD Aggregators

Almost all can provide a means of publishing finding aids, or may serve as an additional distribution channel for collection descriptions. Many also have tools to aid in EAD creation, provide instruction opportunities, and have developed best practice guidelines.

United States

- **Archival Resources in Wisconsin:**
<http://digital.library.wisc.edu/1711.dl/wiarchives>
- **Archives Florida:**
<http://palmm2.fcla.edu/afl/>
- **Arizona Archives Online:**
<http://azarchivesonline.org>
- **Black Metropolis Resources Consortium:**
<http://www.blackmetropolisresearch.org> [forthcoming]
- **Historic Pittsburgh:**
<http://digital.library.pitt.edu/pittsburgh/>
- **Kentuckiana Digital Library:**
<http://kdl.kyvl.org/>
- **Mississippi Digital Library:**
<http://msdiglib.net/>
- **Mountain West Digital Library:**
<http://mwdl.org/index.php/search/results?format=ead>
- **Northwest Digital Archives:**
<http://nwda.wsulibs.wsu.edu>
- **OhioLINK:** <http://ead.ohiolink.edu>
- **Online Archive of California (OAC):**
<http://www.oac.cdlib.org/>
- **Pennsylvania Digital Library:**
<http://padl.pitt.edu/>
- **Rhode Island Archival and Manuscript Collections Online (RIAMCO):**
www.riamco.org [forthcoming]
- **Rocky Mountain Online Archive:**
<http://rmoa.unm.edu>
- **Texas Archival Resources Online (TARO):**
<http://www.lib.utexas.edu/taro>
- **Virginia Heritage Project:**
<http://www2.lib.virginia.edu/small/vhp/>

UK and Continental Europe

- **A2A (Access to Archives, United Kingdom):**
<http://www.nationalarchives.gov.uk/a2a/>
- **Archives Hub (United Kingdom):**
<http://www.archiveshub.ac.uk/>
- **Archives Portal Europe:**
<http://www.apenet.eu/> [forthcoming]
- **MALVINE (Europe):**
<http://www.malvine.org/>
- **National Archival Database of Sweden:**
http://nad.ra.se/static/back_eng.html

Subject based

- **Navigational Aids for the History of Science in Europe (NAHSTE):**
<http://www.nahste.ac.uk/>
- **Guide to Australian Literary Manuscripts:**
<http://findaid.library.uwa.edu.au/>
- **Irish Literature Collections Portal:**
<http://irishliterature.library.emory.edu/>
- **Physics History Finding Aids Web site (PHFAWS):**
<http://www.aip.org/history/nbl/findingaids.html>

Other

- **ArchiveGrid:**
<http://www.archivegrid.org>
OCLC's Archive Grid combines finding aids with MARC records to create one-stop-shopping for users. Heavy representation from US institutions, also representation from outside the US. Contribution is free and open to any institution.

Appendix II. Tools

Templates

Examples of templates include:

- Northwestern University:
<http://staffweb.library.northwestern.edu/dl/ead/template.xml>
- Syracuse University:
http://library.syr.edu/digital/guides/ead/aaa_template.xml
- Indiana University Bloomington:
<http://www.dlib.indiana.edu/services/metadata/activities/eadDocumentation.shtml>

Web-based forms

Following are a few examples of Web-based forms. Some produce only collection-level EAD (that is, they do not include the inventory in the <dsc> section), while others produce complete inventories. They differ in ease of use, complexity, and quality of product. The first two are probably the most complete.

- California Digital Library: http://www.cdlib.org/services/dsc/tools/ead_webtemplates.html

The California Digital Library has created numerous online templates that create EAD from typed or copy-pasted data. One or more of these may be adaptable for your institution's needs.

- ArchivesHubUK: <http://www.archiveshub.ac.uk/eadform2002.html>

Web-based form, renders complete EAD document. Enables editing of uploaded files, creation of new ones, saving of draft file between editing sessions, preview feature. Links to digital surrogates are easy to add, common markup tags (e.g. paragraphs, lists, titles, links) can be added via a right-click menu, and a special characters keyboard is provided. Components can easily be added to create sub-fonds descriptions, and a tree structure will show exactly what the hierarchy looks like.

- Northwestern University: <http://staffweb.library.northwestern.edu/dl/ead/eadchef/template.cgi/ead/nul>
- Notre Dame: <http://classic.archives.nd.edu/ead/ead.htm>
includes both collection level and inventory forms
- Berkeley: <http://sunsite3.berkeley.edu/ead/tools/template/>
cgi Web application; appears to be collection-level only. Last update to page is 2005.
- Western Kentucky: http://pax.uky.edu/template-v1-cgi/template.pl/KNVUA_generic.tmp
- Austin College Xforms tool: <http://www.archivists.org/saagroups/ead/tools.html>
Developed by Justin Banks at Austin College. Requires server that supports Xforms. Additional information from SAA 2007 available here: <http://matienzo.org/saa2007descriptionexpo>.

Standalone forms

- The University of Utah has a java-based EAD authoring tool called xEAD, currently publicly available at https://lsta.lib.byu.edu/lstawiki/index.php/XEAD_Project. The application opens an EAD file in its buffer, allowing users to manipulate the data and then resave.
- The German Bundesarchiv, with assistance from the Mellon Foundation, has developed a tool called Midosa Editor for XML or MEX. It is available in English and German and for both OSX and Windows. MEX provides an authoring environment for creating a variety of levels of EAD records and includes built-in publication to HTML capability. Background information is available at <http://www.bundesarchiv.de/daofind/en>. Downloads and a quick-start guide are available from the MEX SourceForge wiki at <http://mextoolset.wiki.sourceforge.net/>.

Style sheets for authoring finding aids

- Yale University has developed the Yale Finding Aids Creation Tool (FACT), which is a customized version of XMetaL Author. The tool includes a set of style sheets, macros, and other customizations. More information is available at <http://yalefact.pbworks.com>.
- Northwestern Digital Archives: <http://nwda.wsulibs.wsu.edu/tools/ead.css>
- Syracuse University: <http://library.syr.edu/digital/guides/ead/ead.css>

XSLT stylesheets for displaying finding aids

- EAD 2002 Cookbook: <http://www.archivists.org/saagroups/ead/ead2002cookbookhelp.html>
- EAD Help pages' user contributed style sheets: <http://www.archivists.org/saagroups/ead/stylesheets.html>
- UC Berkeley EAD Tools: <http://sunsite3.berkeley.edu/ead/tools/>
- NC Echo EAD Tools: <http://www.ncecho.org/dig/ncead.shtml#tools>
- University of Minnesota: <https://wiki.lib.umn.edu/Staff/FindingAidsInEAD>

Commercial XML tools for EAD encoding

- oXygen: <http://www.oxygenxml.com/>

Commercial XML authoring and editing software. Data entry, copy-paste; can handle DTD or schema

- XMetaL: <http://na.justsystems.com/content-xmetal>

Commercial XML authoring and editing software. Data entry, copy-paste; can handle DTD or schema.

- XMLSpy: <http://www.altova.com>

Commercial XML editing, authoring, development environment.

- NoteTab Pro: <http://www.notetab.com/>

General purpose text editor which can be easily customized to handle EAD. Clip libraries and other add-ons are available here <http://www.archivists.org/saagroups/ead/tools.html>

Content Management Systems for Archives

- Archon: <http://www.archon.org>

Open source full archival management system, developed by University of Illinois with funding from Mellon Foundation. Capable of ingesting MARC or CSV format. Can export EAD and MARC.

- Archivists' Toolkit: <http://www.archiviststoolkit.org>

Open source full archival management system, developed by UCSD, NYU, and Five Colleges Inc. and funded by Mellon Foundation. Capable of ingesting tab-delimited fields and MARC as well as preexisting EAD. Can export EAD, Marc, Dublin Core, MODS, METS.

- MSAccess, FileMakerPro, etc.

Commercial but reasonably priced general database development tool. A database developed in-house with the appropriate fields, in conjunction with XSL, java, or other scripting languages, can generate EAD.

Archival management systems that support publishing

- Adlib (commercial): <http://www.adlibsoft.com/>

- Archon (open source): <http://archon.org/>

- Calm (commercial): <http://www.crxnet.com/page.asp?id=57>

- Cuadra STAR/Archives (commercial): <http://www.cuadra.com/products/archives.html>

- Eloquent (commercial): <http://www.elloquent-systems.com/products/archives.shtml>

- ICA-AtOM (open source): <http://ica-atom.org> [currently available in Beta release]

- Minisis M2A (commercial): <http://www.minisisinc.com/index.php?page=m2a>

Various other commercial archival management systems can import and export EAD, including:

- Minisis: <http://www.MINISISInc.com>
- Eloquent Archives: <http://www.elloquent-systems.com/products/archives.shtml>
- AdLib: <http://www.adlibsoft.com>
- CALM: <http://www.ds.co.uk>

For detailed information on these and other commercial tools, refer to Lisa Spiro's 2009 report for CLIR, Archival Management Software (<http://clir.org/pubs/reports/spiro2009.html>), or to the associated wiki at <http://archivalsoftware.pbwiki.com/>.

XML publishing platforms⁴⁵

- XTF: <http://xtf.wiki.sourceforge.net/>
“A flexible indexing and query tool that supports searching across collections of heterogeneous data and presents results in a highly configurable manner.” XTF supports powerful searching, faceted browsing, and viewing search terms in context.
(Open source; used by California Digital Library and numerous others)
- Mark Logic: <http://www.marklogic.com/>
The University of Chicago is developing an XML publishing infrastructure built on MarkLogic, a native XML database. The front end can be built on any platform and provides flexible display options. The University of Chicago’s code will be available to anyone. Archives that want to use the software will need MarkLogic, but there is a free (limited) version that will suffice for small institutions.
(Commercial; used by University of Chicago)
- PLEADE: <http://www.pleade.org/en/>
“open source search engine and browser for archival finding aids encoded in XML/

EAD. Based on the SDX platform, it is a very flexible Web application.”

(Open source; used by Denver Public Library)

- Cocoon: <http://cocoon.apache.org/>

Cocoon is an open source XML publishing framework that applies XSLT stylesheets to the EAD finding aid to display HTML. Used in tandem with indexing technologies such as Lucene or eXist.

(Open source; used by Ohio State and the Five College Finding Aids Access Project)

- DLXS: <http://www.dlxs.org>

XML-aware search engine (XPAT or XPAT Lite) with DLXS middleware which includes a “class” for finding aids (currently in prototype).

(Open source component, with commercial options available; used by University of Michigan and University of Minnesota Libraries⁴⁶)

- Cheshire3: <http://www.cheshire3.org/>

“Fast XML search engine.” Standards compliant, with support for Open Access Initiative (OAI) protocols and Z39.50. Modular and configurable.”

(Open source, used by University of Liverpool Special Collections and Archives and ArchivesHub)

Specialized migration or conversion tools

- MARC to EAD—MarcEdit: <http://oregonstate.edu/~reeset/marcedit/html/index.php>

If you have MARC records for your manuscript collections, you can quickly and easily generate skeleton (i.e., collection-level) EAD records from it using MarcEdit. MarcEdit uses xsl style sheets which can easily be modified/customized. Developed by Terry Reese at Oregon State University; free.

- Excel to EAD using Mailmerge

Excel is an immensely useful tool for generating the code for lengthy inventories of minimal depth. Text can be entered into a spreadsheet, then columns can be added before and after

the text and populated with the correct EAD elements. For collections with large inventories, either copy-pasting (from Word, RTF, or txt) or entering afresh in Excel may be a workable solution. Excel's MailMerge feature can also be employed to automatically generate coded data from an Excel spreadsheet; see video here <http://archives.state.ut.us/containerlist/containerlist.html>. This could be used in conjunction with MarcEdit which generates the collection-level part of EAD to produce a full EAD inventory. Indiana University has posted detailed instructions and an Excel template for encoding lengthy inventories at <http://www.dlib.indiana.edu/services/metadata/activities/eadDocumentation.shtml> (see the section "Using Excel to assist with encoding").

- **Text to EAD—EAD Conversion:**
<http://agileimage.com/html/ead/>

Reads a text version of inventory and generates an EAD-encoded version. Last update to Web site was in 2004.

- **Text to EAD—MSWord**

Starting with an electronic file of an inventory, a surprising amount of tagging can be done simply using MSWord's search and replace feature with tabs and regular expressions, including locating and tagging `unitdate`, `extent`, long inventories, etc. Good overviews of regular expressions in word are available here:

- <http://office.microsoft.com/en-us/help/HA010873051033.aspx>
- <http://office.microsoft.com/en-us/help/HA010873041033.aspx>
- <http://word.mvps.org/FAQs/General/UsingWildcards.htm>
- Date normalization—tri-XMLdate-normalizer.pl: <http://www.archivists.org/saagroups/ead/tools.html>

Developed by Jason Casden at The Ohio State University; free. For large files, inserting the `normal` attribute for `unitdate` elements can be extremely time-consuming. This Perl script automatically recognizes numerous date formats and inserts the

`normal` attribute. Offers options to overwrite existing values or leave them.

Papers, production guides, case studies, etc.

- **Northwest Digital Archives Standards Working Group Review of Web Templates for EAD Creation** (October 2008): http://www.orbiscascade.org/index/cms-filesystem-action?file=nwda/files/bowmanreview_200810.pdf – excellent and thorough review of available Web-based templates that generate EAD.
- **OhioLINK EAD Starter Package** (April 2008):
http://platinum.ohiolink.edu/dms/ead/contentguide/EAD_starter_packet_v4.pdf
– Draft of paper looking at steps in an EAD conversion project; very rough but good info.
- **University of Indiana EAD guide:** <http://www.dlib.indiana.edu/services/metadata/activities/EADMManual.pdf> – includes detailed instructions for using oXygen.
- **Utah State Archives EAD Project:** <http://www.archives.state.ut.us/research/inventories/ead.html> – includes detailed discussion of how they converted their legacy finding aids using a combination of tools (HTML, Excel, WordPerfect, etc).
- **NYU Archives EAD Production Guide:**
<http://www.nyu.edu/library/bobst/research/arch/eadProduction.htm> – detailed procedures used by NYU, includes detailed instructions for NoteTab Pro.
- **The EAD Help Pages:** <http://archivists.org/saagroups/ead/sitesann.htm> – more than 80 institutions currently implementing EAD, including brief descriptions of each institution's approach and a point-of-contact. Implementations run the gamut from extremely simple (EAD put online with a style sheet) to extremely sophisticated (databases that provide server-side transformations and advanced search capabilities).

Appendix III. EAD Migration, Creation and Publication Paths

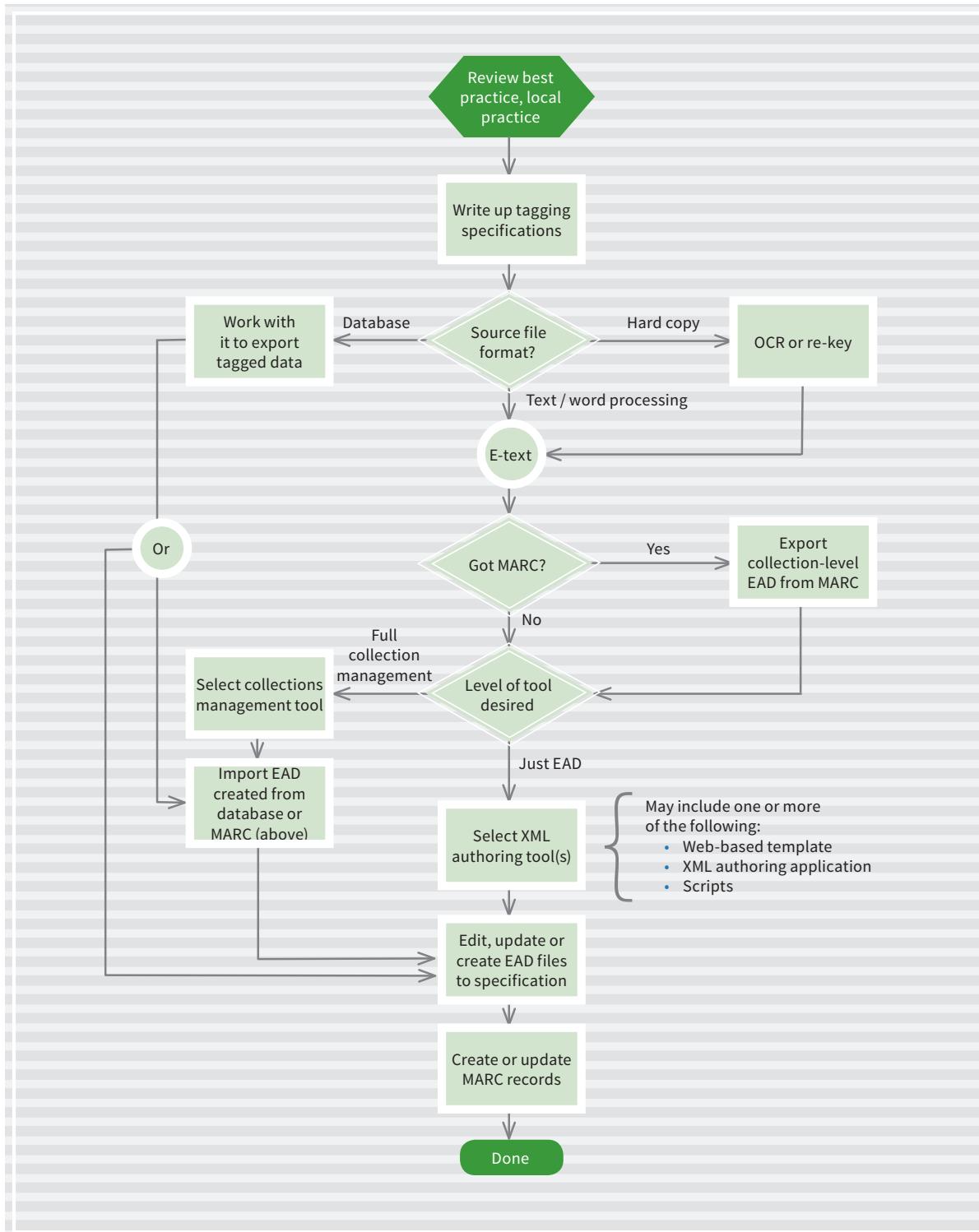


Figure 1. EAD Migration and/or Creation

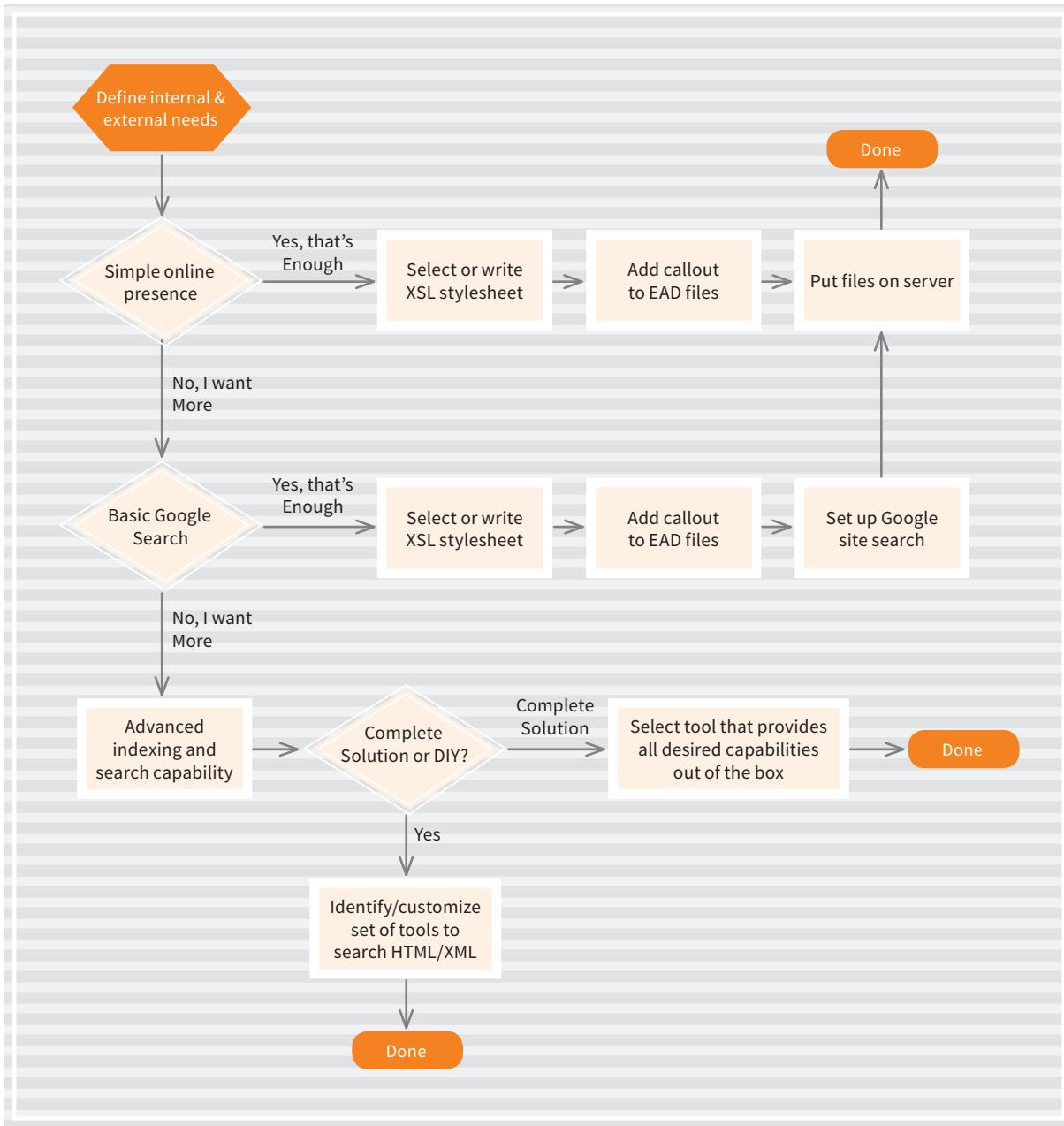


Figure 2. EAD Publishing Paths

Simplest path (conversion from MARC records)

- Export collection-level EAD records from MARC using MarcEdit.
- Validate and correct errors using W3C's online validation tool (<http://validator.w3.org>).
- Select or create an XSL style sheet.
- Put files on server.

Simplest path (starting without MARC)

- Use one of the many existing Web-based templates to generate collection-level EAD files.
- Validate and correct errors using W3C's online validation tool (<http://validator.w3.org>).
- Find or create an XSL style sheet.
- Put files on server.

The above solutions require minimal time, expertise, and money, yet yield online collection-level descriptions that will be “crawled” by Google and other search engines making them discoverable via the open Web, and valid EAD files which can be contributed to consortia (a list of consortia can be found in Appendix I) or aggregators like OCLC’s ArchiveGrid. Files can easily be “upgraded” at some later date without reworking.

More sophisticated path

- Choose one of the full collection-management packages such as Archivists’ Toolkit or Archon.
- Perform data entry and/or import to level desired.
- Install and configure XTF installation for Web availability, browsing, searching, etc.
- Export EAD to be indexed/ searched by XTF system.

This yields a fully-functioning database and sophisticated search capabilities but requires substantial technical knowledge to install and configure the XTF installation.

Notes

All Web links were verified 25 February 2010.

- ¹ EAD Roundtable, EAD (*Encoded Archival Description*) Help pages, sponsored by the Society of American Archivists, <http://www.archivists.org/saagroups/ead/>. In particular, see the section on “What is EAD,” at <http://www.archivists.org/saagroups/ead/aboutEAD.html>.
- ² Archivists Toolkit User Group, “2008 AT User Group Survey Results,” <http://archiviststoolkit.org/sites/default/files/AT%20User%20Group%20SurveyResultsFD.pdf>. Additional breakdown of survey results supplied by Brad Westbrook to the working group via e-mail on 18 December 2008.
- ³ Tatem, Jill, “EAD: Obstacles to Implementation, Opportunities for Understanding,” *Archival Issues* 23,2, (1998): 155-169; Roth, James M., “Serving Up EAD: An Exploratory Study on the Deployment and Utilization of Encoded Archival Description Finding Aids,” *The American Archivist*, 64,2 (2001): 214-237; Dow, Elizabeth H., “EAD and the Small Repository,” *The American Archivist*, 60,4 (Fall 1997): 446-455.
- ⁴ Wisser, Katherine M., *EAD Tools Survey*, Society of American Archivists, EAD Roundtable, (August 2005), <http://www.archivists.org/saagroups/ead/EADToolsSurvey.pdf>; Yaco, Sonia, “It’s Complicated: Barriers to EAD Implementation,” *The American Archivist*, 71,2 (Fall/Winter 2008): 456-475.
- ⁵ A recent report from CLIR documents the current state of archival management software, much of which provide support for EAD: Lisa Spiro, *Archival Management Software, A Report for the Council on Library and Information Resources*, CLIR Reports (January 2009), <http://www.clir.org/pubs/reports/spiro2009.html>.
- ⁶ Children’s Television Workshop, *Sesame Street*, Episode 0299 (December 9, 1971).
- ⁷ Yaco, “It’s Complicated,” 468: “another key barrier is ‘Lack of institutional support’”; Wisser, *EAD Tools Survey*, 19: “Some folks in the organization... are unconvinced of EAD’s staying power and thus stingy when it comes to allocating resources and time for relevant training. Is there an EAD eye-opener kit?”
- ⁸ International Council on Archives. ISAD(G): General International Standard Archival Description, Second Edition, <http://www.ica.org/en/node/30000>; Society of American Archivists, *Describing Archives: A Content Standard (DACS)*, <http://www.archivists.org/governance/standards/dacs.asp>.
- ⁹ Society of American Archivists. “Appendix A: EAD Crosswalks,” Encoded Archival Description Tag Library, Version 2002, The Library of Congress, http://www.loc.gov/ead/tglip/appendix_a.html.
- ¹⁰ Spiro, *Archival Management Software*.
- ¹¹ U.S. Library of Congress Network Development and MARC Standards Office, *EAD (Encoded Archival Description) Electronic List*, <http://www.loc.gov/ead/eadlist.html>; Society of American Archivists, Web site, www.archivists.org; Rare Book School, Web site, <http://www.rarebookschool.org/>.
- ¹² Szary, Richard V., “Encoded Finding Aids as a Transformative Technology in Archival Reference Service,” in *Encoded Archival Description on the Internet*, ed. Daniel V. Pitti and Wendy M. Duff, 187-197. Binghamton, NY: Haworth Information Press (2001).
- ¹³ ArchivesZ, <http://archivesz.com>; Syracuse University Library, “Elastic Lists: EAD demo”, <http://library.syr.edu/digital/guides/ElasticLists-EAD/binv3/index.html>; NNDB Mapper, <http://mapper.nndb.com/>.

- ¹⁴ National Endowment for the Humanities, Division of Preservation and Access, “Frequently Asked Questions: Humanities Collections and Resources,” <http://www.neh.gov/grants/guidelines/collections&resfaqs.html>.
- ¹⁵ National Information Standards Organization, *A Framework of Guidance for Building Good Digital Collections, 3rd Edition*, with support from the Institute for Museum and Library Services (IMLS), <http://framework.niso.org/node/38>.
- ¹⁶ National Historical Publications and Records Commission (NHPRC), “Basic Projects Grant Announcement,” <http://www.archives.gov/nhprc/announcement/basic.html>, and “Detailed Processing Projects Grant Announcement,” <http://www.archives.gov/nhprc/announcement/detailed.html>.
- ¹⁷ EAD Roundtable, “90 Implementors Currently Listed,” *EAD Help Pages* <http://www.archivists.org/saagroups/ead/implementors.html>.
- ¹⁸ Yaco, “It’s Complicated,” 459: “Concern about the quality or completeness of finding aids often causes archivists to rewrite legacy finding aids before EAD encoding.”
- ¹⁹ Voltaire, *La Béguule: Conte moral*, (1772).
- ²⁰ Mark A. Greene and Dennis Meissner, *More Product, Less Process: Pragmatically Revamping Traditional Processing Approaches to Deal with Late 20th-Century Collections*, <http://ahc.uwyo.edu/documents/faculty/greene/papers/Greene-Meissner.pdf>, page 2. (Also published as “More Product, Less Process: Revamping Traditional Archival Processing,” *American Archivist* 68,2 (2005): 208-63.)
- ²¹ Society of American Archivists, “Code of Ethics for Archivists,” http://www.archivists.org/governance/handbook/app_ethics.asp.
- ²² Society of American Archivists, “ALA-SAA Joint Statement on Access: Guidelines for Access to Original Research Materials,” <http://www.archivists.org/statements/alasaa.asp>.
- ²³ Yaco, “It’s Complicated,” 466: “lack of staff...[along with] lack of time.”
- ²⁴ “MarcEdit—Your Complete Free MARC Editing Utility,” (Web site hosted at Oregon State University) <http://people.oregonstate.edu/~reeset/marcedit/html/index.php>.
- ²⁵ Yaco, “It’s Complicated,” 460: “the complexity of EAD itself [is] a deterrent to implementation.”
- ²⁶ Yale University Library FACC: Finding Aids Coordinating Committee, “Yale University EAD Encoding Best Practice Guidelines,” <http://www.library.yale.edu/facc/bpgs.html>.
- ²⁷ Yaco, “It’s Complicated,” 466: “Lack of a plan”.
- ²⁸ Society of American Archivists, Encoded Archival Description Working Group, *Encoded Archival Description Application Guidelines for Version 1.0*, (Encoded Archival Description (EAD), Document Type Definition (DTD), Version 1.0, Technical Document No. 3), (1999), <http://www.loc.gov/ead/ag/aghome.html>.
- ²⁹ Spiro, *Archival Management Software*, 6: “Archival material is so specific that you don’t want to get locked into anything... Ideally, I would want something that would also preserve that information in a format that is able to migrate if needed.”
- ³⁰ RLG EAD Advisory Group, *RLG EAD Best Practice Guidelines for Encoded Archival Description*, (August 2002), <http://www.oclc.org/programs/ourwork/past/ead/bpg.pdf>.
- ³¹ Michele Combs, *EAD Tagging Specs (Revised 10-3-08)*, Syracuse University Library, http://library.syr.edu/digital/guides/ead/tagging_specs.doc.
- ³² World wWde Web Consortium (W3C), *Markup Validation Service*, <http://validator.w3.org>.
- ³³ Spiro, *Archival Management Software*, 12: “Some institutions, however, lack the technical staff to implement open source software. Others may oppose it because of they fear security risks or high maintenance costs.”

- ³⁴ Archon and the Archivists' Toolkit are both open source applications. At this writing, the two groups are looking into the possibility of combining the best features and functionality of each application into a single application.
- ³⁵ Koha, <http://www.koha.org/>; Evergreen, <http://open-ils.org/>.
- ³⁶ Fedora, <http://fedora-commons.org/>; DSpace, <http://dspace.org/>; EPrints, <http://eprints.org/>; Greenstone, <http://www.greenstone.org/>.
- ³⁷ Yaco, "It's Complicated," 461: "Even archives with the resources to create custom, sophisticated encoding computer programs struggle with publishing issues."
- ³⁸ Refsnes Data, [w3schools.com](http://www.w3schools.com), "Displaying XML with XSLT," *XML Basic*, http://www.w3schools.com/Xml/xml_xsl.asp.
- ³⁹ Refsnes Data, [w3schools.com](http://www.w3schools.com), "Displaying XML with CSS," *XML Basic*, http://www.w3schools.com/Xml/xml_display.asp.
- ⁴⁰ Bob DuCharme, "Using XSLT to Deliver XML on Browsers," *bobdc.blog*, <http://www.snee.com/bobdc.blog/2008/09/using-xslt-to-deliver-xml-on-b.html>.
- ⁴¹ Syracuse University Library, Special Collections Research Center, <http://library.syr.edu/information/spcollections/>.
- ⁴² Swish-e, <http://swish-e.org/>.
- ⁴³ Google Inc., Google Site Search, <http://www.google.com/sitesearch/>.
- ⁴⁴ For more information on archival management systems, see the Archival Software wiki, <http://archivalsoftware.pbwiki.com/FrontPage>.
- ⁴⁵ This section adapted from Spiro, *Archival Management Software*.
- ⁴⁶ University of Minnesota Libraries, "Finding Aids in EAD," *Libraries Staff Wiki*, <https://wiki.lib.umn.edu/Staff/FindingAidsInEAD>.

6

Thresholds for Discovery: EAD Tag Analysis in ArchiveGrid, and Implications for Discovery Systems

By M. Bron, M. Proffitt and B. Washburn

The ArchiveGrid discovery system is made up in part of an aggregation of EAD (Encoded Archival Description) encoded finding aids from hundreds of contributing institutions. In creating the ArchiveGrid discovery interface, the OCLC Research project team has long wrestled with what we can reasonably do with the large (120,000+) corpus of EAD documents. This paper presents an analysis of the EAD documents (the largest analysis of EAD documents to date). The analysis is paired with an evaluation of how well the documents support various aspects of online discovery. The paper also establishes a framework for thresholds of completeness and consistency to evaluate the results. We find that, while the EAD standard and encoding practices have not offered support for all aspects of online discovery, especially in a large and heterogeneous aggregation of EAD documents, current trends suggest that the evolution of the EAD standard and the shift from retrospective conversion to new shared tools for improved encoding hold real promise for the future.

Introduction

ArchiveGrid is an aggregation of nearly two million archival material descriptions, including MARC records from WorldCat and finding aids harvested from the web. It is supported by OCLC Research as a corpus for experimentation and testing in text mining, data analysis, and discovery system applications and interfaces. Archival collections held by thousands of libraries, museums, historical societies, and archives are represented in ArchiveGrid. Although roughly 90% of what is in ArchiveGrid are MARC records, as of April 2013 OCLC Research had harvested 124,009 EAD encoded finding aids for inclusion in ArchiveGrid¹. This small segment of ArchiveGrid is important because EAD has been embraced by the archival community since its inception in the 1990s, and is supported by a range of tools designed specifically for archives, such as ArchivesSpace, Archivists' Toolkit, Archon, CALM, and others.

In creating the ArchiveGrid discovery interface, the project team has wrestled with what we can reasonably do with this corpus. For example, it would be useful to be able to sort by size of collection, however, this would require some level of confidence that the `<extent>` tag is both widely used and that the content of the tag would lend itself to sorting. Other examples of desired functionality include providing a means in the interface to limit a search to include only items that are in a certain genre (for example, photographs) or to limit a search by date. Again, we would need to have confidence that the metadata we have will actually support these features, and not leave out potentially important collections simply because

of the absence of certain tags. Specifically, we will consider how the variability of use of elements in finding aids affects discovery considering five different possible dimensions of a discovery system: search, browse, sort, limit, and display.

As a warning to the reader: this paper delves deeply into EAD elements and attributes and assumes at least a passing knowledge of the encoding standard. For those wishing to learn more about the definitions and structure, we recommend the official EAD website or the less official but highly readable and helpful EADiva site².

Related Work

The work that is the most closely related to our research was done by Katherine M. Wisser and Jackie Dean¹. In 2010 Wisser and Dean solicited EAD files repositories from institutions in order to "identify encoding behavior."³ In total, 108 repositories submitted up to 15 finding aids for the analysis; 1,136 finding aids comprise the entire sample. The formal results of their analysis will be published in the Fall 2013 edition of *American Archivist*. We are grateful to the authors for sharing their early work with us, and note with interest that in many cases, their analysis of EAD usage is quite similar. However, in some notable cases, the findings from the two samples diverge dramatically (see for example elements in `<archdesc>` above the `<dsc>` in Table 9). As noted by Wisser and Dean some of this variation can be attributed to the many different ways in which EAD files can be obtained. Wisser and Dean invited a limited contribution (12-15 finding aids) from a wide variety of repositories, including significant contributions

from institutions outside of the US; even though Wisser and Dean carefully articulated that results would be anonymized, there is some chance that the results were somewhat skewed by the process of selecting files for inclusion. By contrast, our data set was assembled by harvesting EAD documents from institutions directly, see below.

Contributing institutions have been motivated to contribute to ArchiveGrid primarily to share information about their collections, not their EAD practices. Additionally, ArchiveGrid is primarily constituted by repositories from the United States, with few institutions from Europe or elsewhere represented in the data set. Either or both of these key differences may account for divergence in findings between our work and that of Wisser and Dean.

The 2010 report, “Implications of MARC Tag Usage on Library Metadata Practices” focused on an analysis of the MARC standard as reflected in World-Cat5. Although the emphasis of the report was, similar to Dean and Wisser, meant to “inform community practice,” a secondary purpose was to draw conclusions about the suitability of MARC data for machine matching and processing, which is similar to our desire to identify functionality (and gaps in functionality) that exist in our current EAD corpus.

OCLC Research regularly harvests EAD documents from contributing institutions to update their representation in the ArchiveGrid index. The update cycle is roughly every six weeks. Institutions are contacted to obtain their permission to harvest and use the data in ArchiveGrid, and to identify the target URLs and rules for selection. For some contributors, the harvesting rules are simple: a directory listing or an HTML page is made available to our crawler, with every link leading to an EAD XML file on the contributor’s server. For other contributors we may make use of a website designed for human visitors, applying custom include and exclude rules to the URLs we find to select only links to EAD documents. Though OAI-PMH repositories and other more specialized harvesting protocols may be available at some contributor sites, we have seen little interest among contributors in their use, and currently we are using only standard HTTP GET requests for all the many hundreds of EAD document providers.

Maintaining the EAD harvesting operation continues to be a significant component of the ArchiveGrid support costs covered by OCLC Research.

Methods

Defining Thresholds

It is difficult to predefine thresholds for the level of usage of an element at which it becomes more or less useful for discovery. Is an element that is used 95% of the time still useful but one that is used 94% not? In this paper we consider the thresholds resulting from working with our sample of documents. We will use the terminology documents and finding aids interchangeably throughout the paper.

As an indicator for usage of an element we use the percentage of documents that contain the element at least once (% uniq). The nested nature of finding aids, however, influences the usage of elements as the absence of a parent element reduces the possibility of the occurrence of child elements. As an alternative indicator for usage we use the percentage of documents that contain an element in the sample of documents that contain the element’s parent element (% uniq in C).

Figure 1 shows how often the percentage of usage of an element falls into certain intervals. Note that we use relative usage (% uniq in C) here.

The distribution of element usage could be roughly divided into 4 groups: (i) usage between 0%-50% or low use; (ii) usage between 51%-80% or medium use; (iii) usage between 81%-95% or high use; (iv) usage between 96%-100% or complete use. Although we will use these levels as a reference point in this document, we do so with a recognition that correlating usage with discovery is an artificial construct. In the absence of a more effective approach, we are using these levels as an initial framework for discussion.

The absence of an element does not directly lead to a breakdown in a discovery system. It is more like a gradual decay of the effectiveness of a discovery system. But not all elements are created equally – in current archival discovery systems, we see a range of functionality that is offered, both in terms

of search and advanced search options, as well as sorting features, and results display. Within smaller aggregations, we might very well expect tag usage to be considerably more internally consistent than is the case in the ArchiveGrid aggregation. But in the case of ArchiveGrid and similar large

aggregations of finding aids, what functionality can be reasonably supported, given the present state of the data? What functionality can we offer with assurance, if we look only at elements that are in the high or complete categories?

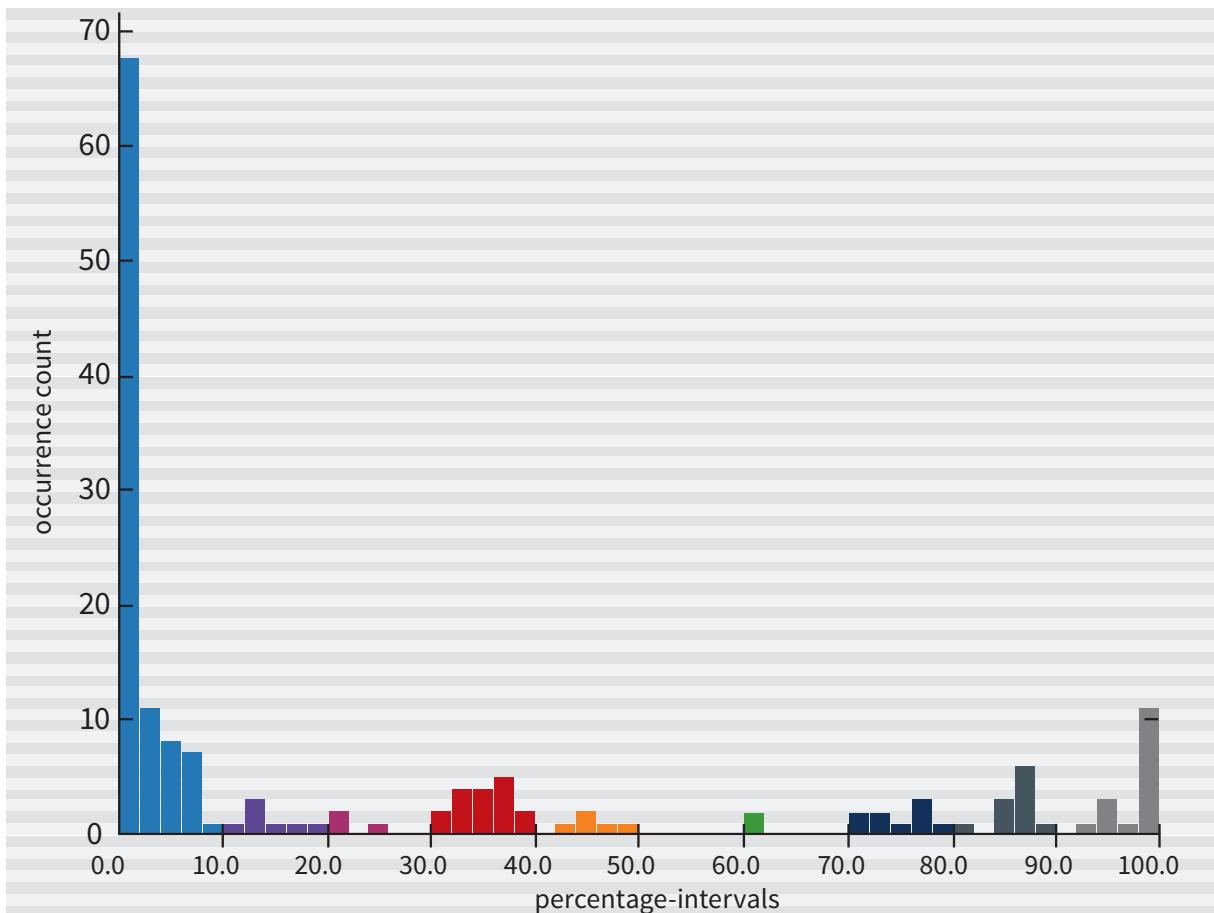


Figure 1: The distribution of percentage of element usage (% uniq in C). Elements are nested and the absence of a parent element influences the occurrence percentage of a child element. For this reason we use the number of element occurrences relative to the occurrences of the parent element (% uniq in C).

Counting Element Occurrences

Finding aids follow the Encoded Archival Description standard, which is a complex XML structure. As an example of the complexity of EAD in implementation, we found more than 26,000 paths in our 129,009 document set. To provide a starting point for obtaining element counts we recreated the many (but not all) tables of element, attribute, and value counts as presented in the report by Wisser et al.⁴ Each table was recreated by performing one or more XPath queries over the corpus of finding aids. In the discussion of our analysis we do not follow the same structure as in Wisser et al.⁴ as our focus is on implications of element usage on discovery and presentation. Where appropriate similarities and differences between element usage in our sample of finding aids and those used in Wisser et al.⁴ are reported.

In the rest of the paper we use the following notation in our tables: (i) N is the total number of occurrences of an element; (ii) N uniq is the number of documents in which the element occurs at least

once; (iii) $\frac{N_{\text{uniq}}}{S}$ is the percentage of documents in our sample of EAD documents (S= 124009) that

contain the element at least once; and (iv) $\frac{N_{\text{uniq}}}{n=...}$ is the percentage of documents that contain the element in the sample of documents (n=...) that contain a certain element. We will provide the size of each particular sample explicitly. For example, when considering the <eadheader> element that occurs in every document we get

$\frac{N_{\text{uniq}}}{n=124009}$, which is the same as $\frac{N_{\text{uniq}}}{S}$.

We use $\frac{N_{\text{uniq}}}{n=...}$ to indicate the percentage of documents that contain the element in the sample of documents that contain a certain element as collected by Wisser. In most cases the sample size will be all documents in Wisser's

sample, i.e., $\frac{N_{\text{uniq}}}{n=1136}$. Finally, we use diff to indicate the percentage point difference between the percentage Nuniq and Nuniqk, i.e., between Wisser's and our sample.

Dimensions for Analysis

Our analysis considered the following dimensions:

- search: all discovery systems have a keyword search function; many also include the ability to search by a particular field or element: examples include name, date, subject.
- browse: many discovery systems include the ability to browse finding aids: examples include browse by repository, browse by material type.
- results display: once a user has done a search, the results display will return portions of the finding aid to help with further evaluation: examples include title, dates, collection size.
- sort: once a user has done a search, they may have the option to reorder the results. Examples include: order by date, order by title, order by size.
- limit by: once a user has done a search, they may have the option to narrow the results to only include results that meet certain criteria. This may be done through presentation of facets: examples include limit by collections with digital material, limit by repository.

Current discovery interfaces

We reviewed a number of different discovery interfaces for finding aids in order to provide an overview of the type of search, browse, sort, limit, and display options that are generally available. Interfaces included are:

- the Online Archive of California (<http://www.oac.cdlib.org/>),
- the Northwest Digital Archive (<http://nwda.orbiscascade.org/>),
- Texas Archival Resources Online (<http://www.lib.utexas.edu/taro/index.html>),
- Arizona Archives Online (<http://www.azarchivesonline.org/xtf/search>),
- the Five Colleges Archives and Manuscripts Collection (<http://asteria.fivecolleges.edu/index.html>),
- the Rocky Mountain Online Archive (<http://rmoa.unm.edu/>),

- the Harvard Library's Online Archival Search Information System (http://oasis.lib.harvard.edu/oasis/deliver/home?_collection=oasis).

The interfaces we surveyed are very traditional in the capabilities they support — this is no doubt in part an outcome of the type of functionality that is supported in EAD 2002. In addition to assessing the suitability of the ArchiveGrid corpus for some general archival-specific discovery interfaces, we wanted to cast our net a little wider and speculate on how well EAD may meet the needs of emerging NextGen (or NowGen!) approaches to discovery that may not be represented in our interfaces surveyed, or supported by 2002 era EAD. Emerging discovery apparatus include:

Support for geo-locating archival locations, subjects of collected materials, and other elements, to serve map-based search interfaces. Examples of map-based discovery interfaces include:

- HistoryPin (<http://www.historypin.com/>),
- WhatWasThere (<http://www.whatwasthere.com/>),
- Historvius (<http://www.historvius.com/>)

Similarly, we see support for event-based retrieval, using timelines or similar devices, as an area in which discovery systems are evolving. Some examples include:

- SIMILE, example project timeline for Jewish History <http://simile.mit.edu/timeline/examples/religions/jewish-history.html>,
- Timeline view, Philippine Archives Collection, NARA <http://www.archives.gov/research/military/ww2/phillipine/timeline.html>
- Zagora Archaeological Project <http://www.powerhousemuseum.com/zagora/timeline/>

Analysis Details

We now take a closer look at which elements might drive each function, how the aggregated data fits this purpose both in terms of meeting our thresholds, and how well the content of key elements are fit for purpose. With each element, we've included a note about how they are used in ArchiveGrid and in other discovery systems.

Date

Our analysis shows use of <unitdate> within the high-level <did> as medium (72.64% — see Table 7); This makes <unitdate> values less than reliable for functions such as sort and limit by. Consider, for example, a scenario where a researcher is interested in material from the Second World War. Filtering by a date range between 1939-1945 will result in only those documents being presented that have a <unitdate> assigned in that period and may lead to the researcher missing potentially relevant documents. Alternatively, only those documents could be excluded that have a date outside of the indicated range. However, with a large amount of EADs missing a <unitdate> field this approach defeats the purpose of filtering.

Investing effort to bring this element closer to high or complete may be warranted; however, to support dimensions beyond just display, the content of the field or contents of the “normal” attribute must be easily parseable. When we look at the content of <unitdate>, we find a wide range of descriptive practices, some of which could pose problems for machine parsing to support use in indexing and retrieval.

Another issue involved in using the <unitdate> field is that it can be used in several places, e.g., on its own in the top level <did> or as a subelement of <unittitle>.

Comparing the usage of <unitdate> in our collection of EAD documents and that of Wisser, we find that it is one of the elements where we see the greatest divergence, i.e., Wisser's sample shows a usage of <unitdate> in the <did> of 97.00%.

In ArchiveGrid, dates are used in:

- search: they are keyword searchable
- display: with the collection title (when available) in brief displays

In other Archival Discovery Systems:

- search
- browse
- sort
- display

Extent

Our analysis shows use of `<extent>` within the high-level `<did>` as medium (70.43% — see Table 8); as with `<unitdate>`, the content of `<extent>` is quite varied and does not easily facilitate sorting, with values ranging from “miscellaneous artifacts” to “2 ceramic heads.” The syntax of the `<extent>` element (with attributes for `@encodinganalog`, `@type`, and `@unit`) does not currently lend itself to structuring data in a way that can be used for sorting without clear guidelines, tools to enforce appropriate encoding, and rigor on the part of institutions; retrospectively refitting to be utilized in sorting could be a daunting challenge for many institutions.

Many documents in the ArchiveGrid corpus have multiple `<extent>` statements, further complicating matters, as the system would need to decide which one to sort, for example. For display, including `<extent>` statements in order to help aid researchers in evaluating results seems fit to purpose.

In ArchiveGrid, extent is used in:

- search: extent values are keyword searchable
- display: presented in brief displays and separately in the display of individual collection descriptions

In other Archival Discovery Systems:

- sort
- display

Collection Title

Our analysis shows use of `<unittitle>` in the high-level `<did>` as complete (99.93% — see Table 7); this would suggest that it is suitable for all uses. However, for sorting and browsing, again, utility depends on the content of the element. If the content of the `<unittitle>` element is something generic like “Records” or “Papers” (in cases where perhaps the creator has been recorded separately in the origination element), then all functions may be less than ideal, but particularly sorting by title or creating browse lists.

Many interfaces either construct browse lists of collections titles, or allow users to sort results

by title, or search within titles. Not surprisingly, we found that the required `<filedesc>` element in the `<eadheader>` to be complete. Although our analysis did not include elements below `<filedesc>`, we can assume that the required `<titlestmt>` and its required child, `<titleproper>` will be similarly complete.

The fact that `<titlestmt>` is fully populated is good news for searching and display; however for sorting and constructing browse lists, we would need to have some assurance that the contents of `<titleproper>` are fit to purpose. This would be an opportunity for further evaluation, although a quick scan of the contents of `<titleproper>` encouragingly revealed that 42% of ArchiveGrid finding aids have a `@type` attribute with the value “filing”, which is rather remarkable as there is no specified list of values for type.

In ArchiveGrid, collection titles are used in:

- search: they are keyword searchable
- display: collection titles appear in brief search results

In other Archival Discovery Systems:

- sort
- browse
- display

Subject

Our analysis shows use of `<controlaccess>` as medium (72.89% — see Table 9); `<controlaccess>` is the parent element of both subject as well as other access points (such as `<corpname>`, `<genreform>`, `<geogname>`, and `<persname>`). Our analysis did not include drilling down to use of `<controlaccess>` subelements. (Given differences in library and archival practices, we would expect control of form and genre terms to be relatively high, and control of names and subjects to be relatively low.)

In ArchiveGrid, subjects are used in:

- limit by: we show `<controlaccess>` values for people, groups, places and topics as Result Overview facets for limiting a search result

In other Archival Discovery Systems:

- search
- browse

Material type

Researchers may wish to limit to or seek out material in a specific format, and our survey of discovery systems reveal that some systems support this functionality. Our analysis did not include the children of `<controlaccess>`, which includes `<genreform>`.

In ArchiveGrid, material type is used for:

- search: material types in `<genreform>` are keyword searchable

In other Archival Discovery Systems:

- search
- browse
- limit by

Names (personal or corporate)

Names can be found in multiple places — for the creator of a collection, is most logically found in `<origination>`, where both `<persname>` and `<corpname>` are child elements. The use of the origination tag is medium (87.78% — see Table 7); our analysis did not include evaluation of the use of `<persname>` and `<corpname>` in origination. Otherwise, personal and corporate names as access points may be found in `<controlaccess>` (see above).

Name elements occur ubiquitously in EAD version 2002, and our analysis did not include a detailed inventory of `<persname>` and `<corpname>` in the many places they can occur. A weakness of the distributed nature of names throughout EAD documents is that without detailed annotations and co-references, discovery systems only have a shallow understanding of names and their relationship to the collection and to one another. Discovery systems are not always able to differentiate between names when used in a creator context versus those covered in the description, which may show up as access points.

In ArchiveGrid, names are used for:

- search: names are keyword searchable
- limit by: names for people, groups and places appear in the Result Overview

In other Archival Discovery Systems:

- Used in search
- Used for limiting

Repository

The name of the repository is found in the high-level did in <repository>. Use of this element falls into the promising complete category (99.46%; see Table 7). However, a variety of practice is in play, with the name of the repository being embellished with <subarea> and <address> tags nested within <repository>. To avoid the difficulties in handling these variations across a range of contributing institutions, ArchiveGrid maintains a separate system to manage the form of the institution name for use in the system.

In ArchiveGrid, <repository> is not used as an access point, though ArchiveGrid's separately administered and controlled form of the repository name is used for search, browse, sort, limit and display.

In other Archival Discovery Systems, used in:

- browse
- limit by

Scope note, biographical note, abstract

Our analysis shows use of <scopecontent> as high (84.41% — see Table 9), while <bioghist> (70.42% — see Table 9) and <abstract> (79.20% — see Table 7) are medium; all three are suitable for search and for display in a results view, although they can be quite lengthy.

For search, its worth noting that the semantics of these elements are different, and may result in unexpected and false “relevance” for matches against descriptions in <bioghist> (about the person) and <scopecontent> and <abstract> (which may be more about the collection).

In ArchiveGrid, these notes are used in:

- search: notes are keyword-searchable
- display: <scopecontent> notes appear (in truncated form if lengthy) in brief search results

In other Archival Discovery Systems, used in:

- search
- display (in snippets or in their entirety)

Collections with digital content

Our analysis did not explore the use of <dao> or <daogrp> elements, which can be used in a variety of places in EAD 2002. Wisser and Dean found that <dao> is used in 7.7% and 9.3% of the documents in their sample, putting both into the low category (see Wisser, Table 26, <dao> elements). However, with growing interest in digitized materials from archival collections, identifying those materials is of increasing importance.

In ArchiveGrid, we provide no mechanism for searching or identifying collections with digital content.

In other Archival Discovery Systems:

- Limiting results to those with digital content
- Flagging collections with digital content

Future Work

In order to make EAD-encoded finding aids more well suited for use in discovery systems, the population of key elements will need to be moved closer to high or (ideally) complete. However, it is not only a matter of populating the elements, but ensuring that the data will reliably power key aspects of discovery systems. This will take concerted effort and tools, both on the part of individual institutions and groups.

In the analysis of “NextGen” discovery services, we noted the use of geolocation-based discovery. Although we would need to do further analysis in <controlaccess> to assess the usage for <geogname> in our document set, the current structure of the <geogname> element does not support geolocation functionality. However, as part of the redesign for EAD3, EAD is becoming more supportive of linked data and linked data structures. This may offer some hope for retrofitting EAD data to be more suited for the task of meeting map-based discovery requirements.

Likewise, the data we have on hand does not suggest good support for event-based discovery, which would draw on well-structured dates, geographic subject terms, and topical subject terms (such as “Battle of Alma” or “Great Depression”). Again, EAD 2002 does not support

the sort of encoding that would be necessary to serve event-based discovery, but EAD3 may provide more appropriate structures.

An Optimum Threshold for Discovery?

The picture for archival discovery and EAD is decidedly mixed. On the one hand, we have elements that are in high or even complete use. On the other hand, we have many elements that are necessary for discovery interfaces that are in medium use; and even with elements that are in high or complete use, the contents of those tags are not always fit to purpose. This can be at least partly explained by EAD's history. In the early days of EAD the focus was largely on moving finding aids from typescript to SGML and XML. Even with much attention given over to the development of institutional and consortial best practice guidelines and requirements, much work was done by brute force and often with little attention given to (or funds allocated for) making the data fit to the purpose of discovery.

Tag analyses such as the work described in this paper can help inform the development and implementation of the EAD schema (indeed the work done by Wisser and Dean was considered in the development of EAD3). But our analysis suggests that the standard has most of the elements and attributes needed to effectively support discovery; what's missing is agreement on and widespread application of best practices tied to supporting discovery.

So, is the container list half empty? If the archival community continues on its current path then the potential of the EAD format to support researchers or the public in discovery of material will remain underutilized. Minimally, collection descriptions that are below the thresholds for discovery will hinder their discovery efforts and maximally will remain hidden from view. Our paper provides suggestions for the elements where additional effort and investment are warranted to improve their utility for discovery systems. (We recognize that for some institutions, that additional effort may not be feasible or warranted; for their purposes they may find that HTML or PDF collection descriptions suffice.)

Or is the container list half full? Perhaps with emerging evidence about the corpus of EAD, continued discussion of practice, recognition of a need for greater functionality, and shared tools both to create new EAD documents and improve existing encoding, we can look forward to further increasing the effectiveness and efficiency of EAD encoding, and develop a practice of EAD encoding that pushes collection descriptions across the threshold of discovery.

Tables

Table 1: (Wisser Table 1): General statistics for EAD finding aids, using queries: /ead/*.

Element	N	N_uniq	% [N_uniq/S]	% [N_uniq/ n=124009]	% [(N_uniqK)/ n=1136]	diff
eadheader	124009	124009	100.00	100.00	100.00	0.00
archdesc	124009	124009	100.00	100.00	100.00	0.00
frontmatter	46115	46115	37.19	37.19	24.60	12.59
eadgrp	0	0	0.00	0.00	0.00	0.00
archdescgrp	0	0	0.00	0.00	0.00	0.00
dscgrp	0	0	0.00	0.00	0.00	0.00

Table 2: (Wisser Table 2): Elements used within eadheader, using query /ead/eadheader/*.

Element	N	N_uniq	% [N_uniq/S]	% [N_uniq/ n=124009]	% [(N_uniqK)/ n=1136]	diff
eadid	124445	124008	100.00	100.00	100.00	-0.00
filedesc	124009	124009	100.00	100.00	100.00	0.00
profiledesc	123103	123103	99.27	99.27	98.10	1.17
revisiondesc	42504	42501	34.27	34.27	32.70	1.57

Table 3: (Wisser Table 3) Attributes used with eadheader, using query //eadheader.

Element	N	N_uniq	% [N_uniq/S]	% [N_uniq/ n=124009]	% [(N_uniqK)/ n=1136]	diff
countryencoding	107412	107412	86.62	86.62	89.50	-2.88

Table 3 continued on next page

Element	N	N_uniq	% [N_uniq/S]	% [N_uniq/ n=124009]	% [(N_ uniqK)/ n=1136]	diff
dateencoding	107377	107377	86.59	86.59	88.20	-1.61
findaidstatus	42910	42910	34.60	34.60	27.80	6.80
langencoding	117641	117641	94.86	94.86	95.00	-0.14
repositoryencoding	106370	106370	85.78	85.78	87.80	-2.02
scriptencoding	95230	95230	76.79	76.79	77.60	-0.81

Table 4: (Wisser Table 4): Attributes used with eadid, using query //eadid.

Element	N	N_uniq	% [N_uniq/S]	% [N_uniq/ n=124009]	% [(N_ uniqK)/ n=1136]	diff
countrycode	108668	108667	87.63	87.63	94.30	-6.67
mainagencycode	105351	105350	84.95	84.95	92.60	-7.65
publicid	45758	45758	36.90	36.90	31.10	5.80
url	38020	38020	30.66	30.66	42.30	-11.64
urn	2312	2312	1.86	1.86	3.90	-2.04
identifier	57260	57260	46.17	46.17	49.30	-3.13

Table 5: (Wisser Table 8): Elements within frontmatter, using query /ead/frontmatter/*.

Element	N	N_uniq	% [N_uniq/S]	% [N_uniq/ n=46115]	% [(N_ uniqK)/ n=279]	diff
titlepage	45726	45726	36.87	99.16	92.80	6.36
div	190	190	0.15	0.41	2.20	-1.79

Table 6: (Wisser Table 9): Values for @level within archdesc, using query //archdesc/@level.

Element	N	N_uniq	% [N_uniq/S]	% [N_uniq/ n=124009]	% [(N_ uniqK)/ n=1,136]	diff
collection	116957	116957	94.31	94.31	90.90	3.41
fonds	135	135	0.11	0.11	4.80	-4.69
class	9	9	0.01	0.01	0.30	-0.29
recordgrp	433	433	0.35	0.35	1.40	-1.05
series	2394	2394	1.93	1.93	0.60	1.33
subfonds	49	49	0.04	0.04	0.30	-0.26
subgrp	526	526	0.42	0.42	1.00	-0.58
subseries	46	46	0.04	0.04	0.00	0.04
file	2446	2446	1.97	1.97	0.40	1.57
item	987	987	0.80	0.80	0.30	0.50
otherlevel	25	25	0.02	0.02	0.10	-0.08

Table 7: (Wisser Table 10): Elements within archdesc/did, using query /ead/archdesc/did/*.

Element	N	N_uniq	% [N_uniq/S]	% [N_uniq/ n=124009]	% [(N_ uniqK)/ n=1,136]	diff
abstract	102792	98218	79.20	79.20	86.60	-7.40
container	5447	3471	2.80	2.80	0.40	2.40
langmaterial	112938	109232	88.08	88.08	89.90	-1.82
materialspec	41	41	0.03	0.03	1.60	-1.57
origination	113684	108853	87.78	87.78	89.00	-1.22
physdesc	135126	122402	98.70	98.70	97.20	1.50
physloc	53564	45620	36.79	36.79	27.80	8.99

Table 7 continued on next page

Element	N	N_uniq	% [N_uniq/S]	% [N_uniq/ n=124009]	% [(N_ uniqK)/ n=1,136]	diff
repository	123343	123330	99.45	99.45	99.60	-0.15
unitdate	97247	90080	72.64	72.64	97.00	-24.36
unitid	119911	114898	92.65	92.65	90.10	2.55
unittitle	123959	123916	99.93	99.93	100.00	-0.07

Table 8: (Wisser Table 11): Elements within archdesc/did/physdesc, using query /ead/archdesc/did/physdesc/*.

Element	N	N_uniq	% [N_uniq/S]	% [N_uniq/ n=124009]	% [(N_ uniqK)/ n=1,136]	diff
dimensions	666	576	0.46	0.46	1.80	-1.34
extent	122613	87339	70.43	70.43	76.30	-5.87
physfacet	2000	1520	1.23	1.23	1.70	-0.47

Table 9: (Wisser Table 12): Elements within archdesc:above the dsc, using query /ead/archdesc/*.

Element	N	N_uniq	% [N_ uniq/S]	% [N_uniq/ n=124009]	% [(N_ uniqK)/ n=1,136]	diff
accessrestrict	55751	55579	44.82	44.82	86.20	-41.38
accruals	694	694	0.56	0.56	7.10	-6.54
acqinfo	40668	40451	32.62	32.62	68.00	-35.38
altformavail	2293	2289	1.85	1.85	12.70	-10.85
appraisal	4613	4602	3.71	3.71	4.80	-1.09
arrangement	40979	40627	32.76	32.76	65.50	-32.74
bibliography	4573	4083	3.29	3.29	10.10	-6.81

Table 9 continued on next page

Element	N	N_uniq	% [N_uniq/S]	% [N_uniq/ n=124009]	% [(N_uniqK)/ n=1,136]	diff
bioghist	89103	87333	70.42	70.42	87.30	-16.88
controlaccess	92124	90390	72.89	72.89	85.00	-12.11
custodhist	8375	8366	6.75	6.75	14.10	-7.35
descgrp	67684	56446	45.52	45.52	32.00	13.52
fileplan	50	44	0.04	0.04	0.60	-0.56
index	1231	656	0.53	0.53	1.20	-0.67
odd	9594	8145	6.57	6.57	9.70	-3.13
originalsloc	988	973	0.78	0.78	3.40	-2.62
otherfindaid	6529	6271	5.06	5.06	11.90	-6.84
phystech	900	897	0.72	0.72	4.20	-3.48
prefercite	49015	48989	39.50	39.50	85.40	-45.90
processinfo	27249	26623	21.47	21.47	0.00	21.47
relatedmaterial	23932	23676	19.09	19.09	40.30	-21.21
runner	10822	10822	8.73	8.73	1.10	7.63
scopecontent	105384	104670	84.41	84.41	93.40	-8.99
separatedmaterial	5789	5691	4.59	4.59	14.80	-10.21
userrestrict	41365	40749	32.86	32.86	68.30	-35.44

Table 10: Table 13: The inclusion of dsc in finding aids, using query //dsc.

Element	N	N_uniq	% [N_uniq/S]	% [N_uniq/ n=124009]	% [(N_uniqK)/ n=1,136]	diff
< dsc >	98663	94473	76.18	76.18	90.30	-14.12

Table 10 continued on next page

Element	N	N_uniq	% [N_uniq/S]	% [N_uniq/ n=124009]	% [(N_ uniqK)/ n=1,136]	diff
multiple < dsc > s	98663	2075	1.67	1.67	2.40	-0.73

Table 11: (Wisser Table 14): dsc type attributes, using query //dsc/@type.

Element	N	N_uniq	% [N_uniq/S]	% [N_uniq/ n=99023]	% [(N_ uniqK)/ n=1,105]	diff
analyticover	3156	3149	2.54	3.18	5.10	-1.92
combined	49205	49184	39.66	49.67	66.50	-16.83
in-depth	36433	35876	28.93	36.23	16.70	19.53
othertype	1725	1572	1.27	1.59	3.50	-1.91

Table 12: (Wisser Table 15): c-c12 tags, using query //c | //c01 | //c02 | //c03 | //c04 | //c05 | //c06 | //c07 | //c08 | //c09 | //c10 | //c11 | //c12.

Element	N	N_uniq	% [N_uniq/S]	% [N_uniq/ n=96548]	% [(N_ uniqK)/ n=1,053]	diff
c	4745698	14440	11.64	14.96	11.10	3.86
c01	1650659	78600	63.38	81.41	88.00	-6.59
c02	7432993	59217	47.75	61.33	72.50	-11.17
c03	6625963	29136	23.50	30.18	41.80	-11.62
c04	2927180	12819	10.34	13.28	20.60	-7.32
c05	1312217	5587	4.51	5.79	10.70	-4.91
c06	598647	2266	1.83	2.35	4.60	-2.25
c07	261648	922	0.74	0.95	2.00	-1.05

Table 12 continued on next page

Element	N	N_uniq	% [N_uniq/S]	% [N_uniq/ n=96548]	% [(N_ uniquK)/ n=1,053]	diff
c08	90401	331	0.27	0.34	0.70	-0.36
c09	21514	110	0.09	0.11	0.30	-0.19
c10	3578	36	0.03	0.04	0.10	-0.06
c11	823	7	0.01	0.01	0.00	0.01
c12	96	2	0.00	0.00	0.00	0.00

Table 13: (Wisser Table 16): Values for level attribute on c, c/@level, using query //c/@level | //c01/@level | //c02/@level | //c03/@level | //c04/@level | //c05/@level | //c06/@level | //c07/@level | //c08/@level | //c09/@level | //c10/@level | //c11/@level | //c12/@level.

Element	N	N_uniq	% [N_uniq/S]	% [N_uniq/ n=96548]	% [(N_ uniquK)/ n=1,053]	diff
collection	13489	4782	3.86	4.95	2.10	2.85
fonds	418	95	0.08	0.10	0.70	-0.60
class	63134	2113	1.70	2.19	1.20	0.99
recordgrp	1535	193	0.16	0.20	0.70	-0.50
series	398727	58480	47.16	60.57	77.70	-17.13
subfonds	3210	637	0.51	0.66	1.70	-1.04
subgrp	5573	430	0.35	0.45	3.10	-2.65
subseries	466366	16974	13.69	17.58	35.30	-17.72
file	11419524	36262	29.24	37.56	56.90	-19.34
item	3480272	20415	16.46	21.14	24.20	-3.06
otherlevel	368942	6225	5.02	6.45	9.10	-2.65

Table 14: (Wisser Table 17): c-c12/did elements, using query //c/did/* | //c01/did/* | //c02/did/* | //c03/did/* | //c04/did/* | //c05/did/* | //c06/did/* | //c07/did/* | //c08/did/* | //c09/did/* | //c10/did/* | //c11/did/* | //c12/did/*.

Element	N	N_uniq	% [N_uniq/S]	% [N_uniq/n=96548]	% [(N_uniqK)/n=1,053]	diff
abstract	1421043	3850	3.10	3.99	2.50	1.49
container	24951558	72377	58.36	74.96	82.50	-7.54
langmaterial	46798	1127	0.91	1.17	6.10	-4.93
materialspec	22870	106	0.09	0.11	1.30	-1.19
origination	1308346	4090	3.30	4.24	8.10	-3.86
physdesc	3967094	37749	30.44	39.10	54.40	-15.30
physloc	1343791	5978	4.82	6.19	5.80	0.39
repository	34923	29	0.02	0.03	0.30	-0.27
unitdate	9613593	41894	33.78	43.39	90.60	-47.21
unitid	7167784	31035	25.03	32.14	46.20	-14.06
unittitle	25228059	92888	74.90	96.21	98.90	-2.69

Table 15: (Wisser Table 18): c-c12/did/physdesc elements, using query //c/did/physdesc/* | //c01/did/physdesc/* | //c02/did/physdesc/* | //c03/did/physdesc/* | //c04/did/physdesc/* | //c05/did/physdesc/* | //c06/did/physdesc/* | //c07/did/physdesc/* | //c08/did/physdesc/* | //c09/did/physdesc/* | //c10/did/physdesc/* | //c11/did/physdesc/* | //c12/did/physdesc/*.

Element	N	N_uniq	% [N_uniq/S]	% [N_uniq/n=96548]	% [(N_uniqK)/n=1,053]	diff
dimensions	144079	1378	1.11	1.43	5.20	-3.77
extent	2401903	24495	19.75	25.37	36.60	-11.23
physfacet	164430	613	0.49	0.63	6.80	-6.17

Table 16: (Wisser Table 19): other elements found in c-c12, using query //c/* | //c01/* | //c02/* | //c03/* | //c04/* | //c05/* | //c06/* | //c07/* | //c08/* | //c09/* | //c10/* | //c11/* | //c12/*.

Element	N	N_uniq	% [N_uniq/S]	% [N_uniq/n=96548]	% [(N_uniqK)/n=1,053]	diff
accessrestrict	600069	4844	3.91	5.02	10.70	-5.68
accruals	12	11	0.01	0.01	0.00	0.01
acqinfo	68066	1477	1.19	1.53	4.50	-2.97
altformavail	252282	766	0.62	0.79	2.70	-1.91
appraisal	48	30	0.02	0.03	0.70	-0.67
arrangement	31945	5746	4.63	5.95	19.00	-13.05
bibliography	2067	48	0.04	0.05	1.50	-1.45
bioghist	12511	1132	0.91	1.17	4.60	-3.43
controlaccess	243134	2149	1.73	2.23	5.10	-2.87
custodhist	26224	181	0.15	0.19	2.20	-2.01
descgrp	2703	31	0.02	0.03	1.80	-1.77
index	386148	835	0.67	0.86	0.70	0.16
note	1180397	11265	9.08	11.67	20.30	-8.63
odd	242182	2663	2.15	2.76	7.20	-4.44
originalsloc	9959	211	0.17	0.22	1.00	-0.78
otherfindaid	1945	247	0.20	0.26	2.30	-2.04
phystech	8439	300	0.24	0.31	1.50	-1.19
prefercite	1995	264	0.21	0.27	0.10	0.17
processinfo	26332	1084	0.87	1.12	3.80	-2.68
relatedmaterial	16727	882	0.71	0.91	4.40	-3.49
runner	0	0	0.00	0.00	0.00	0.00

Table 16 continued on next page

Element	N	N_uniq	% [N_uniq/S]	% [N_uniq/ n=96548]	% [(N_uniqK)/ n=1,053]	diff
scopecontent	1852092	33483	27.00	34.68	61.30	-26.62
separatedmaterial	2784	208	0.17	0.22	0.00	0.22
userrestrict	2993	580	0.47	0.60	3.20	-2.60

Table 17: (Wisser Table 20): content tags in dsc, using query //dsc//*.

Element	N	N_uniq	% [N_uniq/S]	% [N_uniq/ n=96548]	% [(N_uniqK)/ n=1,053]	diff
corpname	373402	6082	4.90	6.30	8.40	-2.10
famname	3644	914	0.74	0.95	1.70	-0.75
function	996	53	0.04	0.05	0.00	0.05
genreform	351956	6988	5.64	7.24	5.10	2.14
geogname	1023771	6653	5.36	6.89	6.30	0.59
name	34339	380	0.31	0.39	1.40	-1.01
occupation	25284	285	0.23	0.30	0.40	-0.10
persname	2610548	11970	9.65	12.40	12.90	-0.50
subject	1239139	2419	1.95	2.51	4.70	-2.19

References

- 1 In April 2013, the ArchiveGrid index contained 1,632,246 MARC records, 119,984 EAD records, 61,551 HTML records, and 4,532 PDF records. The EAD count in the index is lower than the set of documents analyzed, to avoid duplicating their display for certain contributors who supply corresponding MARC records.
- 2 Library of Congress EAD Website: <http://www.loc.gov/ead/index.html>; EADiva: <http://eadiva.com/>.
- 3 E-mail to archives and archivists listserv, November 15, 2010.
- 4 Wisser, Katherine M, and Jackie Dean, EAD Tag Usage: Community analysis of the use of Encoded Archival Description elements, article submitted for publication in the American Archivist
- 5 Smith-Yoshimura, Karen, Catherine Argus, Timothy J. Dickey, Chew Chiat Naun, Lisa Rowlinson de Ortiz, and Hugh Taylor. 2010. Implications of MARC Tag Usage on Library Metadata Practices.

About the Authors

Marc Bron is a researcher at the Intelligent Systems Lab Amsterdam, where he is about to complete his PhD in Information Retrieval. His dissertation focused on improving accessibility to information stored in cultural heritage institutions by developing algorithms and interactive retrieval systems that support exploration and contextualization. During his PhD Marc has published over 20 papers at top tier conferences, journals, and workshops. His current research direction aims to develop new collaborative search methods for users of archival collections.

Bruce Washburn is a Consulting Software Engineer in OCLC Research. He provides software development support for OCLC Research initiatives and participates as a contributing team member on selected research projects. In addition, he provides software development support for selected OCLC Products and Services. At OCLC Washburn has been a part of the product teams that developed and maintain CAMIO, ArchiveGrid, the WorldCat Search API, and OAIster.

Merrilee Proffitt is a Senior Program Officer in OCLC Research. She provides project management skills and expert support to institutions represented within the OCLC Research Library Partnership. Merrilee has authored or co-authored articles, guidelines, and reports for a variety organizations and professional journals. She is frequently an invited speaker at international professional conferences and workshops on topics relating to digital libraries and special collections. Her current projects and interests include: archival description, increasing access to special collections, looking at developing better relationships between Wikipedia and cultural heritage institutions, and how Massively Open Online Courseware (MOOCs) may impact libraries. She is a member of the small but mighty ArchiveGrid team.

7

The Metadata is the Interface: Better Description for Better Discovery of Archives and Special Collections, Synthesized from User Studies

Jennifer Schaffner
Program Officer
OCLC Research

Introduction

Tim Ericson warned that user studies are important, but “they can also be a substitute for more direct action.”¹ We have strong evidence about how to improve discovery of archives and special collections, and we need to start somewhere. These days we are writing finding aids and cataloging collections largely to be discovered by search engines. People expect to find archives and special collections on the open Web using the same techniques they use to find other things, and they expect comprehensive results. Invisibility of archives, manuscripts and special collections may well have more to do with the metadata we create than with the interfaces we build. Now that we no longer control discovery, the metadata that we contribute is critical. In so many ways, the metadata is the interface.²

Structured metadata can be useful internally for collection management and public services, but is not always what users need most to discover primary sources, especially minimally-described collections and “hidden collections.”³ We understand archival standards for description and cataloging, but our users by and large don’t.⁴ Studies show that users often do not want to search for collections by provenance, for example, as important as this principle is for archival collections.⁵ One of several core competencies that special collections metadata librarians must have is “a keen understanding of users’ needs and preferences.”⁶ This is especially important now that discovery happens in multiple environments.⁷ Librarians and archivists need to manage archival collections by provenance, but also must describe what is in the collections for their users.

This essay—part of a series of OCLC Research projects to mobilize unique materials—synthesizes evidence of what descriptive information people say they need for research.⁸ As this literature review got underway, it soon became evident that we already know most of what we need to know in order to get started making changes.

In many contexts over many years, librarians and archivists have studied users with a wide variety of research methods: using surveys and questionnaires, examining statistics and citations,

testing usability of interfaces, studying information-seeking behaviors, listening to focus groups, creating personas, and questioning the efficacy of finding aid portals.⁹ The goal has always been to improve practices in order to help people—not just archivists and librarians—discover archival and rare materials.¹⁰ We still have gaps in our understanding, and comparing different kinds of studies across many years of work is like comparing apples and oranges. Nevertheless, the community has learned from these studies about obstacles between people and unique materials. While there is more to learn, let’s start now by adjusting our practices in order to disclose information about special collections and archives more effectively.

Librarians and Archivists as Gatekeepers

Users work increasingly on their own, while librarians and archivists have expected to mediate research. Most often people want to be autonomous and discover information about primary sources at the network level, not at the institutional level.¹¹ In an Ithaka study of higher education, Roger Schoenfeld and Ross Housewright learned that scholars consider *less* mediation in research and discovery a good thing:

[L]eading-edge libraries are beginning to change their priorities to match those of faculty and students. Still, the mismatch in views on the gateway function is a cause for further reflection: if librarians view this function as critical, but faculty in certain disciplines find it to be declining in importance, how can libraries, individually or collectively, strategically realign the services that support the gateway function?¹²

The more that discovery occurs directly via search engines, the greater the success of considerable efforts to expose “hidden collections.”

Over twenty-five years ago, Mary Jo Pugh challenged the myth of immortal and omniscient archivists, on whom users would rely for access to the contents of archival collections.¹³ Many studies of library catalogs and archival portals

have shown that these days most users start their search for information with Google or Wikipedia, and usually only come to libraries and archives for known items.¹⁴ Now the primary role in discovery is making the collections more visible and staying out of the way:

"Perceptions of a decline in dependence are probably unavoidable as services are increasingly provided remotely, and in some ways these shifting faculty attitudes can be viewed as a sign of library success. One can argue that the library is serving faculty well, providing them with a less mediated research workflow and greater ability to perform their work more quickly and effectively. In the process, however, they may be making their own role less visible."¹⁵

Perhaps ironically, goals to disclose descriptions online and to digitize primary resources have made special collections more visible and roles of archivists and librarians less visible. The more users do not need to consult archivists and librarians for searching, the more successful initiatives to improve description and discovery have been.

Users Search by Subjects and Keywords

Archivists and librarians have often focused on what collections are made up of (*Ofness*), while many users prefer to learn what collections are about (*Aboutness*).¹⁶ Studies report consistently that many users want to find information about contents of collections.¹⁷ For instance, Bill Maher analyzed reference letters to the University of Illinois archives in 1984-85 and found that over one third of the researchers inquired about subjects.¹⁸ One respondent in Jane Stevenson's testing of the UK's Archives Hub said, "I like the subject finder. I'm pleasantly surprised by it."¹⁹ In the most recent Northwest Digital Archives (NWDA) usability test, one user was enthusiastic to discover the subject section: "These will give me an idea of what this collection is about."²⁰ In a previous NWDA usability study, one person recommended controlled subject vocabulary and wanted subject terms linked to other collections and catalogs.²¹ Wendy Duff concluded, in more than one study, that users wanted "what is it about?" to appear at first glance.²² Louise

Gagnon-Arguin found 41% of queries in Québécois archives were for subjects or themes.²³ A study in 1976 of registration forms at the Michigan Historical Society showed that, "Roughly half of all users, regardless of preparation, began with a subject searching approach."²⁴ For thirty years, people have reported that they want to discover archival materials using subject information.

Content is more important than format.²⁵ Over fifteen years ago, Jackie Dooley cautioned that without subject access to records about archival collections, users are reduced to known-item searching.²⁶ An example of this surfaced in recent usability testing of WorldCat Local at the University of California (UC). Faculty and graduate student participants only searched UC's union catalog for known items, not for discovery, when they were working in their areas of expertise.²⁷ In an example of good intentions, the Online Archive of California (OAC) hoped to add subject searching until they learned that only "60% of the finding aids used controlled access tags."²⁸ Richard Szary and Lawrence Dowler recommended "direct indexing of the content of historical materials" to improve access.²⁹ For discovery, Aboutness is a very important element of description.

While users want to find subjects, they generally search using keyword techniques, rather than by using structured terminology. For example, research shows that keywords are important to historians searching for known items.³⁰ Likewise, NWDA usability testers observed that searches were completely unstructured.³¹ In November 2008, the French CALAMES project reported 40% frequency of searching full text, 34% by personal name, and 19% by various subject elements and attributes.³² Susan Hamburger's research yielded different proportions: 78% by keywords, 31% by names and 23% by subjects.³³ Chris Prom also found that users of the University of Illinois's electronic finding aids primarily used non-fielded keyword search terms, along with structured browsing.³⁴

Recent work addressed phrase-searching techniques. Phrase searches have been shown to be more effective than keyword searches when using search engines to find finding aids.³⁵ People don't search that way, however, according many studies, including OAC usability testing in 2001.³⁶ In another example, 8 out of 9 participants searched by

keywords—not phrases—in NWDA usability testing.³⁷ Kristina Southwell used statistical reports from search engines to demonstrate that the University of Oklahoma's Web pages for manuscripts were typically found through keyword searches, although some people used subject phrases, too.³⁸ Based on research with users at six major research libraries, Susan Hamburger recommended offering searching on both keywords and subject terms in catalog records and finding aids.³⁹

A wide range of research shows that keyword searching is important specifically for humanities scholars, who often search using name, place, title and discipline-specific terms.⁴⁰ Jihyun Kim examined EAD finding aids themselves, rather than users, precisely because historians and humanists search for primary sources by names of people and places. Kim reported that few finding aids used “controlled access headings.”⁴¹ Wendy Duff and Catherine Johnson interviewed ten historians and concluded they search names primarily because names are the easiest way into collections. Social historians desired subject access to collections: “There has to be a way that people can find things without having to know who generates them, so keywords will search across different provenances of things” (participant 6).⁴² Using keyword searching techniques for topics—such as farm women—can be problematic, because archives are organized primarily by the names of the creators, not the subject content of the collections.

There is no common understanding of what users and testers mean when they use words like “keyword,” “subject,” “known item,” “name,” “phrase” and “browse.” Without that common understanding, it is difficult to compare findings from separate studies. Is a keyword search technique in effect a subject search, from a user’s standpoint?⁴³ In one test, while Archives Hub participants favored subject searching, they were confused by a browse list composed of access points.⁴⁴ Do testers consider natural language searching to be keyword or subject searching, even if the user’s search includes names? Wendy Duff and Catherine Johnson, for example, consider a search by name keywords to be a known-item search.⁴⁵ Users do not always distinguish clearly between names and subjects. For instance, two of the participants in the Archives of American Art

usability study never found the Joseph Cornell collection because they searched by keywords rather than browsing an alphabetical list of collections.⁴⁶ RLG learned from focus groups that many participants combine keywords with names, subjects and dates.⁴⁷ Richard Lytle speculated that many kinds of searches might be disguised subject searches: “Requests for records by proper name, geographical area, date or form may conceal a subject request. Does the user really prefer to ask for documents by name...?”⁴⁸

Elsie Freeman memorably posited that good subject information is a large component of discovery experiences that are simple, elegant and intuitive.⁴⁹ Users want to search names by keyword, search for subjects by browsing, and browse by keyword or name, too. When it comes to using descriptive metadata to discover archival materials and special collections, users want it all. This is problematic because significant principles of archival theory and practice have been provenance and description of what the collection is made up of, its *Ofness*.⁵⁰ For users, research shows that important elements of description, especially minimum-level description, are keywords and terms that indicate *Aboutness*.

Users Expect Results Ranked by Relevance

While researchers consider it important to know the relative importance of collections, archivists and librarians rarely create metadata that can be used to rank relevance. In 1987, Avra Michelson argued that scholars using primary sources expected relevant results when doing research in exhaustive listings of collections.⁵¹ Over twenty years later, students at the University of Maryland were overwhelmed by large result sets retrieved by keyword searches; they expected relevance ranking of results such as that returned by Google and other search engines.⁵² Chris Prom learned—using the interface for the University of Illinois Archives—that hits sorted by provenance confused his participants, who were largely expecting search results to be ranked by relevance.⁵³ Andrea Rosenbusch concluded, after studying a dozen archival online databases, that, “The relevance of provenance as the main access point to records is becoming questionable...”⁵⁴

As it stands now, identifying relevant primary resources often requires educated guesswork. All of the participants in Sara Snyder's study at the Archives of American Art said that relevance ranking was essential, especially for large results.⁵⁵ On the other hand, in Jane Stevenson's Archives Hub study, relevance ranking of the results of a subject search puzzled some people, who then wanted to know how relevance worked and why some hits were more relevant.⁵⁶ When redesigning ArchiveGrid for improved usability, RLG determined that the order of search results was important, and relevance—not title—was the desired order.⁵⁷

Several tactics have been proposed that could indicate the relative importance of special collections in discovery experiences. Extent or physical description elements can be useful for some researchers trying to sort out relevance for themselves. ("Just one quick question. Does anybody understand what twelve metres of textual records means?" "Means two weeks in the archives!"⁵⁸) Andrea Rosenbusch suggested relevance could be leveraged from multi-level description, by restricting queries to top-level descriptions: "The aim [of ISAD(G)] is to enable users to identify fonds or even whole collections which have the highest relevancy to them."⁵⁹ Systems don't exist yet that use standards-based descriptions and extent statements in this way.

Search engine optimization strategies could leverage metadata for sorting search results by relevance. Based on keyword density analysis of UC Irvine's finding aids, Michelle Light advocated enhancing discovery by describing collections more strategically—by using more keywords and concepts than folder lists and material types.⁶⁰ Taking another tactic, the NWDA Working Group recommended experimenting with algorithms to combine use statistics with the frequency of index terms in order to produce relevance ranking like in search engines.⁶¹ Recommender systems for discovery of archival collections might provide indications of relevance. Improvements will require imaginative use of available Web 2.0 tools, such as tags for important collections on a topic, or "link paths" like those demonstrated in the Polar Bear project.⁶²

Over twenty years ago, Avra Michelson called for study of search questions, in order to identify successful patterns. Michelson recommended

subsequent improvements in our use of subject terms in description in order to improve what she called "retrieval capabilities."⁶³ More recently, Karen Markey has similarly suggested we would learn a great deal from studying people's search terms.⁶⁴ Many user studies for archives and special collections have focused on discovery within local systems designed for archival materials.⁶⁵ Now that close to 90% of searching behavior begins in search engines,⁶⁶ it is time to evaluate search behaviors at the network level, in order to develop descriptive strategies for ranking the relevance of primary resources.

Comprehensive Coverage

Increasingly, archivists and librarians are acutely aware that many researchers expect comprehensive coverage. A student in the Maryland study expected that "the universe of primary sources is a finite, absolute body of material that can and has been already labeled and categorized for him."⁶⁷ Chris Prom, too, learned that many inexperienced users assume that everything is available.⁶⁸ Jane Stevenson's study with Archives Hub confirmed that some people assumed their search results were comprehensive.⁶⁹ In a usability study of the Lilly Library's Web site, Erika Dowell found that users doubted the utility of the online catalog when cautioned (responsibly) that only 45% of the Lilly's holdings were included.⁷⁰ In a related study in UK museums, the Research Information Network concluded that "what researchers need above all is online access to the records in museum and collection databases to be provided as quickly as possible, whatever the perceived imperfections or gaps in the records."⁷¹

Some researchers have substantiated a "More Product, Less Processing" (MPLP) approach to description and digitization.⁷² At the University of Wisconsin, Joshua Ranger and Krystyna Matusiak are experimenting with a less expensive, streamlined process for mass digitization of archival collections. The students they interviewed all preferred more description, not less. However, when the comparative costs of full and minimal records were explained, all of the participants said streamlined description was preferred: "Better than not having it at all."⁷³ The American Heritage Center at the University of Wyoming

surveyed 600 respondents for their satisfaction with minimal processing. Asked to rank archival priorities, respondents most often chose “putting more resources into creating basic descriptions for all collections.”⁷⁴ The MPLP approach matches users’ acceptance of minimum-level description because they would like to discover more materials online. When such decisions are made to describe more collections at a minimal level, archivists and librarians need to indicate, however briefly, what the contents of collections are about.

Users Know How to Scan and Scroll

Archivists and librarians have worried about confusing users by presenting different kinds and amounts of metadata, while users mostly care more about what is in the collections.⁷⁵ Diverging desires for less or more information appear often in user research. One example of this variety is that Maryland students were able to use long finding aids easily, despite difficulties navigating specific tools to discover them. On the other hand, in the same test one Maryland student reported that “too much information hindered the reading of the display.”⁷⁶ In another contradictory example, some of Wendy Duff’s participants preferred to see shorter abstracts and scope-and-content notes, disparaging long biographical notes (ranked 16th in order of preference) or administrative histories (ranked 23rd). However, a different participant in the same study said anyone interested in long notes can scroll down through the display.⁷⁷

There are many more examples of preferences for both brief and for full displays that support arguments for both minimal and full description. In the RLG rapid iterative interface testing, most participants found a brief scope-and-content note most useful.⁷⁸ On the other hand, studies also report that users know how to skim long pages of records, when they want to. In Jane Lee’s usability testing for the 2008 OAC redesign, she noted that they chose a long display format for search results because, as one participant said, “it’s nice to have a little more information” when browsing.⁷⁹ In NWDA usability testing, “the majority of the users started the search for information by skimming or scrolling through the finding aid page; most said they weren’t reading for content, rather were scanning for key terms.”⁸⁰ Genealogists in RLG’s

Archival Resources focus groups preferred to scroll through large result sets.⁸¹ These conflicting recommendations suggest that minimum description may come as a relief to some users, but others prefer a full description. If a collection is fortunate enough to have full description, it will not necessarily get in a user’s way when scanning and scrolling through results. Users support concise minimum-level description, which can also be effective for discovery when it is done well.

Users’ Lack of Awareness

Archivists and librarians have created catalogs and portals, but many users don’t use them or don’t know they exist. “The greatest barrier to use is lack of awareness.”⁸² Often it isn’t easy to find rare and unique library and archival materials because successful discovery currently requires people to understand what they are looking for and how to find it.⁸³ Karen Markey says rare and unique materials are invisible: “Thousands of special collections that make up the invisible Web feature their own unique search engines because their content is not accessible via general Web search engines.”⁸⁴ Louise Gagnon-Arguin concluded that the key to access is fragile in the context of electronic information.⁸⁵ In order to find primary resources, people need to know too much about how collections are described and where those descriptions are lodged. That isn’t good enough.

Catalogs don’t seem to do the trick. “It is unlikely that researchers approach doing research by looking for a tool for doing research.”⁸⁶ In the RIN user study of UK museums, “most researchers are unaware of the online catalogues...”⁸⁷ Beth Yakel, Susan Hamburger, Bill Maher and others have found that the majority of researchers do not use utilities such as ArchivesUSA, OCLC, RLIN or NUCMC.⁸⁸ While a percentage of people in Kristina Southwell’s Oklahoma survey found manuscript collections by searching the Web, only one person (0.4% in 230 responses!) used RLIN’s AMC.⁸⁹ Southwell was surprised that only 11.3% of respondents discovered manuscript collections using the online catalog, leading her to wonder about the considerable investment creating MARC records. 17.9% found collections from html finding aids on the Web site, 25.1% used footnotes and bibliographies, while another

8.6% used a published guide to the repository for know-item discovery. Users may search on the open Web, but often they find archives indirectly.

So are finding aids best for discovery? Bill Maher questioned out loud our tacit belief that better finding aids will automatically result in better access.⁹⁰ Most participants in Jane Stevenson's study of the Archives Hub "did not mention any kind of cross-searching networks."⁹¹ Kathleen Feeney concluded that "electronic finding aids may not be well suited to serve as pointers to archival collections," based on her 1999 study of retrieval of full-text finding aids by search engines. Feeney concluded that "MARC records remain a more valuable and reliable means of locating archival resources" because of problems with relevance rankings at the network level.⁹²

For successful discovery, what are the lessons learned about our choices for description? Early on, Rob Spindler and Richard Pearce-Moses argued for adapting description methods—based on their case study with Arizona State University patrons—expressly to improve comprehension of AMC records in an integrated online environment.⁹³ More recently, Michelle Light suggested strategies to adapt description than can "enhance retrieval possibilities" at the network level: use long-tail keywords, repeat names and keywords (bending rules for description), put the most important content at the top, say more with less.⁹⁴ If students now don't look in library catalogs or archival portals for primary materials, why spend resources that way? Let's put the right descriptive metadata in the right places.

Conclusion

I argue that some thirty years of user studies teach that *Aboutness* and relevance matter most for discovery of special collections, especially now that discovery happens elsewhere.⁹⁵ Unfortunately, there is a gap between the expectations of users and historical descriptive practices in archives and special collections. Changes must be made to description because researchers rarely look in library catalogs or archival portals for primary resources. These changes are even more important for collections that have been selected for minimal processing and description. Ensuring that "hidden collections" can be discovered requires appropriate description, not just expert processing, cataloging and cross-searching networks. It would be heartbreaking if special collections and archives remained invisible because they might not have the kinds of metadata that can easily be discovered by users on the open Web.

In a 1986 article on "The Use of User Studies," Bill Maher described archivists with instincts about how their collections are used—but without data to support their instincts—as "working in the dark."⁹⁶ Since then, research demonstrates recurring observations of users' needs and preferences when they search for special collections and archives. Over time, users have adapted their research tactics: from discovery only by visiting repositories and by consulting printed catalogs or guides, then discovery using online catalogs and portals, and now discovery on the Web. All along, user studies have demonstrated that descriptive metadata indicating *Aboutness* and relevance matters significantly for discovery. Twenty years later, we are not working in the dark any more.

Notes

- ¹ Timothy L. Ericson, "Preoccupied with Our Own Gardens: Outreach and Archives," *Archivaria* 31 (winter 1990-1991): 120. Richard Cox expressed similar concerns: "User studies and citation analyses float isolated from any practical applications." See "Researching Archival Reference as an Information Function: Observations on Needs and Opportunities," *RQ* 31, no.3 (spring 1992): 393.
- ² This aphorism was coined by Arnold Arcolio.
- ³ Judith M. Panitch conducted the first survey that catalyzed the community around the significance of "hidden collections": *Special Collections in ARL Libraries* (Washington DC: Association of Research Libraries, 2001). Published online at: http://www.arl.org/bm~doc/spec_colls_in_arl.pdf. See also Barbara M. Jones, "Hidden Collections, Scholarly Barriers" (June 6, 2003). Published online at: <http://www.arl.org/bm~doc/hiddencollswhitepaperjun6.pdf>.
- ⁴ I have borrowed this idea from Steve Hensen.
- ⁵ Wendy M. Duff and Catherine A. Johnson, "Accidentally Found on Purpose: Information-Seeking Behavior of Historians in Archives," *Library Quarterly* 72, no. 4 (2002): 477, footnote 5.
- ⁶ ACRL Rare Books and Manuscripts Section Task Force on Core Competencies for Special Collections Professionals, "RBMS Competencies for Special Collections Professionals," *College and Research Libraries News*, 69, no. 10 (November 2008): 622.
- ⁷ Lorcan Dempsey, "The Library Catalogue in the New Discovery Environment: Some Thoughts," *Ariadne* 48 (July 2006). Published online at: <http://www.ariadne.ac.uk/issue48/dempsey/>.
- ⁸ I focus attention on metadata for discovery, not on research methodologies, nor on interfaces, nor on interface terminology, nor on information-seeking behavior writ large, nor on what users do with what they find, much as these topics are fascinating and closely related to the relationship between description and discovery.
- ⁹ Many more studies have been done than are reported: "In recent years archive services in many parts of the world have undertaken surveys of their users. Valuable work has been done, but most of it suffers from lack of availability—many surveys remain unpublished and largely inaccessible..." Geoffrey Yeo, "Understanding Users and Use: A Market Segmentation Approach," *Journal of the Society of Archivists* 26, no. 1 (April 2005): 43-44.
- ¹⁰ Lawrence Dowler called for a national study of the use and users of archives to inform archival practices: "The Role of Use in Defining Archival Practice and Principles," *American Archivist* 51 (winter/spring 1988): 74-95. A stellar example is the National Council on Archives in the UK, which established standards requiring repositories to give users a choice in access to content, and to open up that content to a wider range of users: David Mander, ed., *Standard for Access to Archives* (National Council on Archives [UK]: Public Services Quality Group, 1st ed. 1999, 2nd ed. 2003, 3rd ed. 2008), 37, 46. The Auditor General of Canada is taking a similar tactic: Jean Dryden, "Do We Care What Users Want? Evaluating User Satisfaction and the LibQUAL+™ Experience," *Journal of Archival Organization* 4, no. 4 (2004): 83-84.
- ¹¹ Louise Gagnon-Arguin applied the term "autonomy" to users doing research: "Les questions de recherché comme matériau d'études des usagers en vue du traitement des archives," *Archivaria* 46 (winter 1998): 86-102.

- ¹² Ross Housewright and Roger Schoenfeld, *Ithaka's 2006 Studies of Key Stakeholders in the Digital Transformation in Higher Education* (August 2008). Published online at: <http://www.ithaka.org/research/Ithakas%202006%20Studies%20of%20Key%20Stakeholders%20in%20the%20Digital%20Transformation%20in%20Higher%20Education.pdf>.
- ¹³ Mary Jo Pugh, "The Illusion of Omniscience: Subject Access and the Reference Archivist," *American Archivist* 45, no. 1 (winter 1982): 38. Dowler also targeted mediation with humor: "To archivists, mediation has generally meant the satisfying vision of the erudite archivist leading a grateful scholar by the hand through the uncharted forest of records to precisely the right material." "The Role of Use in Defining Archival Practice and Principles," *American Archivist* 51 (winter/spring 1988): 82.
- ¹⁴ For example, Jennifer Ward, Pam Mojfeld, and Steve Shadle, "WorldCat Local at the University of Washington Libraries," *Library Technology Reports* 44, no. 6 (August /September 2008). Published online at: <http://www.techsource.ala.org/ltr/worldcat-local-at-the-univ-of-washington-libraries.html>.
- ¹⁵ Ross Housewright and Roger Schoenfeld, *Ithaka's 2006 Studies of Key Stakeholders in the Digital Transformation in Higher Education* (August 2008). Published online at: <http://www.ithaka.org/research/Ithakas%202006%20Studies%20of%20Key%20Stakeholders%20in%20the%20Digital%20Transformation%20in%20Higher%20Education.pdf>.
- ¹⁶ Sarah Shatforth (Layne) first applied the concepts Ofness and Aboutness to cataloging: "Analyzing the Subject of a Picture: a Theoretical Approach," *Cataloging & Classification Quarterly* 6 (1986): 39-62. Users and user testers have not used the terms.
- ¹⁷ For example, a recent study of users of online library catalogs found that users desired "more subject information" second only to more online content: Karen Calhoun, *Online Catalogs: What Users and Librarians Want* (Dublin, Ohio: OCLC, 2009), 13, 43, 47. Published online at: <http://www.oclc.org/reports/onlinecatalogs/fullreport.pdf>.
- ¹⁸ William J. Maher, "Use of User Studies," *Midwestern Archivist* 11 (1986): 21.
- ¹⁹ Jane Stevenson, "'What Happens If I Click on This?' Experiences of the Archives Hub," *Ariadne* 57 (October 2008). Published online at: <http://www.ariadne.ac.uk/issue57/stevenson/>.
- ²⁰ Northwest Digital Archives (NWDA), "Executive Summary: Usability Test #5" (November 6, 2008), 6. Published online at: <http://www.orbiscascade.org/index/usability-design-working-group-reports>.
- ²¹ NWDA, "Executive Summary: Usability Testing Round 4" (March 12, 2008), 6, 7.
- ²² Wendy Duff, and Penka Stoyanova, "Transforming the Crazy Quilt: Archival Displays from the Users Point of View," *Archivaria* 45 (spring 1998): 63. "Many of the historians interviewed, particularly the social historians, suggested that subject indexes, keyword searches, or identification of themes would help them with their research": Wendy Duff and Catherine Johnson, "Accidentally Found on Purpose: Information-Seeking Behavior of Historians in Archives," *Library Quarterly* 72, no. 4 (2002): 490.
- ²³ 21% were by name, and 33% were by the type of document. Louise Gagnon-Arquin, "Les questions de recherché comme matériau d'études des usagers en vue du traitement des archives," *Archivaria* 46 (winter 1998): 92.
- ²⁴ Mary Jo Pugh, "The Illusion of Omniscience: Subject Access and the Reference Archivist," *American Archivist* 45, no. 1 (winter 1982): 40.

- ²⁵ Dana Chisnell, *Report: RLG Archival Resources Service Redesign Focus Groups* (Usability Works, 2 December 2004), 18. For similar results from recent user studies in a library context, see Karen Calhoun et al., *Online Catalogs: What Users and Librarians Want*, (Dublin, Ohio: OCLC, 2009). Published online at: <http://www.oclc.org/reports/onlinecatalogs/fullreport.pdf>. See also 2004 *Information Format Trends: Content, Not Containers* (Dublin, Ohio: OCLC, 2004). Published online at: http://www5.oclc.org/downloads/community/2004infotrends_content.pdf.
- ²⁶ Jackie M. Dooley, "Subject Indexing in Context," *American Archivist* 55 (spring 1992): 351. Richard Smiraglia proposed methodology for subject access in "Subject Access to Archival Materials Using LCSH," *Cataloging & Classification Quarterly* 11, no. 3/4 (1990): 63-90.
- ²⁷ Arnold Arcolio and Felicia Poe, *WorldCat Local at the University of California: Usability Testing, Round One* (Spring 2008), 4.
- ²⁸ Michelle Light, "The Endangerment of Trees," EAD @ 10 (August 31, 2008): 1, forthcoming. Published online at: <http://www.archivists.org/publications/proceedings/EAD@10/Light-EAD@10.pdf>.
- ²⁹ Lawrence Dowler quotes Richard Szary saying that provenance information and content indexing are each half of an approach to improving description and retrieval: "The Role of Use in Defining Archival Practice and Principles," *American Archivist* 51 (winter/spring 1988): 83.
- ³⁰ Wendy M. Duff and C.A. Johnson, "Accidentally Found on Purpose: Information-Seeking Behavior of Historians in Archives," *Library Quarterly* 72, no. 4 (2002): 472-496.
- ³¹ NWDA, "Executive Summary: Usability Testing Round 4" (March 12, 2008), 6, 7.
- ³² E-mail correspondence 12 November 2008.
- ³³ Here more than one answer was possible, hence a total of over 100%. Susan Hamburger, "How Researchers Search for Manuscript and Archival Collections," *Journal of Archival Organization* 2, no. 1/2 (2004): 84.
- ³⁴ Christopher Prom, "User Interactions with Electronic Finding Aids in a Controlled Setting," *American Archivist* 67, no. 2, (fall/winter 2004): 263.
- ³⁵ Helen Tibbo and Lokman I. Meho, "Finding Finding Aids on the World Wide Web," *American Archivist* 64, no. 1 (spring/summer): 61.
- ³⁶ California Digital Library, *OAC Usability Test Summary* (October 2001), 1.
- ³⁷ Northwest Digital Archives, "Executive Summary: Usability Test #5" (November 6, 2008), 2. Published online at: <http://www.orbiscascade.org/index/usability-design-working-group-reports>.
- ³⁸ Kristina Southwell, "How Researchers Learn of Manuscript Resources in Western History Collections," *Archival Issues* 26, no. 2 (2002):99.
- ³⁹ Susan Hamburger, "How Researchers Search for Manuscript and Archival Collections," *Journal of Archival Organization* 2, no. 1/2 (2004): 91.
- ⁴⁰ Carole L. Palmer, Lauren C. Teffieu and Carrie M. Pirmann, *Scholarly Information Practices in the Online Environment: Themes from the Literature and Implications for Library Service Development* (Dublin, Ohio: OCLC Research, 2009): 10. Published online at: <http://www.oclc.org/programs/publications/reports/2009-02.pdf>.
- ⁴¹ Jihyun Kim, "EAD Encoding and Display: A Content Analysis," *Journal of Archival Organization* 2, no. 3 (2004): 50.
- ⁴² Wendy Duff and Catherine Johnson, "Accidentally Found on Purpose: Information-Seeking Behavior of Historians in Archives," *Library Quarterly* 72, no. 4 (2002): 493-494.

- ⁴³ Neither users nor testers ever suggest that Library of Congress Subject Headings (LCSH) are what they mean by “subjects;” LCSH is never even mentioned by testers or users.
- ⁴⁴ Jane Stevenson, “What Happens If I Click on This?: Experiences of the Archives Hub,” *Ariadne* 57 (October 2008). Published online at: <http://www.ariadne.ac.uk/issue57/stevenson/>.
- ⁴⁵ Wendy Duff and Catherine Johnson, “Accidentally Found on Purpose: Information-Seeking Behavior of Historians in Archives,” *Library Quarterly* 72, no. 4 (2002): 485.
- ⁴⁶ Sara Snyder et al., [Archives of American Art] Web Usability Study Report: Round 1, Subject Experts, (April 2008).
- ⁴⁷ Dana Chisnell, *Report: RLG Archival Resources Service Redesign Focus Groups* (Usability Works, 2 December 2004), 14-15.
- ⁴⁸ Richard Lytle, “Intellectual Access to Archives: I. Provenance and Content Indexing Methods of Subject Retrieval,” *American Archivist* 43 (winter 1980): 68.
- ⁴⁹ Elsie T. Freeman, “In the Eye of the Beholder: Archives Administration from the User’s Point of View,” *American Archivist* 47, no. 2 (spring 1984): 116. Wendy Duff and Catherine Johnson added the requirement that systems be intuitive: “Accidentally Found on Purpose: Information-Seeking Behavior of Historians in Archives,” *Library Quarterly* 72, no. 4 (2002): 472.
- ⁵⁰ “[R]ecords in an archives are not records about an activity, but records of an activity.”[Emphasis added.] For a concise summary of how historians use provenance information for subject queries, see Wendy Duff and Catherine A. Johnson, “Accidentally Found on Purpose: Information-Seeking Behavior of Historians in Archives,” *Library Quarterly* 72, no. 4 (2002): 477, including footnote 5.
- ⁵¹ Avra Michelson, “Description and Reference in the Age of Automation,” *American Archivist* 50 (spring 1987): 199.
- ⁵² Joanne Archer, Ann M. Hanlon and Jennie Levine, “Investigating Primary Source Literacy,” *forthcoming [in Journal of Academic Librarianship* 35, no. 5 (September 2009): 13. Archer, Hanlon and Levine report students’ searches, using keywords, returned thousands of results and very few students refined or limited their results.
- ⁵³ Christopher Prom, “User Interactions with Electronic Finding Aids in a Controlled Setting,” *American Archivist* 67, no. 2, (fall/winter 2004): 254.
- ⁵⁴ Andrea Rosenbusch, “Are Our Users Being Served? A Report on Online Archival Databases,” *Archives and Manuscripts* 29, no. 1 (2001): 50.
- ⁵⁵ Sara Snyder et al., [Archives of American Art] Web Usability Study Report: Round 1, Subject Experts (April 2008).
- ⁵⁶ Jane Stevenson, “What Happens If I Click on This?: Experiences of the Archives Hub,” *Ariadne* 57 (October 2008). Published online at: <http://www.ariadne.ac.uk/issue57/stevenson/>.
- ⁵⁷ Dana Chisnell, *ArchiveGrid Rapid Iterative Testing Report* (Usability Works, 28 February 2006), 12, 17.
- ⁵⁸ Wendy Duff, and Penka Stoyanova, “Transforming the Crazy Quilt: Archival Displays from the Users Point of View,” *Archivaria* 45 (spring 1998): 57.
- ⁵⁹ Andrea Rosenbusch, “Are Our Users Being Served? A Report on Online Archival Databases,” *Archives and Manuscripts* 29, no. 1 (2001): 49.
- ⁶⁰ Michelle Light, “The Endangerment of Trees,” *EAD @ 10* (August 31, 2008): 2; *forthcoming*. Published online at: <http://www.archivists.org/publications/proceedings/EAD@10/Light-EAD@10.pdf>.
- ⁶¹ NWDA, “Executive Summary: Usability Testing Round 4” (March 12, 2008), 4.

- ⁶² Elizabeth Yakel et al., "Creating the Next Generation of Archival Finding Aids," *D-Lib* 13, no. 5/6 (May/June 2007). Published online at: <http://www.dlib.org/dlib/may07/yakel/05yakel.html>. For an intriguing example of possibilities for one specific finding aid, see Magia Ghetu Krause and Elizabeth Yakel, "Interaction in Virtual Archives: The Polar Bear Expedition Digital Collections Next Generation Finding Aid," *American Archivist* 70, no. 2 (fall/winter 2007): 282-314.
- ⁶³ Avra Michelson, "Description and Reference in the Age of Automation," *American Archivist* 50 (spring 1987): 192, 200.
- ⁶⁴ Karen Markey, "Twenty-Five Years of End-User Searching, Part 2: Future Research Directions," *Journal of the American Society for Information Science and Technology*, 58 no. 8 (2007): 1126.
- ⁶⁵ An outstanding counter-example is Jody DeRidder's use of sitemaps and links in static html pages to increase dramatically discovery of finding aids at the University of Tennessee: "Googlizing a Digital Library," *The Code4Lib Journal* 2 (March 24, 2008). Published online at: <http://journal.code4lib.org/articles/43/>.
- ⁶⁶ Cathy De Rosa et al., *College Students' Perceptions of Libraries and Information Resources* (Dublin, Ohio: OCLC, 2006), 1-1. Published online at: <http://www.oclc.org/reports/pdfs/studentperceptions.pdf>.
- ⁶⁷ Joanne Archer, Ann M. Hanlon and Jennie Levine, "Investigating Primary Source Literacy," *forthcoming [in Journal of Academic Librarianship* 35, no. 5 (September 2009)]: 11.
- ⁶⁸ Christopher Prom, "User Interactions with Electronic Finding Aids in a Controlled Setting," *American Archivist* 67, no. 2, (fall/winter 2004): 247.
- ⁶⁹ Jane Stevenson, "'What Happens If I Click on This?': Experiences of the Archives Hub," *Ariadne* 57 (October 2008). Published online at: <http://www.ariadne.ac.uk/issue57/stevenson/>.
- ⁷⁰ Erika Dowell, "Web Site Usability for Rare Book and Manuscript Libraries," *RBM* 9, no. 2 (fall 2008): 175, 179.
- ⁷¹ Research Information Network, *Discovering Physical Objects: Meeting Researchers' Needs* (London, October 2008): 5. Published online at: <http://www.rin.ac.uk/objects>.
- ⁷² Mark A. Greene and Dennis Meissner, "More Product, Less Process: Revamping Traditional Archival Processing," *American Archivist* 68, no. 2 (fall/winter 2005): 208-263.
- ⁷³ Joshua Ranger, "Mass Digitization of Archival Manuscripts," (2008). Published online at: <http://www.pacsclsurvey.org/documents/ranger/04ranger.ppt>.
- ⁷⁴ Mark Greene, "[Draft Grant Report to NHPRC:] Basic Processing User Survey Overview," (2008).
- ⁷⁵ For example, RLG learned this when MARC records were mixed with archival finding aids. See Dana Chisnell, *Report: RLG Archival Resources Service Redesign Focus Groups*, (Usability Works, 2 December 2004), 18.
- ⁷⁶ Joanne Archer, Ann M. Hanlon and Jennie Levine, "Investigating Primary Source Literacy," *forthcoming [in Journal of Academic Librarianship* 35, no. 5 (September 2009)]: 17.
- ⁷⁷ Wendy Duff, and Penka Stoyanova, "Transforming the Crazy Quilt: Archival Displays from the Users Point of View," *Archivaria* 45 (spring 1998): 52-54, 59.
- ⁷⁸ Dana Chisnell, *ArchiveGrid Rapid Iterative Testing Report* (Usability Works, 28 February 2006), 11.
- ⁷⁹ Jane Lee, *OAC First round Usability Test Findings* (OAC Redesign Project, 11 September 2008), 15.
- ⁸⁰ NWDA, "Executive Summary: Usability Testing Round 4" (March 12, 2008), 6.
- ⁸¹ Dana Chisnell, *Report: RLG Archival Resources Service Redesign Focus Groups* (Usability Works, 2 December 2004), 17.

- 82 RLG in-house Archival Resources correspondence from 2003. See also Cathy De Rosa et. al., *Perceptions of Library and Information Resource* (Dublin, Ohio: OCLC Research, 2005), ix. Published online at: http://www.oclc.org/reports/pdfs/Percept_all.pdf.
- 83 "Generally, participants did not talk about tools for doing research...It is unlikely that researchers approach doing research by looking for a tool for doing research." Dana Chisnell, Report: RLG Archival Resources Service Redesign Focus Groups (Usability Works, 2 December 2004), 13.
- 84 Karen Markey, "Twenty-Five Years of End-User Searching, Part 1: Research Findings," *Journal of the American Society for Information Science and Technology*, 58 no. 8 (2007): 1071.
- 85 "Le traitement des fonds demeure cependant la préoccupation des centres d'archives et le lien entre l'information décrite et la clé d'accès fournie est d'autant plus fragile dans un contexte informatique," Louise Gagnon-Arguin, "Les questions de recherché comme matériau d'études des usagers en vue du traitement des archives," *Archivaria* 46 (winter 1998): 100.
- 86 Dana Chisnell, Report: RLG Archival Resources Service Redesign Focus Groups (Usability Works, 2 December 2004), 13.
- 87 Research Information Network, *Discovering Physical Objects: Meeting Researchers' Needs* (London, October 2008), 5. Published online at: <http://www.rin.ac.uk/objects>.
- 88 Elizabeth Yakel, "Listening to Users," *Archival Issues* 26:2 (2002): 118. Susan Hamburger, "How Researchers Search for Manuscript and Archival Collections," *Journal of Archival Organization* 2, no. 1/2 (2004): 83, 89. William J. Maher, "Use of User Studies," *Midwestern Archivist* 11 (1986): 23.
- 89 Kristina Southwell, "How Researchers Learn of Manuscript Resources in Western History Collections," *Archival Issues* 26, no. 2 (2002):100.
- 90 Maher, William J., "Use of User Studies," *Midwestern Archivist* 11 (1986): 22.
- 91 Jane Stevenson, "'What Happens If I Click on This?': Experiences of the Archives Hub," *Ariadne* 57 (October 2008). Published online at: <http://www.ariadne.ac.uk/issue57/stevenson/>.
- 92 Kathleen Feeney, "Retrieval of Archival Finding Aids Using World-Wide-Web Search Engines," *American Archivist* 62:2 (1999): 224.
- 93 Spindler, Robert P. and Richard Pearce-Moses, "Does AMC Mean 'Archives Made Confusing? Patron Understanding of USMARC AMC Catalog Records," *American Archivist* 56 (spring 1993): 340.
- 94 Michelle Light, "The Endangerment of Trees," EAD @ 10 (August 31, 2008): 3-5, *forthcoming*. Published online at: <http://www.archivists.org/publications/proceedings/EAD@10/Light-EAD@10.pdf>. Dryden is also concerned that we have—for years—developed standards for archival description without consulting "consumers" of these descriptions: Jean Dryden, "Do We Care What Users Want? Evaluating User Satisfaction and the LibQUAL+™ Experience," *Journal of Archival Organization* 4, no. 4 (2004): 84-85.
- 95 Lorcan Dempsey says that his aphorism, "Discovery happens elsewhere," "...is influenced by Steve Rubel's use of the phrase, 'traffic happens elsewhere' in discussion of the 'Cut and Paste Web' <<http://www.micropersuasion.com/2007/08/the-cut-and-pas.html>>. Published online at: <http://orweblog.oclc.org/archives/001430.html> (September 16, 2007).
- 96 William J. Maher, "Use of User Studies," *Midwestern Archivist* 11 (1986): 17.

8

“Capture and Release”: Digital Cameras in the Reading Room

Lisa Miller, Steven K. Galbraith,
and the RLG Partnership Working Group
on Streamlining Photography and Scanning

Introduction

Digital cameras are revolutionizing special collections reading rooms and the research process, much as photocopy machines did for a previous generation. Reference routines focused on the photocopier are embedded in workflows of every repository; photocopying is accepted by repositories, tolerated by rights holders, and expected by researchers. Now technology is forcing repositories to confront change again. The ubiquity of digital cameras and other mobile capture devices has resulted in researchers desiring and expecting to use cameras in reading rooms. While some librarians and archivists have resisted digital cameras, others have embraced them—and rightfully so. The benefits to researchers, repositories, and collection materials are undeniable.

Benefits

Digital cameras are gentler on collection materials—Upending collection materials to position them on a photocopy machine, even when done with the utmost care, risks more damage to materials than photographing them in the reading room while they are face up and appropriately supported. The materials are not subjected to the intense light of a photocopier, but rather are usually easily photographed with ambient lighting.

Digital cameras facilitate use—Researchers with limited time can cover more collection materials during their visit by photographing relevant materials for in-depth study later. We preserve these materials so that they can be used. More use allows us to report higher reference figures and significant research use to our resource allocators.

Digital cameras increase researcher satisfaction—Researchers must take time from work and school to travel to our reading rooms during our limited business hours, often at great expense. Just as libraries and archives struggle with tighter budgets in these challenging economic times, so, too, do researchers. Digital cameras maximize their precious time in the reading room and end their wait for copies. Depending on the nature of the repository's camera use policy, patrons may

also save money and eliminate time spent on photocopy request paperwork. They may also make copies of a broader universe of materials, like oversize materials and bound volumes that are excluded from the photocopy policies of many repositories, and they can make color copies. Given a choice between two repositories, one that has more generous policies and one that does not, researchers may make choices accordingly.

Digital cameras reduce repository workload—Depending on the repository's photocopy and digital camera policies, allowing personal digital cameras outsources duplication tasks to the user, freeing staff to perform other work in these times of increased demands, expectations, and workloads. In addition, cameras may reduce photocopier maintenance and supplies.

Digital cameras enhance security and save reading room checkout time—Digital cameras decrease the number of photocopies leaving the reading room in the hands of researchers, reducing checkout time and the opportunity for theft. With twentieth- and twenty-first-century collections, it is frequently difficult to distinguish between copies and originals.

Digital cameras save paper and photocopy toner—Photographing materials is an effortless way to reduce our environmental impact.

Repositories stay current and resolve an ongoing issue—Repositories remain largely analog outposts, in contrast to the 24/7 online world that most people live and work in. As much as we would like to deliver collection materials to all online, it is still beyond our grasp. Digital cameras are research tools that reach across this online/offline divide, one researcher at a time.

Digital cameras reduce liability for copyright infringement—Digital cameras lessen the repository's risk profile, especially if it maintains a "hands-off" approach towards the use of personal cameras. When a repository makes copies of copyrighted documents for users or provides equipment on which users can make their own copies, it runs the risk of engaging in direct and indirect copyright infringement.

Duplication, Copyright and the Web

Section 108 of U.S. copyright law allows repositories to make digital copies of textual material for private study, scholarship, or research. If making a copy of an entire book or manuscript item or a substantial part of it for a user, the repository must determine that a copy is not available at a fair price.¹ For non-textual material, such as photographs, the law allows repositories to make copies only if the repository concludes that the user's request is a fair use—a difficult, and potentially risky, assumption.² By allowing patrons to use their digital cameras, the repository removes itself from the duplication process and eliminates the risk associated with making copies.

Section 108(f)(1) protects a repository from secondary liability for the “*unsupervised* use of reproducing equipment located on its premises” (emphasis added), provided that the equipment displays a notice that making copies may be subject to copyright law.³ Ironically, supervised use of reproduction equipment, such as requiring users to seek permission before making any copies, increases the repository’s risk of liability.

Rather than place a notice on cameras, the Section 108 Study Group recommended that a notice be posted prominently in public areas stating that making copies may be subject to copyright law.⁴ Such a statement should also appear on digital camera use agreements signed by researchers.

Reading room photography does not lead inexorably to collection materials inappropriately ending up online. This issue is already managed by each repository’s publication policy. Many repositories have been providing digital reproductions to patrons for years under existing duplication and publication policies. Given how easy it is to digitize analog reproductions, drawing distinctions between analog and digital copies makes little sense. Some repositories consider responsible reuse of images on the Web as good outreach.

Suggested Practices for Cameras in the Reading Room

To synthesize a core of suggested practices, the RLG Partnership working group reviewed the current policies of thirty-five repositories comprised of academic libraries, independent research libraries, historical societies, government archives, and public libraries (see *Appendix A. Policies Reviewed*). Below are the most commonly shared elements, arranged in categories for administration and handling of collection materials.

Administration

- Require camera users to complete and sign an application/policy/terms-of-use form agreeing that images of sensitive and copyrighted materials will only be used for study, teaching, or research purposes and will be used in compliance with copyright law. Some agreements also stipulate that the user cannot reproduce images without permission from the institution. A few forms require the user to list specifically what he or she is digitally reproducing. This allows the institution to keep statistics on what and how much is being digitized and to check whether any of the materials already exist in the institution’s digital repository, though it increases liability for copyright infringement.
- Staff reviews collection materials prior to photography. This ensures that items are not too fragile to be reproduced and allows staff to note any copyright or donor restrictions, though it also places the institution at greater risk of liability.
- Limit the number of shots, when appropriate, to a quantity determined by institutional policy and/or in accordance with copyright policies.
- Watermark digital reproductions by requiring that each item be photographed with a streamer, transparency, or card that identifies the item and its holding institution and, if applicable, displays a copyright notice. Patrons are responsible for properly citing their copies, but repositories may provide citation guidelines.

- Digital photography must not disturb other users or staff. All audio functions on digital cameras must be turned off and users may not photograph other patrons, staff, or the reading room.

Handling Collection Materials

- No flash photography. It is a distraction to other users.
- As with any method of duplication, camera use is considered only if it will not damage collection materials. Users are instructed on how to handle items during photography.
- In an effort to monitor how users handle items during photography, several institutions designate specific work areas where items may be photographed or have a staff member present during shooting. Some provide or require use of an in-house camera stand. Some policies make a point of prohibiting users from bringing their own tripods or lighting equipment.

Evolving Practices for Digital Cameras in Reading Rooms

- Beyond the suggested practices above, many facets of digital camera use continue to develop and can be implemented independently along sliding scales represented in Table 1. A repository can mix and match from these modules according to its nature, needs, and inclination.
- Established photocopy policies and processes often form the baseline for a repository's digital camera policy. If the staff performs all photocopying, an appointment and designated workstation for digital photography, supervised by the photocopy staff and with time charged to the researcher, may be the logical approach. As an alternative, the digital camera policy could steer researchers toward some goal of the repository, such as reducing the staff's photocopy workload or achieving a paperless duplication system. A repository might encourage both of these goals by continuing to charge its standard rate for photocopies while not charging for copies made with digital cameras. The repository can swap out one facet for another as it experiments with cameras, and gradually settle on a policy that works for it.

Table 1. Faceted Camera Use Grid

Facet	Shutter-bug	Exposed	Camera-shy
Traditional photocopying (possible baseline for digital camera policy)	Self-service	Self-service after staff review	All copying done by staff
Equipment	<ul style="list-style-type: none">• No flash, no lights• Allow flatbed scanners• Allow and/or provide copy stand, tripod, extension cords, stepstool, etc.• Repository supplies camera or self-service overhead book scanner in addition to allowing patron's camera	<ul style="list-style-type: none">• Patron's camera• Limits on supporting equipment (copy stands, tripods, cords, etc.)—some pieces allowed, others not• No flatbed scanners	Repository's camera only (and possibly other equipment supplied by repository)
Photography space	In reading room at any station	In reading room at designated stations, usually close to reference desk	Separate room
Photography rules	<ul style="list-style-type: none">• No standing on tables or chairs• No rearrangement of furniture• No materials on floor• Remain behind table, facing forward at all times• Set camera to "mute"• Do not disturb others• No photographs of reading room, staff, or patrons		
Appointments	Appointment not required	Appointment made during visit	Appointment made in writing in advance

Table 1 continued on next page

Facet	Shutter-bug	Exposed	Camera-shy
Staff review of collection materials	Part of standard staff surveillance of patrons in reading room	Patron must verbally notify reference attendant each time camera is used and show attendant the materials being photographed	<ul style="list-style-type: none"> • Patron must formally indicate and curatorial staff formally review all materials • Camera stays in locker until approval is given • Same-day approval may not be possible
Materials handling rules	<ul style="list-style-type: none"> • Preservation needs always trump photography needs • Do not manipulate materials to achieve a better image • Do not press down on materials or bindings • Manuscript materials must always be flat on the table and not held up in air • Loose materials must remain in their folder and in order at all times • Photograph materials from one folder at a time • Volumes should not be laid flat—book cradles will be provided • Weight bags and snakes are available • Do not fold pages • Do not remove fasteners—ask for staff assistance • Do not remove items from sleeves, mats, etc. 		

Table 1 continued on next page

Facet	Shutter-bug	Exposed	Camera-shy
Quantity limits	No limits	<ul style="list-style-type: none">• No more than 50 pages or 20 percent (whichever is smaller) of any manuscript or book• No entire book, manuscript box, or collection, nor substantial portions of them• Please limit number of photographs to a reasonable amount• Photographs are meant to alleviate photocopying and supplement note taking, not to create a complete personal copy	<ul style="list-style-type: none">• Limit to established number of shots per day• Patron's images may be reviewed during checkout to enforce quantity limit
Other limits		<ul style="list-style-type: none">• Oversize items or anything that does not safely fit on table• Fragile or damaged items• No materials received on interlibrary loan, unless lending library permits• Only materials checked out to the patron using the camera• If not allowed, staff may digitize at standard fees• Repository reserves right to deny permission for any collection materials at its discretion	
Copyright		<ul style="list-style-type: none">• Copyright notice (and citation) in all shots (paper strip or transparency)• Digital copies are for personal research use only• Repository displays a copyright warning where digital camera requests are accepted and on digital camera policy forms	
Paperwork (in addition to forms completed by all patrons)	<ul style="list-style-type: none">• Camera use agreement included on registration form• Separate camera use agreement with copyright declaration (renewed annually/ per visit/per day)	<p>Patron provides list of collections (plus camera use agreement)</p>	<ul style="list-style-type: none">• Patron provides list of each item (plus camera use agreement)• Written request before visit

Table 1 continued on next page

Facet	Shutter-bug	Exposed	Camera-shy
Fees	None	<ul style="list-style-type: none"> Minimal fee (per visit, per shot) Fees for equipment supplied by repository 	Fee equals or exceeds cost of photocopies
Publication	Publication requires permission of the copyright holder	Images for publication or distribution must be ordered through the library at set fees	<ul style="list-style-type: none"> Images taken by patron may not be published in print or on Internet Publication requires written permission from repository
Citations		<ul style="list-style-type: none"> Patron is responsible for recording complete citations for each shot Subsequent orders for high-resolution images cannot be processed without complete citations Source repository template in all shots (paper strip or transparency, often included with copyright notice) 	
Other		<ul style="list-style-type: none"> Camera privileges can be revoked at any time if rules are not followed Provide tips on taking good images and creating complete citations In some cases, repository receives copies of all photographs 	

Many repositories charge for photocopies and this can form a basis for charging for digital camera use. When a repository establishes such fees, it should follow the guidelines noted in the *ALA/SAA Joint Statement on Access to Research Materials in Archives and Special Collections Libraries*. This document states in part, "A repository should facilitate access to collections by providing reasonably priced reproduction services that are administered consistently in accordance with legal authority, including copyright law, institutional access policy, and repository regulations. These services . . . should be clearly stated in a publicly accessible written policy."⁶ Charging fees for reproductions of copyrighted material may place the institution at greater risk for copyright infringement. If the fees are determined to provide "direct or indirect commercial advantage" to the repository, its Section 108 exemptions are lost and maintaining a "fair use" defense becomes much harder.

A few repositories have introduced particularly unique facets to their digital camera policies, as noted in the "other" section of the grid. Some ask for copies of all digital images, with citations, and add them to the repository's collection of digital assets. In these cases, the repository may wish to include a statement to that effect in the digital camera use agreement.

To assist researchers in obtaining usable photographs and citations, some repositories provide photography tips to their patrons.⁷

Conclusion

Digital cameras are the newest research tool, but they will not be the last. The next generation of archivists, librarians, and curators will view digital cameras the way we currently view photocopy machines, as essential components of our reference system. The issues of new technology are wrongly framed as a threat or a challenge for repositories to remain relevant.⁸ Rather, digital cameras should be considered from the perspective of our most fundamental goals—improving conditions for our collections materials, facilitating greater research economically and efficiently, and resolving competing demands for resources and maximizing the productivity of our staff. By adopting this mindset with our digital camera policies, we are poised to evaluate objectively the technology that will replace digital cameras in the next generation—or sooner.

Peter Hirtle, Jim Kuhn, Merrilee Proffitt, Jackie Dooley and Ricky Erway reviewed early versions of this report. The final document is better as a result of their comments, which are greatly appreciated.

Appendix A. Policies Reviewed

- American Antiquarian Society
- Arizona State University, Arizona Historical Foundation
- Brigham Young University, L. Tom Perry Special Collections
- California Historical Society
- Cornell University, Division of Rare and Manuscript Collections
- Dallas Theological Seminary
- Duke University, Rare Book, Manuscripts, and Special Collections Library
- Emory University, Pitts Theology Library Archives and Manuscripts Department
- Folger Shakespeare Library
- Frick Art Reference Library
- Getty Research Institute
- Harvard University, Houghton Library
- The Huntington Library, Arts Collections, and Botanical Gardens
- Indiana University Bloomington, Lilly Library
- Library of Congress, Prints & Photographs Division
- Library of Virginia
- Minnesota Historical Society
- The National Archives at College Park, Maryland
- The National Archives, United Kingdom
- New York Public Library, Manuscripts and Archives Division, Berg Collection, and Schomburg Center
- New York University, Fales Library
- The Newberry Library
- San Francisco Public Library, San Francisco History Center
- Stanford University, Hoover Institution Archives
- Stanford University, Special Collections and University Archives
- Syracuse University
- Tulane University, Louisiana Research Collection
- University of Alaska Anchorage & Alaska Pacific University Consortium Library, Archives & Special Collections
- University of California, Berkeley, Robbins Collection
- University of California, Irvine, Langson Library Special Collections
- University of California, Los Angeles, Charles E. Young Research Library Department of Special Collections
- University of Maryland at College Park, Special Collections
- University of Miami, Special Collections and University Archives
- University of Texas at Austin, The Harry Ransom Center
- University of Virginia, Albert and Shirley Small Special Collections Library

Appendix B.

Draft Modular Form: Camera Use in Reading Room

This sample form can be adapted by a repository by deleting irrelevant sections or inserting additional specific requirements. It is available as a standalone editable document on the OCLC Research Web site at <http://www.oclc.org/research/activities/photoscan/policy.doc>.

Camera Use Policy

Researchers may take photographs of collection materials for study purposes only, and as allowed by the library, based on the physical condition of the materials, copyright law, donor restrictions, and reading room rules.

I agree to the following conditions:

Repository procedures [delete or add as needed]

- I will obtain permission from library staff before taking any photographs.
- I will indicate all items to be photographed and show them to library staff for approval.
- I will provide a list of all [collections or items] photographed.
- I will take photographs at designated stations only.
- I will not photograph more than [50 pages or 20 percent of any book or manuscript (whichever is smaller), 100 pages per collection, other arbitrary limit].
- I will use my personal camera only—not portable scanners, [phone cameras, other].
- I will include in each photograph a strip provided by the library stating [repository name, copyright notice, and/or citation].
- It is my responsibility to keep accurate citations for all items photographed, which I will need when ordering publication-quality images or requesting permission to quote.

Materials handling rules [delete or add as needed]

- I will handle the materials with care and according to library rules.
- I will not bend, press down, or otherwise manipulate or rearrange materials to get a better photograph.
- I will keep materials flat on the table or in the stand/cradle provided.
- I will ask library staff for assistance with fastened items.
- I will not remove items from their plastic sleeves.
- I will not stand on chairs, tables, or other furniture.
- I will turn off the flash and sound on my camera.
- I will not use special lights [other prohibited equipment].
- I will not take photographs of the staff, reading room, or other researchers.
- I understand that the library reserves the right to deny permission to photograph collection materials at its discretion.

Copyright [delete or add as needed]

- I will use the photographs for my private study, scholarship and research only.
- I will not publish the photographs in print, post them on the Internet, nor exhibit them.
- I will not donate, sell, or provide the photographs to another repository.
- I will request publication-quality images from the library at its standard fees.
- It is my responsibility to obtain permission to publish from copyright owners.

Repository gets copies of all photographs taken [delete or add as needed]

- I will provide copies of all of my photographs and citations to the library, and I assign any intellectual property rights that I may possess in them to the repository.

WARNING CONCERNING COPYRIGHT RESTRICTIONS

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be "used for any purpose other than private study, scholarship, or research." If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use," that user may be liable for copyright infringement.

This institution reserves the right to refuse a copying order if, in its judgment, fulfillment of the order would involve violation of copyright law.

I agree to indemnify and hold harmless [repository name], its agents and employees against all claims, demands, costs and expenses incurred by copyright infringement or any other legal or regulatory cause of action arising from the use of these photographs.

I have read and agree to abide by the terms and conditions above. I understand that my failure to follow them may result in the termination of my camera privileges.

Signature

Date

Name (Please print)

List of collections photographed. Please print clearly. [delete or add as needed]

1.	
2.	

List of items photographed. Please print clearly. [delete or add as needed]

Item	Collection	Box	Folder	Item Description
1				
2				

Appendix C.

Members of the RLG Partnership Working Group on Streamlining Photography and Scanning

- Anne Blecksmith
Getty Research Institute
- Eleanor Brown
Cornell University
- Paul Constantine
University of Washington
- Gordon Daines
Brigham Young University
- Tiah Edmunson-Morton
Oregon State University
- Cristina Favretto
University of Miami
- Steven K. Galbraith
Folger Shakespeare Library
- Susan Hamson
Columbia University
- Sue Kunda
Oregon State University
- Jennie Levine Knies
University of Maryland
- Suzannah Massen
Frick Art Reference Library
- Dennis Massie
OCLC Research
- Dennis Meissner
Minnesota Historical Society
- Elizabeth McAllister
University of Maryland
- Lisa Miller
Hoover Institution Library and Archives, Stanford University
- Timothy Pyatt
Duke University
- Jennifer Schaffner
OCLC Research
- Shannon Supple
Robbins Collections
University of California, Berkeley
- Francine Snyder
Solomon R. Guggenheim Museum
- Mattie Taormina
Stanford University
- Cherry Williams
Lilly Library, Indiana University Bloomington

Notes

- 1 “Limitations on exclusive rights: Reproduction by libraries and archives,” Section 108, Chapter 1, Title 17, United States Code, accessed February 1, 2010, at http://www.law.cornell.edu/uscode/17/usc_sec_17_00000108---000-.html.
- 2 For more on direct and indirect copyright infringement and its relevance for reading room practice, see: Peter B. Hirtle, Emily Hudson, and Andrew T. Kenyon, *Copyright and Cultural Institutions: Guidelines for Digitization for U.S. Libraries, Archives, and Museums* (Ithaca, NY: Cornell University Library, 2009): pp 78-83, accessed February 1, 2010, at <http://hdl.handle.net/1813/14142>.
- 3 “Limitations on exclusive rights: Reproduction by libraries and archives,” Section 108, Chapter 1, Title 17, United States Code, accessed February 1, 2010, at http://www.law.cornell.edu/uscode/17/usc_sec_17_00000108---000-.html.
- 4 Copyright Office and National Digital Information Infrastructure and Preservation Program of the Library of Congress, *Section 108 Study Group Report* 91-92 (2008), accessed February 1, 2010, at <http://section108.gov/docs/Sec108StudyGroupReport.pdf>.
- 5 For a study of the effects of light on various materials, see: Terry T. Schaeffer’s *Effects of Light on Materials in Collections: Data on Photoflash and Related Sources* (Los Angeles, CA: Getty Publications 2001).
- 6 *ALA/SAA Joint Statement on Access to Research Materials in Archives and Special Collections Libraries*, adopted by the SAA Council on June 1, 2009, and the ACRL Board during the ALA Annual Conference, July 2009, accessed February 1, 2010, at <http://www.archivists.org/statements/ALA-SAA-Access09.asp> and <http://www.ala.org/ala/mgrps/divs/acrl/standards/jointstatement.cfm>.
- 7 Lyndon Baines Johnson Library and Museum, “Instructions for Use of Digital Cameras in the Reading Room,” May 20, 2004, accessed February 1, 2010, at http://www.lbjlib.utexas.edu/johnson/archives.hom/digital_camera_info.shtm, and Kirklin Bateman, Sheila Brennan, Douglas Mudd, and Paula Petrik, “Taking a Byte Out of the Archives: Making Technology Work for You,” *Perspectives*, 43,1 (January 2005), accessed February 1, 2010, at <http://www.historians.org/Perspectives/Issues/2005/0501/0501arc1.cfm>.
- 8 Richard Cox with the University of Pittsburgh archives students, “Machines in the Archives: Technology and the Coming Transformation of Archival Reference,” *First Monday*, 12,11 (November 2007), accessed February 1, 2010, at <http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/2029/1894>.

9

Tiers for Fears: Sensible, Streamlined Sharing of Special Collections

Dennis Massie
Program Officer
OCLC Research

Acknowledgments

Jennifer Schaffner, Program Officer, OCLC Research, has from the beginning been the driving force behind this work. Her passion for bringing special collections materials to a wider audience continues to be both transformative and inspirational.

Special thanks go to the members of the Special Collections Delivery Steering Committee—Cristina Favretto of the University of Miami, Susan Hamson of Columbia University, and Mattie Taormina of Stanford University—who in 2009 identified the physical sharing of special collections as one of a set of key cooperative projects that could transform the way special collections materials are brought together with the researchers who need them.

The Sharing Special Collections Working Group did all the heavy lifting and deserves the lion's share of the credit for this report:

- Jennifer Block, Princeton University
- Scott Britton, Boston College (formerly of University of Miami)
- Barbara Coopey, Pennsylvania State University
- Aimee Lind, Getty Research Institute
- Sandra Stelts, Pennsylvania State University
- Dennis Massie and Jennifer Schaffner, OCLC Research

The Sharing Special Collections Advisory Group consulted extensively with the working group, guided the conversation about the physical lending of special materials, and served as expert sounding board:

- Eleanor Brown, North Carolina State University (formerly of Cornell University)
- Laura Carroll, Emory University
- Margaret Ellingson, Emory University
- Cristina Favretto, University of Miami
- Paul Constantine, University of Washington
- Suzan Hallgren, University of Minnesota
- Elizabeth Nielson, Oregon State University

- Shannon Supple, University of California, Los Angeles (formerly of University of California, Berkeley)

Our working and advisory groups enjoyed a special collaborative relationship with the RBMS Task Force responsible for the 2012 revision of ACRL's Guidelines for Interlibrary and Exhibition Loans of Special Collections Materials, thanks to chair Hjordis Halvorson of the Newberry Library, Christian Dupont of Atlas Systems, and our operative in common, Shannon Supple of UCLA.

Members of the 2012 SHARES Executive Group (particularly Lesliediana Jones of the George Washington University Law School Library) improved the SHARES Facility Trust Checklist and added several additional use cases.

Ricky Erway, Senior Program Officer, OCLC Research, significantly improved both the text and the organization of this report.

Executive Summary

This report presents strategies for providing efficient and affordable interlending of actual physical items from special collections for research purposes, as well as advice on determining if a loan is the most appropriate way to fulfill a particular request.

The lending of physical items for exhibition purposes has long been a core activity of archivists and special collections curators. Now, with the increased visibility of special collections, requests for research loans are multiplying. There are legitimate instances—based on the nature of the material, the type of research question, or the need for extended access by a distant scholar—when only the loan of a physical item from special collections can satisfy a researcher's request.

Prudent approaches to lending rare and unique materials are justified, and providing a digital surrogate is usually the answer. But such thinking is not appropriate for every item in special collections, or for every request, and often results in time-consuming, overly cautious procedures. Streamlining such procedures is critical. Labor-intensive processes and policies can be simplified to fit the nature of the material, institutional resources, the circumstances of requests, and the risk tolerance of curators and administrators.

Lending physical items ranks among the most divisive issues in the field of archives and special collections, perhaps the one most likely to bring out equal parts raw emotion and well-reasoned professional opinion. But solid evidence indicates that the practice of lending physical items from special collections is becoming as common as not doing so. While an increasing number of curators are willing to consider the physical loan of materials under their stewardship, the workflows for considering and executing such loans tend toward unscalable. In order for curators to cope with the uptick in requests and arrive at a well-considered and professionally-responsible “yes” as often as possible, new workflows and new ways of thinking about lending physical items from special collections must be established.

From 2009 through 2011, a working group made up of resource sharing supervisors and special collections curators from OCLC Research Library Partnership institutions studied this issue. The most significant activity of the working group was creating a set of tools that will help institutions reconsider and streamline their processes for handling loan requests for special collections materials.

These tools include:

- a tiered approach to streamlining workflows associated with lending special collections, outlining minimal, moderate and maximum amounts of effort and overhead, to be invoked based on
 - the material
 - the request
 - the risk tolerance of curators and administrators
- a model written policy on sharing special collections
- a “trust” checklist to serve as a conversation starter between a prospective lender and an institution interested in borrowing an item from special collections

This report contains a complete description of the working group's activities, plus all of the tools listed above, and advice on how best to use them. The report's principles intentionally dovetail with the Association of College and Research Libraries' 2012 revision of Guidelines for Interlibrary and Exhibition Loans of Special Collections Materials (ALA 2012).

Introduction

Enhanced discoverability of special collections has led to increased interest from researchers. Concurrently, advances in scanning technology have helped make the provision of such materials in digital form fairly routine. There are instances, however—due to the nature of the material, the type of research question, or the need for extended access by a distant scholar—when only the loan of a physical item from special collections can satisfy a request.

“Says who?” you might ask. Says two-thirds of community practitioners, according to a survey conducted for this report.

Ten years ago, requests for loans of special items for exhibition purposes were routine. Loans for research purposes were rare. Many institutions refused to consider such requests. Those that did turned each request into what amounted to a special project, requiring multiple internal consultations and extensive contacts between staff at the borrowing and lending institutions. Each step of the process, including packing and unpacking, required the participation of specially-trained experts.

With the increased visibility of special collections, requests for physical loans have multiplied. They arrive at prospective lending institutions in two separate streams, directly to the special collections curators and also via interlibrary loan departments. While an increasing number of curators are willing to consider the physical loan of materials under their stewardship, the workflows for considering and executing such loans don’t scale well. In order for curators to cope with the increased volume in requests and arrive at a professionally-responsible “yes” as often as possible, new workflows and new ways of thinking about lending physical items from special collections must be established.

This report presents strategies for determining if a loan of the original item is the most appropriate way to fulfill a particular request for special collections material and offers techniques for providing efficient and affordable delivery of physical items. Cautious approaches to lending rare and unique materials, while justified, are not necessarily appropriate for every item in special collections and often result in time-consuming procedures. Labor-intensive processes and policies can be streamlined to fit institutional resources, the circumstances of requests, and the risk tolerance of curators and administrators.

Let's Get Physical

The work described in this report got its initial spark from the same steering committee that previously championed allowing cameras in the reading room and providing scan-on-demand services for users of special collections materials—both topics, in those days (2009), rather controversial ideas in themselves. But this issue always stood apart. The physical lending of special collections was put on the table, whisked off, and then nudged back on again. “I know we should be talking about this,” said one committee member, “if only because it makes me feel so uncomfortable.” Another agreed: “We have a professional responsibility to push at our boundaries and question our comfort zones.” The third added, “I love the idea of lending from special collections. But I would never be allowed to bring it up at my own institution.”

The idea has been around for a long time. Some prestigious institutions have been doing it for years, almost completely without mishap. The Historical Society of Wisconsin, for instance, has since the early 1970s operated a statewide network of regional research centers which moves archival materials around so that researchers can use them close to where they live (Erney and Ham 1972). More recently, in 2010, Elaine Engst of Cornell University sent an entire archival collection to Columbia University so that a Manhattan-based FBI agent could, over many months, search for crucial provenance evidence in an effort to recover letters allegedly stolen from a special collection at the New York Public Library (2012). NYPL had no item-level description of the collection, but decades ago a Cornell Ph.D. candidate consulted it and extensively described many of the letters in notes made while preparing his dissertation. Those notes were the key to the case. This represents a classic instance where only prolonged access to a complete set of original archival material at a spot near the user’s home base could adequately satisfy the need.

But emotions on this issue can run high, and professional peer pressure can be intense. I offer one example from my own experience:

In 2003, following a well-received Research Libraries Group program called *Sharing the Wealth*, where staff from dozens of institutions in the US and a few from the UK came together in Washington, D.C., to talk about their experiences with sharing physical items from special collections, I formed a working group to develop a pilot project that would promote such loans. Within a few weeks, I was pulled aside by the director of a top-tier ARL library who said, only half-jokingly, “Can’t you find something else to work on? This sharing special collections business has my staff yelling at each other in the hallways.” The working group’s only UK representatives soon begged off, because they felt their peers were not ready for a rational conversation about the topic. The working group ended up gathering some interesting examples of documentation and best practices for lending special collections, but the pilot project itself never materialized.

Fast forward to today . . .

Lending physical items from special collections for research purposes is finally an idea whose time has come—for some. It remains among the most divisive issues in the field of archives and special collections, perhaps the one most likely to bring out equal parts raw emotion and well-reasoned professional opinion. But solid evidence indicates that the practice of lending physical items from special collections is becoming more commonplace than not doing so. The *Sharing Special Collections Working Group’s 2010 survey of 88 special collections and archives departments in North America, Europe, Australia, and Africa*, found that 57.4% of respondents will lend physical items from their special collections within a consortium, while another 10.3% will lend even beyond their favored group (See figure 1.). That’s 67.7% of respondents who lend physical items from special collections at least some of the time.

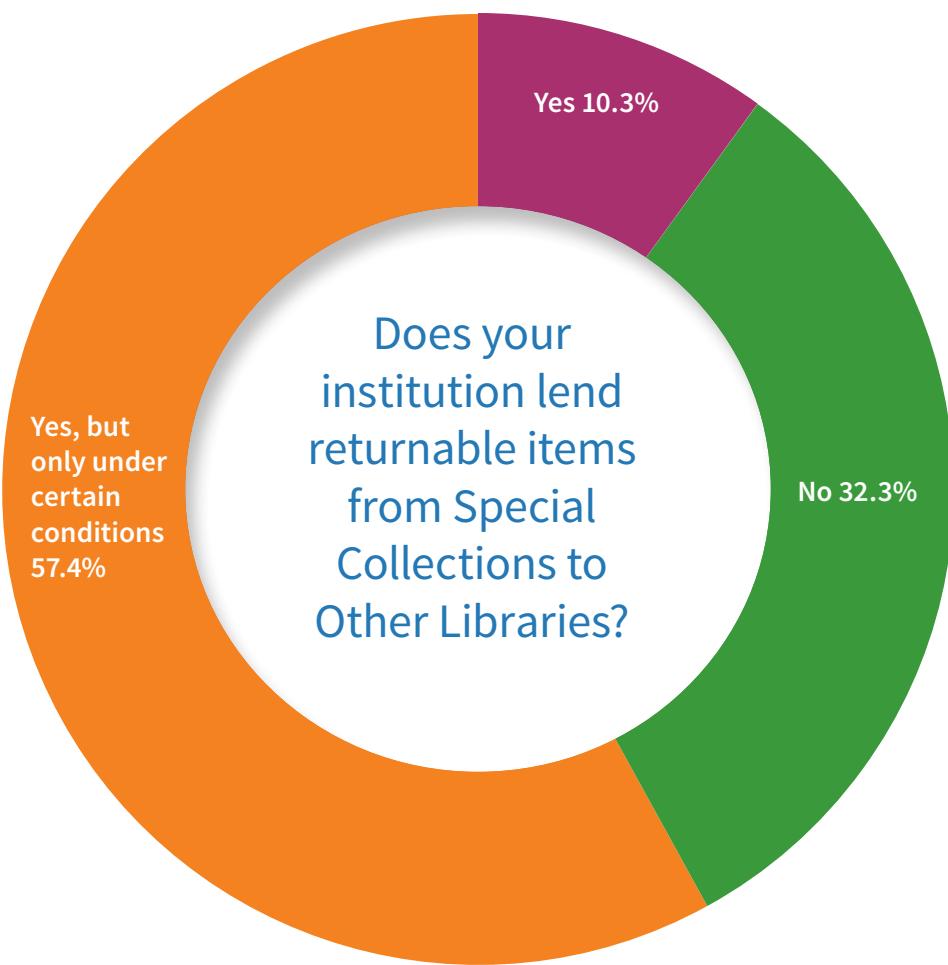


Figure 1. Most respondents (67.7%) physically lend special collections items

Getting to “Yes”

The working group was made up mostly of teams of special collections curators and interlibrary loan supervisors and included seasoned veterans of lending special collections, those who had experimented with the practice, and one team considering doing so for the first time. One of our initial tasks was to develop a set of “first principles” to guide our exploration of the issues:

- Lending a physical item from special collections is an exception, appropriate only when providing a surrogate copy would fail to satisfy the request.
- Considering a loan from special collections often requires a flip in mindset from “Why?” to “Why not?”
- Not everything held in special collections is equally special.
- Not every requester of special collections material realizes that the item is held in special collections.
- Let those who are best positioned to do something do it.
- Lending physical items from special collections requires trust, both internally and externally.
- An interlibrary loan (ILL) of special collections material counts as use.
- Borrowers of special collections should give serious consideration to being lenders; lenders of special collections should be entitled to some expectation of success in borrowing.

The working group devoted time and energy to a number of activities designed to promote the physical lending of special collections:

- Compiling a glossary for use by the working and advisory groups (the main contribution of which was to establish that by “special collections” we meant any material held in formal special collections or archives departments).
- Conducting a survey (sent via international discussion lists) of current practices and attitudes regarding the sharing of special collections, targeting both special collections and interlibrary loan practitioners.

- Producing a webinar, *Treasures on Trucks*, which featured a recent history of sharing special collections and a panel discussion featuring grizzled veterans alongside newcomers to the practice (Schaffner and Massie 2009).
- Supporting and informing the work of the RBMS Task Force that, in 2011, revised the ACRL guidelines on sharing special collections for exhibit and for research, with our main contribution being to ensure that sufficient numbers of interlibrary loan professionals and archivists commented on the draft guidelines (See this report’s list of references on page 39 for a link to the revised guidelines, which have since been endorsed by the Association of College and Research Libraries’ board of directors and the Society of American Archivists Council). (ALA 2012)

The most significant activity of the working group was creating a set of tools that will help institutions reconsider and streamline their processes for handling loan requests for special collections materials. These tools include:

- A tiered approach to streamlining workflows associated with lending special collections, outlining minimal, moderate and maximum amounts of effort and overhead, to be invoked based on
 - the material
 - the request
 - the risk tolerance of curators and administrators
- A model written policy on sharing special collections
- A “trust” checklist to serve as a conversation starter between a prospective lender and an institution interested in borrowing an item from special collections

This report contains a complete description of the working group’s activities, plus all of the tools listed above, and advice on how best to use them. Let the sharing begin. And continue.

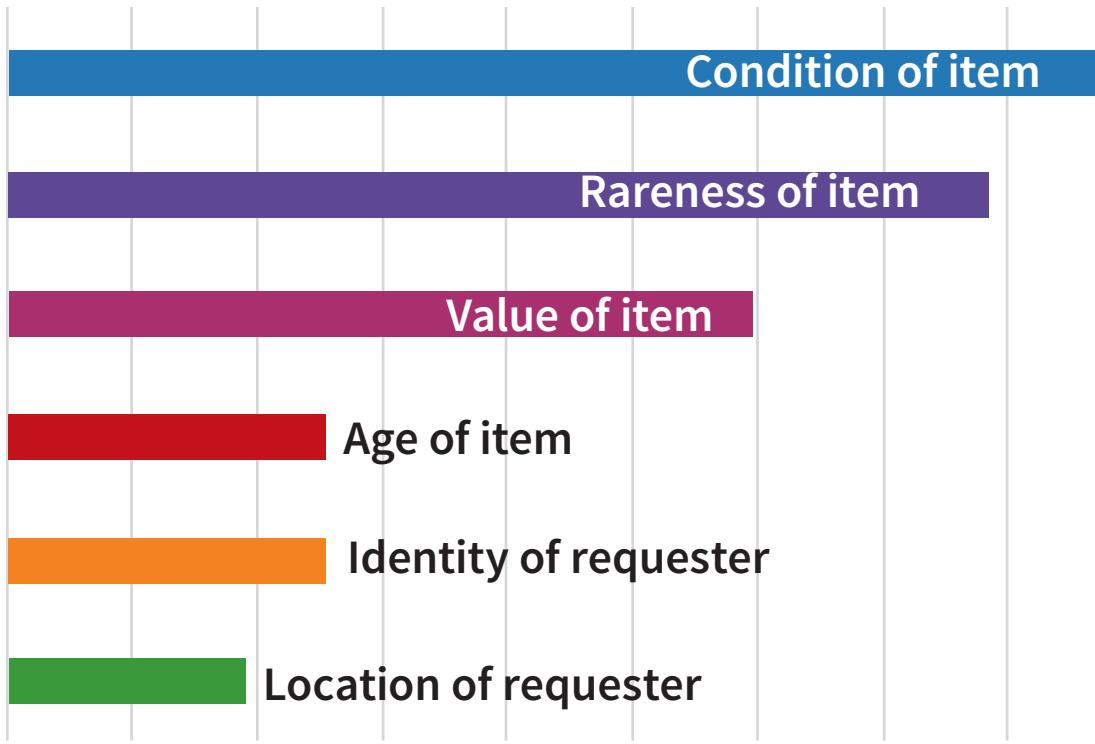
Highlights of the 2010 Sharing Special Collections Working Group Survey

In April 2010, the working group conducted a survey in order to solicit current attitudes, practices, policies, and priorities regarding the lending of special collections materials for research purposes. We cast a wide net, announcing the survey on major primary sources and interlibrary loan discussion lists and inviting any library with a special collections department to reply. Survey instructions encouraged respondents to have ILL and special collections staff members work together in answering the questions.

We received 88 responses. Types of responding institutions varied greatly and included academic, national, and public libraries, plus museums and historical societies. The overwhelming majority of responses came from North America, with four from continental Europe and one each from Africa and Australia. Respondent job titles included a mix of special collections curators, reference or access heads, and interlibrary loan supervisors, along with a few university archivists and library directors.

Major revelations included:

- Lending physical items from special collections is now more common (67.7%) than not doing so, at least within consortia.
- Digitizing on demand has become routine.
- Condition of the item is still the key to the lending decision.
- Attitudes toward unpublished materials are more restrictive than toward published.
- 36% indicated they have written policies for sharing special collections (but no one had an overall policy; each example covered a particular aspect or format).
- “Too risky” (69%) is by far the most common reason for not sharing returnable special collections (i.e., original items held in special collections that must be returned at the conclusion of the loan period).
- “Because we never have” and “Not part of our mission” each got more votes than “Lack of staff resources” as main reasons not to lend returnable special collections materials.
- Most interesting comment: “We were able to borrow things we would not be able to provide to others.”



Coopey and Stelts. 2011 (Revised with permission)

Figure 2. Issues involved in physically lending special collections items

The first section of the survey focused on policy. It was impossible to find any sort of consensus in the responses, other than one sizable camp being willing to lend even its treasures to trusted partners, while another sizable camp expresses an aversion to risk that at times sounds more like fear.

One respondent (an ILL staff person) expressed surprise at learning through completing the survey that special collections staff often receive and fill requests directly without ILL involvement; the ILL'er felt that such requests should be routed through the established resource sharing channels, because a willingness to lend such items creates a valuable reserve of good will for that library out in the community when it comes time to borrow.

In response to an open-ended question about how the decision is made to lend or not to lend, we mostly received confirmation of what was learned from the multiple-choice questions: condition matters most, with other factors such as rarity, value, popularity, and proximity (of the requester to the supplier, or of the requester to other copies of the same material) carrying significant weight. But one response was so thoughtful and

comprehensive in approach—while so perfectly capturing the spirit of considering each request on its own merits—that it deserves to be quoted in full:

We look at WorldCat to see how many other libraries have the item and where they are located. If the requester lives within a day's to-and-fro driving distance, we would usually prefer that the researcher come to us to use the book. If we find via WorldCat that the requester lives nearby another library (a researcher from the University of Chicago who wants a book that is owned by the Newberry Library, for example) we would usually decline to lend, particularly if the title is scarce. We think about the rigors of traveling and how the journey might affect the condition of the book. Some items are just too frail to lend and must be used under curatorial supervision. We think about the type of book—novelty books with pull tabs, fragile pop-up books, etc.—and turn down requests to borrow. When we are reluctant to lend, we often look up researchers in their university directory and email them to ask what specifically they are looking for. Sometimes we can fill a request by photocopying the table of contents or the index, or perhaps a few relevant

pages. When requesters learn that the book is in a special collections library, we find that some say, "Oh, never mind. It isn't terribly important" or "I wouldn't want you to lend a rare book." Sometimes we find that researchers have gone on to another topic or no longer need the book because of a deadline. It sometimes makes a difference whether a researcher is engaged in an initial fishing expedition on a topic or whether the book would contribute to a major project. If we learn that our book is a vital part of a researcher's work, then we will go out of our way to accommodate the request. For books that are not particularly valuable or scarce, we check in other online catalogs to see if the same title might be in the circulating stacks of another library. We have lots of books (science fiction, utopias, or works by certain authors, for example) that are not particularly rare but that in OUR library are housed in Special Collections because of their subject or provenance. We wouldn't lend one of our utopian works if circulating copies are easily available from other institutions. We think of what it might mean to us if our book is damaged or lost. (We have on occasion lost books through lending.) There's a cost involved in two senses: What would it literally cost us to replace a book (if we could)? What would be the cost to our collections and researchers if we couldn't replace a book? If we are asked for a particularly scarce book, I look to see if there is a copy currently on the market. (This is often how we establish insurance values as well.) Could it be easily replaced? Was it given by a donor who would be angered by our having loaned it? Is it a key item within Special Collections that we couldn't afford to lose because it is so closely identified with us? Is the book unique (a signed copy or an association copy, for example), or does it have a particularly fine binding? We think about our local use patterns. We have an unwritten policy not to lend county histories or county atlases, for example, because they are so often consulted in our own reading room. It would be a hardship to our users (particularly genealogists on the road) to come here to find that a book that is supposed to be non-circulating is at another institution.

I hope someone thinking like this will be processing my own ILL request for special collections materials.

The survey closed with the open-ended question, "Is there anything you'd like to tell us about sharing special collections materials that wasn't addressed by the survey, or any point you'd like to emphasize?" We received 22 responses. A few mentioned that their institutions do not lend special collections and have no plans to review their policies; others lauded the increasing emphasis on access; a few wanted to hear more about the experiences of those institutions that are successfully lending entire archives. One respondent pushed the idea of digitizing as much as possible and making it available online as the best means of providing access. Another wrote, "We receive for our patrons materials that are similar to items we would not provide."

In other words, the survey showed us what we suspected already: that there is currently no consensus on any aspect of sharing special collections.

Survey Implications

When community practice is all over the map or split down the middle, the time is ripe for someone with a strong point of view to step forward and lay out a prospective path for that community. The Sharing Special Collections Working Group studied the survey results and decided to leap into the void.

Taking the "Scare" Out of Lending the Scarce

For members of our working group, volunteering for this assignment meant continually having to confront their own fears about the physical lending of special collections materials for research purposes. (See appendix 1, a case study of Pennsylvania State University staff involvement in this process, for an explicit example.) These fears became more manageable as we built up a core set of working principles.

Principal #1: The appropriate answer is still usually “No.”

No one is going to lend *The Book of Kells*—except perhaps for the most major exhibitions. In all cases, making a surrogate of the item, digital or otherwise, will be the first option in answer to an external request to borrow something from special collections. But will a surrogate be useful to the researcher?

Principle #2: Not all special collections material is equally special.

Many items are in special collections not because they are rare or valuable, but because they were written by a certain author or type of author, about a certain place or topic, in a specific genre, or at a certain time. Some special collections items, such as transcripts of oral history tapes, are easily replaceable and even easier to copy.

Principle #3: Not all requests for special collections material are created equal.

Often a borrower doesn't realize that a requested item is held in special collections. Sometimes the researcher really needs to see the original; in other cases, a copy of only part of the material will suffice. A researcher may be under a crippling deadline or may have all the time in the world. Sometimes the requested item is absolutely critical; at other times the researcher is merely satisfying an idle bit of curiosity and wouldn't want to put anyone to any special trouble.

Principle #4: Interlibrary loan staff knows how to lend things and get them back safely.

It's what they do. They are meticulous. They've spent decades perfecting infrastructure and techniques. They're aware that existing national and international ILL codes serve as implied contracts that cover any and all interlending transactions. They know how to double back to the requester and find out exactly what is needed. They established the community practice of sending out a surrogate instead of lending the original. They're experienced in making sure material is handled properly. With a little coaching, they can expertly handle even the rarest or most fragile material. They will be judicious about when it's time to confer with special collections experts. In short, special collections and archives staff can trust them.

A Tiered Approach

With these principles in mind, working group members set about creating a flexible system for considering loans of special collections (See figure 3.). A flexible approach acknowledges differences in user needs, collections, institutions, and resources. As always, institutions will bring to bear professional judgment regarding when to scale up effort and investment. Delivery of special collections material, whether of the actual item or a surrogate, is the goal, no matter the combination of tiers chosen.

We borrowed the concept of three tiers, or three levels of effort and overhead, from the work presented by Jennifer Schaffner, Francine Snyder, and Shannon Supple in their April 2011 OCLC Research report, *Scan and Deliver: Managing User-Initiated Digitization in Special Collections and Archives*. We listed the main steps in processing external requests for research loans of special collections: review, decide, lend, and return. Next we laid out tiers with three distinct levels of effort and overhead that may be chosen and combined based on decisions about the value, condition, rarity, format, rights status or popularity of the requested item; the identity, location, and controlled environment of the borrowing institution; the status, needs, and point in the research process of the researcher; and the policies, staff capabilities, and available resources of both institutions. Knowing what questions to ask and which level of staff to involve at each stage of the process are important first steps in streamlining processes, establishing effective communication among cooperating departments, and ensuring appropriate handling for materials regarded as “special.”

	Routine Workflow	Cooperative Workflow	Exceptional Workflow
REVIEW			
Request	Via ILL system	Collaboration between Special Collections (SC) and ILL	Directly to SC
Is material held in a special collection?	ILL staff	Collaboration between borrowing and lending institutions	Lending institution
Reference Interview	At borrowing institution—reference desk and ILL staff	Collaboration of ILL and SC staff in both institutions	By lending institution—SC staff
Inter-institutional communication how?	ILL system	ILL system and email/phone	Direct contact between two SC's
Internal communication how?	ILL system	ILL system and email/phone	Direct contact between SC/ILL staff and other departments
Stipulate for Research Use?	Implicit	Consider emphasizing	Explicit criteria
Reviewing Infrastructure	Written guidelines	Collaboration between borrowing and lending departments	Elaborate decision tree, multiple staff, institutional level decision
Mutual disclosure of ILL and SC facilities	We trust you	Approved checklist	Facilities report
Forms	ILL transaction work form and IFM	Extra insurance and/or forms for special handling	Use agreement, insurance forms, art museums loan agreement, etc.
DECIDE			
Decision Maker	ILL staff	ILL and SC consult when necessary	SC staff, curator, possibly director
Original or Surrogate?	Surrogate or predetermined originals	Prefer to lend surrogate, consider original	Case-by-case consideration

Table continued on next page

	Routine Workflow	Cooperative Workflow	Exceptional Workflow
Published/unpublished?	Some published and predetermined unpublished material types	Some published OK. Unpublished material on a case-by-case basis	Consider lending published and unpublished materials
Use Rights	Borrower's responsibility	What any reasonable SC staffer would do	Search, monitor and control thoroughly
Trust and Training	ILL training and expertise	ILL and SC cross-training on handling fragile materials	SC training and experience only
LEND			
Oversees loan transaction	ILL staff	Staff in ILL and SC	SC specialists
Quality Control	Usual packager, usual shipper, mailroom or ILL	Special ILL or SC packager	SC/preserv staff prepare special supports and deliver with the material
RETURN			
Deliver	Usual shipper, with use/handling conditions	Expedited shipper, extra insurance, special handling instructions	Deliver from SC to SC—call me when you get it

Figure 3. Tiered approach to sharing special collections, with varying degrees of effort and staff involvement

The Way We Were—and the Way We Could Be

While the survey results indicate condition of the material as the primary consideration when deciding whether to lend a special collections item, discussion among working group members revealed that the dominant factor for determining which tiers one will use is attitude toward risk.

Every research request used to be exceptional. Longstanding practice for those institutions that considered lending items from special collections was to treat each request according to the far-right “exceptional” tier. Most often requests were received directly by special collections staff; indeed, if the ILL office received a request for an item in special collections, common practice was to respond negatively and advise the borrowing institution to contact the special collections department directly. Often multiple staff members consulted about whether to lend the item. Typically special collections staff contacted the borrowing institution to talk about the patron’s needs and the borrowing staff’s ability to handle a loaned special collections item professionally. In some cases, use agreements and special insurance arrangements were required before a special item was be loaned. Preservation staff sometimes contributed special containers and support structures to protect the material while on loan.

Surely, working group members reasoned, there must be another kind of workflow appropriate to processing such requests. Surely there must be whole classes of special collections holdings about which an interlibrary loan person could be relied upon to make lending decisions, beyond a blanket negative. Perhaps there could even be middle-ground just beyond the obvious cases that could be decided cooperatively; special collections and interlibrary loan staff could come to an understanding about classes of material where a minimal amount of consultation would be appropriate, not necessarily to the level of bringing in curators or directors every time, and always with an eye toward providing a surrogate rather than lending the actual item whenever a copy would be sufficient. Surely a system could be put in place where the deluxe take-no-chances approach is saved for those few situations that actually require it.

Take a look at the tracks in figure 3. Think about the mindset at your institution, the prevailing attitude toward lending special collections originals, the tolerance for risk. Meet with your colleagues in special collections and interlibrary loan. What classes of material make sense for each track at your shop? For what material does it make sense to blend tracks, taking some steps in the Cooperative Workflow and others in the Routine Workflow? What materials push you outside your comfort level? What do you do when that happens? Proceed directly to the Exceptional Workflow?

Remember to breathe.

Then have the conversation about the tracks again.

The purpose of this report is to bring you to tiers.

Tool 1: Model Local Policy for Lending Special Collections Materials for Research Purposes

Working group members were excited when nearly a third of the respondents to our survey reported having developed a written policy statement that guided the sharing of items from their special collections. While no single institution possessed the kind of comprehensive policy statement that we sought, enough survey respondents provided examples of written policies on sharing particular formats that we were able to borrow the language needed to develop a comprehensive model policy statement on sharing special collections. The lion's share of this work was done by OCLC Research Program Officer Jen Schaffner, and working group member Scott Britton (then at the University of Miami). Our approach was to provide a multiple-choice template that special collections staff could customize for local use, adding and deleting elements to fit local practice.

Lending and Borrowing Special Collections for Research Purposes: Model Local Policy

Mission statement [example; add or delete as needed]:

The [institution name] Special Collections unit supports an active program of loans from its collections. We take local demand for special collections into consideration when deciding whether or not loan. The benefit of increased public access to its collections is measured against internal programs and the demands of preparation, packing, and transportation, with special consideration to the physical conditions of the work must endure throughout the loan. Accordingly, all loan requests are subject to a formal approval procedure. All requests are [considered.] [considered on merit.] [considered for their contribution to scholarship/human knowledge.] [considered for their public purpose.] [etc.]

Formats [add or delete as needed]:

- Formats of materials that will be considered for loan include: [microforms], [rare books], [manuscripts], [maps], [archives], and [videos] [etc.].
- Items and collections for loan must be in stable condition that will not be damaged by the move, change of environment, or even supervised handling by the Borrower.
- Items that are fragile, expensive or oversized may circulate with special packaging, handling instruction and insurance.

Requests [add or delete as needed]:

- Inquiries regarding Interlibrary Loan policy and procedures for special collections should be directed to [ILL email or special collections email] or by telephone at [phone number].
- Researchers must channel loan requests through a qualified institution [university or college library, historical society, public library, archives, museum, etc.].
- Preliminary research concerning a request should be carried out well in advance so that the formal request can be made in a timely fashion.
- Requests accepted via: [ALA], [OCLC], [fax], [email], and [telephone].
- The preferred requesting method is [ILL system] [link to forms][extraordinary circumstances and forms].
- The institution charges what is charged for ILL, except in extraordinary cases. Any preparation requested by the Borrower or required by the Lender which is at variance with normal practice will attract additional charges. These will be negotiated on a case by case basis. Additional shipping/insurance costs may also be charged.
- Unless otherwise specified in writing, all works will be released from and returned to [your mailing information here].

Terms and conditions of loan [add or delete as needed]:

- No item may be re-loaned by Borrower to a third party.
- The borrowing period shall be for [x days or weeks] with a [x days or weeks] renewal period.
- Long-term loans will be considered on a case-by-case basis.
- A researcher may borrow up to [x] items at one time and may not request additional loans until previously borrowed items have been returned.
- Researchers must be in good standing at their home institution.
- In the event that there is a local request for the loaned material, it will be recalled.
- The work must be stored in a space equipped to protect it from fire, smoke, or flood damage; under 24-hour physical and/or electronic security; and protected from humidity and temperature extremes, excessive light, and from insects, vermin, dirt, or other environmental hazards.
- No statement of valuation will be given an item in any manner to individuals or to the general public.
- The Loanning institution recognizes that a Borrower may cancel a loan, or other circumstances may prevent the loan from taking place as planned. Once remitted, loan-processing fees are non-refundable, regardless of circumstance.

Terms and Conditions of use [add or delete as needed]:

- All loaned materials must be used in the Borrowing library, in a reading room monitored by special collections staff.
- Staff of the Borrowing institution will ensure that the lender's regulations for use of [rare books, manuscripts, special collections, photographs and/or archives, etc.] are enforced during the loan period.
- Researchers must handle materials gently, taking care in a manner that avoids damage and excessive wear and tear.
- Permission for reproduction, including electronic formats, must be obtained from the Loanning institution. Permission may also need to be obtained from the copyright holder, if any.
- For specific digitization and publication use questions, please [visit the website] [contact staff].
- Each reproduction must be labeled and credited to the Loanning institution [as specified].
- Some material may not be available for reproduction due to preservation, copyright or other permission restrictions.

Packing, shipping and handling [add or delete as needed]:

- Only qualified staff may unpack, handle and repack the work/s.
- Any instructions given by the Lending institution regarding unpacking, handling and repacking are to be followed.
- The Borrowing institution will keep the packing materials for return shipment, and the work will be repacked using the same protective methods and materials.
- The Lending and the Borrowing institutions will ship the materials by a courier with tracking capabilities, such as UPS or Federal Express.
- The Borrowing institution may be required to bear costs associated with the shipping of the work/s including crating, packing, transportation, etc., in both directions.
- The Borrowing institution is responsible for returning the materials in the same condition as received.
- No work may be altered, cleaned, or repaired without prior written permission.
- Any damage, deterioration or loss to the work/s must be reported to the Lending institution immediately. The work/s should not be moved or treated until further instruction from the institution unless necessary to prevent further damage.
- If irreparable damage or loss occurs at any time, the Borrowing institution must meet all costs of replacement, or appropriate compensation.

Tool 2: The SHARES Facility Trust Checklist

Staff at institutions that lend physical items from special collections report that, upon receiving a borrowing request from another library, they often pick up the phone and initiate a conversation with the special collections or interlibrary loan practitioner at the borrowing library. In 2011, a SHARES working group compiled a set of core questions that the prospective lender typically asks of the borrower during such a conversation. This work was led by Aimee Lind of the Getty Research Institute.

The aim was to establish a set of core criteria that, when met by an institution requesting special collections material, will allow the curator to lend with confidence that the material will be handled safely and professionally. The 2012 SHARES Executive Group agreed that providing a list of such criteria to prospective borrowers and lenders is a valuable first step in promoting the sharing of special collections materials.

Working group and SHARES Executive Group members identified these potential use cases for the checklist:

- For a borrowing institution to cite compliance in interlibrary loan requests for special collections materials, as an indication to lenders that the material will be handled safely and professionally.
- For a borrowing institution to use to convince its own administration that upgrades in facilities and professional competencies are required in order to borrow materials essential to researchers.
- For a lending institution to send to a prospective borrowing institution that has requested special collections material through interlibrary loan, to confirm that the borrower has the facilities and competencies necessary to ensure safe handling of the borrowed item.
- In cases where the borrowing institution does not meet all the criteria, to use as a “conversation starter” with prospective lenders who may be willing to be flexible or to provide certain classes of material if a subset of the criteria are met.

SHARES Facility Trust Checklist: Baseline Criteria for Sharing Special Collections Materials

Institution name and address _____

Contact info for ILL _____

Contact info for Special Collections _____

- 1._____ My institution employs staff trained in handling special collections materials.
- 2._____ My institution maintains a supervised and secure reading room.
- 3._____ My institution's supervised reading room is climate-controlled.
- 4._____ My institution has a locked storage area or vault for housing special materials.
- 5._____ My institution's locked storage area or vault is climate-controlled.
- 6._____ The bags of those leaving my building are inspected, and/or patrons are required to leave bags in a locker before visiting special collections.
- 7._____ My institution's special collections area has intrusion detection equipment.
- 8._____ My institution's special collections area has a fire detection system.
- 9._____ My institution's special collections area has a fire suppression system.
- 10._____ My institution has insurance covering loss of borrowed materials due to damage or theft.
- 11._____ My building has a secure mail receiving room.
- 12._____ Incoming and outgoing special collections materials are received, unpacked, packaged, and shipped by staff trained in handling special collections materials.

Appendix I: Case Study—Pennsylvania State University

(This account draws heavily upon materials prepared by Sandra Stelts and Barbara Coopey; see list of references at the end of this report for specifics.)

- Who: Sandra Stelts, Curator of Rare Books and Manuscripts
Barbara Coopey, Assistant Head, Access Services; Head, Interlibrary Loan
Pennsylvania State University Libraries
- What: Embraced the idea of considering requests for loans of their special collections materials, including unpublished material.
- When: After attending the 2009 OCLC Research webinar, *Treasures on Trucks and Other Taboos: Rethinking the Sharing of Special Collections*, organized by the group that did the work described in this report. (Schaffner and Massie 2009)
- Why: To quote Sandra, “We became alternately intrigued and alarmed by the suggestion that special collections curators should consider lending more and more materials—including original archival and manuscript collections. Such loans on the surface seem contrary to our perceived mission and have put special collections curators’ desire to protect unique material at odds with interlibrary loan librarians who want to fulfill these requests for these materials.”
- How: Joined the OCLC Research Sharing Special Collections Working Group, helped to develop practices to streamline the process of sharing special collections materials, and then applied these concepts to improve their own workflow.

In applying the thinking of the working group to the situation at their home institution, Sandra and Barbara found that the following questions particularly resonated with Penn State’s concerns:

- Collections are for use; how can we share?
- Does the user know the material is in a special collection?

- Should the request go to Special Collections directly or through ILL?
- How does the lending institution staff determine that the requester actually needs the special material?
- How do we build trust—not only between borrowing and lending institutions but also between Special Collections and ILL?
- What can be loaned under what circumstances?
- What can be digitized and added to the collections for others to access and use?

One almost paralyzing worry was that Penn State would be overwhelmed with requests for materials held in their Special Collections, especially items they considered special because of subject or provenance, but that other institutions would keep in their general collections. A real breakthrough for Penn State came during an advisory group conference call when Eleanor Brown, then of Cornell University, reported that her ILL department sends a conditional response to all who request special collections material through interlibrary loan: “This item is held in our Special Collections. If you cannot locate this material elsewhere, please try us again.” Once more, quoting Sandra, “It’s so simple, and it has helped us to focus on the requests that are unique to our institution. We have also asked our own ILL staff to tell us when Penn State is the only location on a request. We know to take those requests particularly seriously and to make every effort to lend or make surrogates.”

Liberated by the “conditional response” strategy, Penn State staff proceeded to examine the workflow between ILL and Rare Books and Manuscripts to ensure careful transport of material between the units. They acquired distinctive tubs (See note, figure 4) that both protected special material while in transit and set it apart from other items being moved in and out of ILL. (This has led to some instances of “tub envy” from staff of other Special Collections units; after some quiet negotiations, archival materials being handled for ILL purposes are now permitted to ride in the same tub as Rare Books and Manuscripts materials.)

Detail	History	OCLC	Z39.50	PubMed/Dcline																																																																																																																																																				
General Request Information <table border="1"> <tr> <td>Transaction Number</td> <td>1500066</td> <td><input type="radio"/> Article</td> <td><input checked="" type="radio"/> Loan</td> </tr> <tr> <td>Username</td> <td>UPM</td> <td>Service Type</td> <td colspan="2"></td> </tr> <tr> <td>Transaction Date</td> <td>11/2/2012 11:17 AM</td> <td>Not Wanted After</td> <td colspan="2">06/28/2012</td> </tr> <tr> <td>Delivery Method</td> <td></td> <td>Site</td> <td colspan="2"></td> </tr> <tr> <td>Service Level</td> <td>Regular</td> <td>Shipping Options</td> <td colspan="2"></td> </tr> <tr> <td>Billing Account</td> <td></td> <td>Doc Type</td> <td colspan="2"></td> </tr> </table>					Transaction Number	1500066	<input type="radio"/> Article	<input checked="" type="radio"/> Loan	Username	UPM	Service Type			Transaction Date	11/2/2012 11:17 AM	Not Wanted After	06/28/2012		Delivery Method		Site			Service Level	Regular	Shipping Options			Billing Account		Doc Type																																																																																																																									
Transaction Number	1500066	<input type="radio"/> Article	<input checked="" type="radio"/> Loan																																																																																																																																																					
Username	UPM	Service Type																																																																																																																																																						
Transaction Date	11/2/2012 11:17 AM	Not Wanted After	06/28/2012																																																																																																																																																					
Delivery Method		Site																																																																																																																																																						
Service Level	Regular	Shipping Options																																																																																																																																																						
Billing Account		Doc Type																																																																																																																																																						
<table border="1"> <tr> <td>Article Info</td> <td>Loan Info</td> <td>Collections</td> <td>Local Holdings</td> <td>Imported Request</td> <td>Copyright</td> <td>Invoice</td> <td>Adj.</td> <td>></td> </tr> <tr> <td colspan="9"> <table border="1"> <tr> <td>Title</td> <td colspan="8">Phantom Africa manuscript and letters - Please return box with book</td> </tr> <tr> <td>Author</td> <td colspan="8">Leiris, Michel, 1901-1990.</td> </tr> <tr> <td>Publisher</td> <td colspan="8"></td> </tr> <tr> <td>Place</td> <td colspan="8"></td> </tr> <tr> <td>Date</td> <td colspan="8"></td> </tr> <tr> <td>Edition</td> <td colspan="8"></td> </tr> <tr> <td>Original Loan Author</td> <td colspan="8"></td> </tr> <tr> <td>Original Loan Title</td> <td colspan="8"></td> </tr> </table> </td> </tr> <tr> <td colspan="5"> <table border="1"> <tr> <td><input type="checkbox"/> Accept Alternate Edition</td> </tr> <tr> <td><input type="checkbox"/> Accept Non English</td> </tr> <tr> <td><input checked="" type="checkbox"/> Allow Copies?</td> </tr> <tr> <td><input type="checkbox"/> Copyright Already Paid?</td> </tr> <tr> <td><input type="checkbox"/> Allow Renewals?</td> </tr> <tr> <td><input checked="" type="checkbox"/> Library Use Only?</td> </tr> <tr> <td><input type="checkbox"/> Replacement Pages?</td> </tr> <tr> <td><input type="checkbox"/> Priority Shipping</td> </tr> <tr> <td><input type="checkbox"/> Ariel</td> </tr> <tr> <td><input checked="" type="checkbox"/> Odyssey (PDF Enabled)</td> </tr> </table> </td> <td>Item Information</td> </tr> <tr> <td colspan="5"></td> <td> <table border="1"> <tr> <td>Call Number</td> <td>1986-0080R PS-V-MS-64 Vault</td> </tr> <tr> <td>Location</td> <td>RBR VAULT</td> </tr> <tr> <td>Due Date</td> <td>8/17/2012</td> </tr> <tr> <td>Reason For Cancellation</td> <td></td> </tr> <tr> <td>Item Num/Ref Num</td> <td></td> </tr> <tr> <td>ISXN</td> <td></td> </tr> <tr> <td>Special Instructions</td> <td></td> </tr> <tr> <td>Max Cost/Pieces</td> <td>100.00JFM</td> </tr> </table> </td> </tr> <tr> <td colspan="5"></td> <td>Citation Information</td> </tr> <tr> <td colspan="5"></td> <td> <table border="1"> <tr> <td>Cited In</td> <td></td> </tr> <tr> <td>Title</td> <td></td> </tr> <tr> <td>Date</td> <td></td> </tr> <tr> <td>Volume/Pages</td> <td></td> </tr> </table> </td> </tr> </table>					Article Info	Loan Info	Collections	Local Holdings	Imported Request	Copyright	Invoice	Adj.	>	<table border="1"> <tr> <td>Title</td> <td colspan="8">Phantom Africa manuscript and letters - Please return box with book</td> </tr> <tr> <td>Author</td> <td colspan="8">Leiris, Michel, 1901-1990.</td> </tr> <tr> <td>Publisher</td> <td colspan="8"></td> </tr> <tr> <td>Place</td> <td colspan="8"></td> </tr> <tr> <td>Date</td> <td colspan="8"></td> </tr> <tr> <td>Edition</td> <td colspan="8"></td> </tr> <tr> <td>Original Loan Author</td> <td colspan="8"></td> </tr> <tr> <td>Original Loan Title</td> <td colspan="8"></td> </tr> </table>									Title	Phantom Africa manuscript and letters - Please return box with book								Author	Leiris, Michel, 1901-1990.								Publisher									Place									Date									Edition									Original Loan Author									Original Loan Title									<table border="1"> <tr> <td><input type="checkbox"/> Accept Alternate Edition</td> </tr> <tr> <td><input type="checkbox"/> Accept Non English</td> </tr> <tr> <td><input checked="" type="checkbox"/> Allow Copies?</td> </tr> <tr> <td><input type="checkbox"/> Copyright Already Paid?</td> </tr> <tr> <td><input type="checkbox"/> Allow Renewals?</td> </tr> <tr> <td><input checked="" type="checkbox"/> Library Use Only?</td> </tr> <tr> <td><input type="checkbox"/> Replacement Pages?</td> </tr> <tr> <td><input type="checkbox"/> Priority Shipping</td> </tr> <tr> <td><input type="checkbox"/> Ariel</td> </tr> <tr> <td><input checked="" type="checkbox"/> Odyssey (PDF Enabled)</td> </tr> </table>					<input type="checkbox"/> Accept Alternate Edition	<input type="checkbox"/> Accept Non English	<input checked="" type="checkbox"/> Allow Copies?	<input type="checkbox"/> Copyright Already Paid?	<input type="checkbox"/> Allow Renewals?	<input checked="" type="checkbox"/> Library Use Only?	<input type="checkbox"/> Replacement Pages?	<input type="checkbox"/> Priority Shipping	<input type="checkbox"/> Ariel	<input checked="" type="checkbox"/> Odyssey (PDF Enabled)	Item Information						<table border="1"> <tr> <td>Call Number</td> <td>1986-0080R PS-V-MS-64 Vault</td> </tr> <tr> <td>Location</td> <td>RBR VAULT</td> </tr> <tr> <td>Due Date</td> <td>8/17/2012</td> </tr> <tr> <td>Reason For Cancellation</td> <td></td> </tr> <tr> <td>Item Num/Ref Num</td> <td></td> </tr> <tr> <td>ISXN</td> <td></td> </tr> <tr> <td>Special Instructions</td> <td></td> </tr> <tr> <td>Max Cost/Pieces</td> <td>100.00JFM</td> </tr> </table>	Call Number	1986-0080R PS-V-MS-64 Vault	Location	RBR VAULT	Due Date	8/17/2012	Reason For Cancellation		Item Num/Ref Num		ISXN		Special Instructions		Max Cost/Pieces	100.00JFM						Citation Information						<table border="1"> <tr> <td>Cited In</td> <td></td> </tr> <tr> <td>Title</td> <td></td> </tr> <tr> <td>Date</td> <td></td> </tr> <tr> <td>Volume/Pages</td> <td></td> </tr> </table>	Cited In		Title		Date		Volume/Pages	
Article Info	Loan Info	Collections	Local Holdings	Imported Request	Copyright	Invoice	Adj.	>																																																																																																																																																
<table border="1"> <tr> <td>Title</td> <td colspan="8">Phantom Africa manuscript and letters - Please return box with book</td> </tr> <tr> <td>Author</td> <td colspan="8">Leiris, Michel, 1901-1990.</td> </tr> <tr> <td>Publisher</td> <td colspan="8"></td> </tr> <tr> <td>Place</td> <td colspan="8"></td> </tr> <tr> <td>Date</td> <td colspan="8"></td> </tr> <tr> <td>Edition</td> <td colspan="8"></td> </tr> <tr> <td>Original Loan Author</td> <td colspan="8"></td> </tr> <tr> <td>Original Loan Title</td> <td colspan="8"></td> </tr> </table>									Title	Phantom Africa manuscript and letters - Please return box with book								Author	Leiris, Michel, 1901-1990.								Publisher									Place									Date									Edition									Original Loan Author									Original Loan Title																																																																																
Title	Phantom Africa manuscript and letters - Please return box with book																																																																																																																																																							
Author	Leiris, Michel, 1901-1990.																																																																																																																																																							
Publisher																																																																																																																																																								
Place																																																																																																																																																								
Date																																																																																																																																																								
Edition																																																																																																																																																								
Original Loan Author																																																																																																																																																								
Original Loan Title																																																																																																																																																								
<table border="1"> <tr> <td><input type="checkbox"/> Accept Alternate Edition</td> </tr> <tr> <td><input type="checkbox"/> Accept Non English</td> </tr> <tr> <td><input checked="" type="checkbox"/> Allow Copies?</td> </tr> <tr> <td><input type="checkbox"/> Copyright Already Paid?</td> </tr> <tr> <td><input type="checkbox"/> Allow Renewals?</td> </tr> <tr> <td><input checked="" type="checkbox"/> Library Use Only?</td> </tr> <tr> <td><input type="checkbox"/> Replacement Pages?</td> </tr> <tr> <td><input type="checkbox"/> Priority Shipping</td> </tr> <tr> <td><input type="checkbox"/> Ariel</td> </tr> <tr> <td><input checked="" type="checkbox"/> Odyssey (PDF Enabled)</td> </tr> </table>					<input type="checkbox"/> Accept Alternate Edition	<input type="checkbox"/> Accept Non English	<input checked="" type="checkbox"/> Allow Copies?	<input type="checkbox"/> Copyright Already Paid?	<input type="checkbox"/> Allow Renewals?	<input checked="" type="checkbox"/> Library Use Only?	<input type="checkbox"/> Replacement Pages?	<input type="checkbox"/> Priority Shipping	<input type="checkbox"/> Ariel	<input checked="" type="checkbox"/> Odyssey (PDF Enabled)	Item Information																																																																																																																																									
<input type="checkbox"/> Accept Alternate Edition																																																																																																																																																								
<input type="checkbox"/> Accept Non English																																																																																																																																																								
<input checked="" type="checkbox"/> Allow Copies?																																																																																																																																																								
<input type="checkbox"/> Copyright Already Paid?																																																																																																																																																								
<input type="checkbox"/> Allow Renewals?																																																																																																																																																								
<input checked="" type="checkbox"/> Library Use Only?																																																																																																																																																								
<input type="checkbox"/> Replacement Pages?																																																																																																																																																								
<input type="checkbox"/> Priority Shipping																																																																																																																																																								
<input type="checkbox"/> Ariel																																																																																																																																																								
<input checked="" type="checkbox"/> Odyssey (PDF Enabled)																																																																																																																																																								
					<table border="1"> <tr> <td>Call Number</td> <td>1986-0080R PS-V-MS-64 Vault</td> </tr> <tr> <td>Location</td> <td>RBR VAULT</td> </tr> <tr> <td>Due Date</td> <td>8/17/2012</td> </tr> <tr> <td>Reason For Cancellation</td> <td></td> </tr> <tr> <td>Item Num/Ref Num</td> <td></td> </tr> <tr> <td>ISXN</td> <td></td> </tr> <tr> <td>Special Instructions</td> <td></td> </tr> <tr> <td>Max Cost/Pieces</td> <td>100.00JFM</td> </tr> </table>	Call Number	1986-0080R PS-V-MS-64 Vault	Location	RBR VAULT	Due Date	8/17/2012	Reason For Cancellation		Item Num/Ref Num		ISXN		Special Instructions		Max Cost/Pieces	100.00JFM																																																																																																																																			
Call Number	1986-0080R PS-V-MS-64 Vault																																																																																																																																																							
Location	RBR VAULT																																																																																																																																																							
Due Date	8/17/2012																																																																																																																																																							
Reason For Cancellation																																																																																																																																																								
Item Num/Ref Num																																																																																																																																																								
ISXN																																																																																																																																																								
Special Instructions																																																																																																																																																								
Max Cost/Pieces	100.00JFM																																																																																																																																																							
					Citation Information																																																																																																																																																			
					<table border="1"> <tr> <td>Cited In</td> <td></td> </tr> <tr> <td>Title</td> <td></td> </tr> <tr> <td>Date</td> <td></td> </tr> <tr> <td>Volume/Pages</td> <td></td> </tr> </table>	Cited In		Title		Date		Volume/Pages																																																																																																																																												
Cited In																																																																																																																																																								
Title																																																																																																																																																								
Date																																																																																																																																																								
Volume/Pages																																																																																																																																																								

Figure 4. ILL transaction record of Pennsylvania State University lending a manuscript to Columbia University

They reviewed paperwork that accompanies loaned material, including instructions for shipping, insurance, and safe handling. They increased the number of filled requests by scanning with an overhead scanner to protect fragile material. They made paper “preservation” copies of fragile items under the copyright law’s fair use provisions and lent the copy. They improved measures to ensure the safety of room-use-only materials borrowed from other institutions—as well as their own materials—by moving the photocopier to a location directly next to the reference desk to ensure more direct staff supervision and compliance with no-photocopying rules. Other renovations to the reference and reading rooms will improve sight lines from the reference desk, and the security cameras have been upgraded.

Sandra and Barbara soon learned that trust was the key ingredient in the sharing of special collections material; as Barbara put it, “Trust should exist not only between borrowing and lending institutions but also between ILL and Special Collections.” Penn State staff put major effort into building trust between the ILL and Special Collections units by increasing communication and paying more attention to the process. Both units now better understand the concerns and needs of the

other and, in fact, find that they share many of the same needs and concerns, such as effectively balancing the pressing needs of researchers with the library’s imperative to protect the material.

Barbara recently had a query from a librarian in Japan who wanted to know what sort of security the Penn State library offered in the reading room before deciding to lend them a book. Barbara just happened to have photos taken for a presentation about the Sharing Special Collections working group and was able to document the layout and security regime of the reading room. The librarian in Japan loaned the book. Working group members agreed that having such photos on hand to share discreetly during the course of an ILL transaction would be quite useful.

Once involved with lending special collections materials via ILL, Penn State staff warmed to the task. They discovered early on that it was often useful to be in touch directly with the other library’s patron to find out exactly what was required, and how vital the need actually was. Sometimes they could satisfy a researcher’s request by simply photocopying a table of contents, or a single chapter, or an illustration, rather than lending the whole book. Staff discovered that

researchers were often sensitive to the curators' concerns, saying "Oh, never mind, I wouldn't want you to ship a rare book" or "I can try to find it on my next trip to Italy—let's hold off for now." When the need was truly urgent and could not be satisfied from other sources, Penn State staff went to great lengths to find a way to fill it.

All of this work building upon the accomplishments of the Sharing Special Collections Working Group has, in Sandra's opinion, led to an increased alignment of the Penn State library with the institutional mission. And to quote her one last time: "I bask in praise after a successful transaction, such as 'Oh! You are just too good!'"

Appendix 2: In-depth Analysis of the 2010 Sharing Special Collections Working Group Survey

In April 2010, the working group conducted a survey in order to solicit current attitudes, practices, policies, and priorities regarding the lending of special collections materials for research purposes. We cast a wide net, announcing the survey on major primary sources and interlibrary loan discussion lists and inviting any library with a special collections department to reply. Survey instructions encouraged respondents to have ILL and special collections staff members work together in answering the questions.

We received 88 responses overall, with 64 completing the entire survey. Types of responding institutions varied greatly and included academic, national, and public libraries, plus museums and historical societies. The overwhelming majority of responses came from North America, with four from continental Europe and one each from Africa and Australia. Respondent job titles included a mix of special collections curators, reference or access heads, and interlibrary loan supervisors, along with a few university archivists and library directors.

ILL Lending Policies for Special Collections Materials

The first section of the survey focused on policy. It was impossible to find any sort of consensus in the responses, other than one sizable camp being willing to lend even its treasures to trusted partners, while another sizable camp expresses an aversion to risk that at times sounds more like fear.

Over two dozen respondents (36.8% of the total) claimed to have a written policy on lending special collections. When the working group followed up, however, we found that not a single institution had an overall written policy covering all special collections and archives. Rather, they had a written policy on some aspect of sharing, such as microfilms or digitizing out-of-copyright materials. In the end, we borrowed language from several of these narrowly-focused policies to create a model overall policy for sharing special collections materials. (See tool 1.)

Nearly half of respondents (48.5%) have different policies for lending published special collections materials than for unpublished, while 35.3% do not, and 16.2% “sometimes” have different policies. Comments revealed that many have the “same” policies for both because they do not lend anything from special collections, published or unpublished. The differences in policy usually centered around being sometimes willing to lend published material but not unpublished, or to copy published material but not unpublished. There was no consensus.

Only 10.3% of respondents indicated that they lend physical items from special collections to other libraries, with another 57.4% reporting that they will do so “under certain conditions,” for a total of 67.7% who share physical items at least sometimes. About a third (32.3%) never lend physical items from special collections. Comments revealed that many lend only for exhibition, others only to fellow participants in the SHARES resource sharing program, still others only published materials. One library reported experimenting with loans of entire archival collections to other libraries within their state. Again, there was no consensus on best practice.

Those who **do lend** physical items from special collections to other libraries were asked to choose the top three issues involved in the decision-making, from a list of eight that included “Other—please specify.” By far the most important issues were “condition of item” (noted by 87.2% of respondents), “rareness of item” (mentioned by 78.7%), and “value of item” (noted by 59.6%). No other answer—age of item, identity of requester, location of requester, how busy we are, or other—was chosen by more than a quarter of respondents. “Other” choices put forward included how heavily the item is used at the home institution, the value to the home institution aside from monetary value, and the quality of the environmental conditions at the borrowing library.

Those who **do not lend** physical items from their special collections to other libraries were asked to rank the reasons why they don’t, from a list

of seven that included “Other—please specify.” The most popular reasons were “too risky” (69.2%), “other” (51.3%), and “items needed onsite (30.8%). “Other” reasons included “items are irreplaceable,” “have loaned previously and gotten back damaged items,” “resistance on the part of special collections staff,” “resistance on the part of branch managers,” and, my personal favorite, “an atmosphere of mistrust and fear.” As previously mentioned, the reasons “not part of our mission” and “because we never have” each was chosen twice as often as “lack staff resources,” which the working group members had anticipated being an oft-cited reason for not lending.

Nearly half of respondents (47.8%) reported lending surrogates of special collections materials to other libraries, while another 35.8% said that they do “under certain conditions,” for a total of 83.6% lending surrogates (compared with 67.7% lending physical items). Only 16.4% reported not lending surrogates of special collections materials to other libraries. Comments centered mostly on the condition of the original item and the proportion of the work being requested. One respondent wrote, “We desire to keep our collections, and make our repository valuable to researchers, so we don’t create duplicate collections for storage by other repositories.”

Those who **do supply** surrogates of special collections materials were asked how they supplied them, choosing all methods that apply from a list of five, including “Other—please specify.” “Scan and send as file” (81.0%) and “photocopy and provide hard copy” (74.1%) were by far the most popular methods, with “scan, add to own digital collection, and provide a link” (41.4%) being the only other choice cited by more than a quarter of respondents.

Those who **do not supply** surrogates of special collections materials were asked why not, with up to three reasons to be chosen from a list of seven that included “Other—please specify.” “Risk of damage to material” (52.6%) and “Other” (47.4%) were the only choices selected by more than a third of respondents. Comments indicated that some respondents interpreted the question as being specifically about providing a surrogate of the entire special collections item, and they either lacked the resources to do so or felt that such a request would violate copyright in most cases.

Workflows for Managing ILL Requests for Special Collections Materials

The divide in the community about sharing special collections materials continued when we looked at workflows for managing incoming requests. More than half (57.1%) have interlibrary loan staff manage library-to-library requests for special collections materials, while 9.5% manage such requests in special collections; 33.3% receive and manage such requests in both departments.

By far the most popular method for managing and tracking ILL requests for special collections materials was ILLiad (41.0%), the ILL management software created by Atlas Systems, with the next popular being paper files (16.4%). Other methods included spreadsheets, integrated library systems, and Clio, an ILL management package designed by Clio Software. One respondent reported using Aeon, an online request system for archives and special collections designed by Atlas Systems.

In response to an open-ended question about how incoming ILL requests for special collections materials are “triaged,” answers varied from “we don’t lend” to “we only lend within our consortium” to “we check with the archivist” to “the director reviews the request.” The preferred method seemed to be related to the size of the staff handling requests and the volume of requests coming in; busier places saw more of a need to automate and streamline processes; at less busy places or sites where one staff member handles all incoming requests, procedures were more informal, epitomized by the comment, “When I get a request for special collections materials, I set it aside until I have a minute to go and see the archivist.”

In answer to a question about the preferred modes of communication among staff processing and reviewing ILL requests for special collections materials, the most popular were “email” and “face-to-face” (36.1% each). Only 13.1% use the ILLiad client for such communication. Very few use paper forms and the telephone. One commenter emphasized, “The answer is no. Always.”

When asked an open-ended question about the effectiveness of current procedures for handling incoming ILL requests for special collections materials, most expressed satisfaction. A few suggested that more automation would be helpful,

while others noted that key staff outages can throw the system into disarray. One respondent (an ILL staff person) expressed surprise at learning through completing the survey that special collections staff often received and filled requests directly without ILL involvement.

Workflows for Processing ILL Requests to Physically Lend Special Collections Materials

The next section of the survey focused on workflows associated specifically with lending physical items from special collections to other libraries. Most respondents (59.6%) indicated that a curator makes the final decision about whether a particular item will be physically loaned, with ILL staff making the decision in only 11.5% of the responses; at nearly a third of the surveyed institutions (28.8%), it is a group decision. In response to an open-ended question about how the decision is made to lend or not to lend, we mostly received confirmation of what was learned from the multiple-choice questions about basic processing of requests for special collections materials: condition matters most, with other factors such as rarity, value, popularity, and proximity (of the requester to the supplier, or of the requester to other copies of the same material) carrying significant weight.

Most potential lenders of physical items from special collections don't require any specific knowledge ahead of time about the borrowing patron, with a few respondents asking to know the name and/or patron status. In response to an open-ended question about what potential lenders of such material might want to know ahead of time about the borrowing institutions, most mentioned the security and environmental controls in place, or the presence of professional staff to supervise use of the items. A few would want to know if the borrowing institution was a fellow member of a consortium such as SHARES. A very few indicated that they would not lend special collections materials to a public library.

As for packaging special collections materials for loans to other libraries, respondents were almost evenly split between assigning this task to ILL staff (34.0%) and special collections staff (30.0%). Only 10.0% of respondents delegated such packaging to the mail room. "Other, please specify" responses comprised more than a

quarter of the total (26.0%); they varied from a division of labor (conservator makes special boxes, ILL staff does packaging) to a case-by-case approach based on condition or format.

Workflows for Processing ILL Requests for Surrogates of Special Collections Materials

The final section of the survey focused on workflows associated specifically with lending surrogates of special collections items to other libraries. As with the decision-making process for lending special collections items themselves, most respondents (44.8%) indicated that a curator makes the final decision about whether a surrogate will be sent (compared with 59.6% having curators decide when loaning the actual item); ILL staff make the decision on lending a surrogate in 25.9% of the responses (compared to only 11.5% having ILL staff decide on lending the actual item). As with the decision-making for lending actual items, nearly a third of the surveyed institutions (29.3%) make providing a surrogate of a special collections item a group decision.

The survey closed with the open-ended question, "Is there anything you'd like to tell us about sharing special collections materials that wasn't addressed by the survey, or any point you'd like to emphasize?" We received 22 responses, mostly reaffirming points made elsewhere in the survey. A few mentioned that their institutions do not lend special collections and have no plans to review their policies; others lauded the increasing emphasis on access; a few wanted to hear more about the experiences of those institutions that are successfully lending entire archives. One respondent pushed the idea of digitizing as much as possible and making it available online as the best means of providing access. Another wrote, "We receive for our patrons materials that are similar to items we would not provide."

The survey showed us what we suspected already: that there is currently no consensus on any aspect of sharing special collections.

When community practice is all over the map or split down the middle, the time is ripe for someone with a strong point of view to step forward and lay out a prospective path for that community. The Sharing Special Collections Working Group studied the survey results and decided to leap into the void.

References

ALA (American Libraries Association). 2012. "ACRL/RBMS Guidelines for Interlibrary and Exhibition Loan of Special Collections Materials." Approved by the ACRL Board of Directors January 2012. <http://www.ala.org/acrl/standards/specialcollections>.

Coopey, Barbara, and Sandra Stelts. 2011. "SHARES Special Collections Workflow Group and Penn State's Efforts to Share." Presentation at the Association of College and Research Libraries Conference, Philadelphia, Pennsylvania, April 2011. <http://www.learningtimes.net/acrl/2011/when-worlds-collide-interlibrary-loan-and-special-collections/> (presentation starts at about the six minute mark)

Engst, Elaine. 2012. "Baseball, True Crime, the FBI and I(LL): Interlibrary Loan for Archival Collections Revisited." Presentation at the Society of American Archivists Annual Meeting, San Diego, California, August 2012. <http://files.archivists.org/conference/sandiego2012/508-Engst.pdf>

Erney, Richard A., and F. Gerald Ham. 1972. "Wisconsin's Area Research Centers." *American Libraries*. (February):135-140.

Schaffner, Jennifer, and Dennis Massie. 2009. *Treasures on Trucks and Other Taboos: Rethinking the Sharing of Special Collections*. Webinar produced by OCLC Research. 5 June. <http://www5.oclc.org/downloads/programsandresearch/parcasts/20090528Schaffner&Massie.mp4>.

Schaffner, Jennifer, Francine Snyder, and Shannon Supple. 2011. *Scan and Deliver: Managing User-initiated Digitization in Special Collections and Archives*. Dublin, OH: OCLC Research. <http://www.oclc.org/research/publications/library/2011/2011-05.pdf>.

Stelts, Sandra, and Barbara Coopey. 2012. "ARL Submission For Special Issue of Research Library Issues (RLI)." Unpublished proposal to submit a written account of Pennsylvania State University's efforts to share special collections materials, to the Association of Research Library's journal RLI.



For more information about our work related to research collections and support, please visit:
www.oclc.org/research/themes/research-collections.html



6565 Kilgour Place
Dublin, Ohio 43017-3395

T: 1-800-848-5878
T: +1-614-764-6000
F: +1-614-764-6096
www.oclc.org/research

ISBN: 1-55653-493-0
978-1-55653-493-5
1504/215371, OCLC