SPEC Kit 307

Manuscript Collections on the Web

October 2008

Donnelly Lancaster Walton
Archival Access Coordinator, W.S. Hoole Special Collections Library
University of Alabama
SURVEY RESULTS

Executive Summary ............................................................................................................. 11
Survey Questions and Responses ......................................................................................... 19
Responding Institutions ......................................................................................................... 109

REPRESENTATIVE DOCUMENTS

Manuscript Collection Web Sites
University of Alabama
    Manuscript Collections .................................................................................................. 114
Arizona State University
    Manuscript Collections ................................................................................................. 115
University of British Columbia
    Manuscript Collections .................................................................................................. 116
University of California, Santa Barbara
    Alphabetical List of Collections ...................................................................................... 117
University of Georgia
    Richard B. Russell Library for Political Research and Studies. Collections .................... 118
Indiana University Bloomington
    The Collections: Manuscripts ....................................................................................... 120
Johns Hopkins University
    Historical Manuscripts Collection .................................................................................. 121
Library of Congress
    Manuscript Reading Room ............................................................................................. 122
McMaster University
    Archives and Manuscripts .............................................................................................. 123
University of Massachusetts Amherst
    UMarmot :: The Catablog of SCUA ................................................................................. 124
National Library of Medicine
    Archives and Modern Manuscripts Program .................................................................. 125
Finding Aid Web Sites

University of Alabama
What is a Finding Aid? ................................................................. 138

University of British Columbia
sample list of fonds...................................................................... 140

University of California, Los Angeles
Finding Aid Search ........................................................................ 141

University of California, San Diego
Finding Aids for Manuscripts & UCSD Archives ......................... 142

Georgia Institute of Technology
Finding Aids ................................................................................ 143

Indiana University Bloomington
Indiana University Finding Aids .................................................... 144

University of Iowa
Finding Aids ................................................................................ 145

Johns Hopkins University
Historical Manuscripts ............................................................... 146

Library of Congress
Manuscript Division Finding Aids Online ..................................... 147

McMaster University
Fonds/Collections Descriptions & Finding Aids ......................... 148

University of Manitoba
Archives & Special Collections’ Holdings ..................................... 149
Rhode Island Archival and Manuscript Collections Online

About RIAMCO .............................................................................................................174

Rocky Mountain Online Archive

Welcome to the Rocky Mountain Online Archive .........................................................175
Browse by Institutions, Browse by Subjects ............................................................176

Texas Archival Resources Online

About TARO .................................................................................................................177

Virginia Heritage

Guides to Manuscript & Archival Collections in Virginia ..............................................178

Arrangement and Description Guidelines

Brigham Young University

Processing Case File Checklist ...................................................................................180

University of Washington

Accessioning, Processing & Encoding Manual for Manuscript and Archival Collections ....182

Processing for the Web

University of Missouri

Policy Number 10: Posting Files to the Web Server ....................................................188

University of California, Riverside

Workflow for Adding Manuscript Collection Information to the Web .........................191

New York State Library

Web Publication Checklist ............................................................................................192

SELECTED RESOURCES

Documents ....................................................................................................................197

Web Sites ....................................................................................................................197
SURVEY RESULTS
EXECUTIVE SUMMARY

Introduction
There is growing demand from users, administrators, and donors to have manuscript collection information available on the Web. In their OCLC Programs and Research report “Shifting Gears: Gearing Up to Get into the Flow,” Ricky Erway and Jennifer Schaffner state, “In a world where it is increasingly felt that if it’s not online it doesn’t exist, we need to make sure that our users are exposed to the wealth of information in special collections.”1 Their report speaks specifically to digitizing collections, but the philosophy holds true for information about manuscript collections as well as digital facsimiles of them. This study explores Web resources that provide information about these collections, rather than the facsimiles.

Many of those who are responsible for the arrangement and description of manuscript and archival materials suffer from chronic backlogs and often lament the lack of resources (staff and time) to deal with their workloads. How do libraries accomplish the task of getting information about their valuable resources online? This survey investigated how many manuscript collections are held in ARL member libraries; what percentage of these collections are represented on the Web; what types of information about the collections are available in finding aids and on the Web; what formats are used for finding aids on the Web; how many library staff are working on manuscript collections, the challenges and benefits of migrating collection information to the Web, and whether and how usage of manuscript collection information is tracked. It was distributed to the 123 ARL member libraries in February 2008. Seventy-two libraries completed the survey by the March 31 deadline for a response rate of 59%.

Background
Using the Society of American Archivists’ (SAA) definition of a manuscript collection as a “collection of personal or family papers”2, the survey first asked whether the library held manuscript collections. The majority of respondents (69 or 97%) answered, “Yes.” Many of the respondents clarified how their answers may not reflect all of the manuscripts held at their institution since these collections are dispersed across several units.

Staffing
All 69 respondents identified the unit, department, or library that is responsible for arranging and describing manuscript collections; in some cases, there is more than one. Typically, it is an archive, special collections, and/or rare books department or library. Sixty-seven respondents answered basic questions about the number and types of staff (including archivists, librarians, other professional staff, support staff, and students) in the unit. They reported a total of 1297 individuals: 584 student assistants, 265 support staff, 218 archivists, 152 librarians, 47 other professionals, and 31 other staff, such as volunteers. The FTE totals indicate that permanent staff largely work full-time and that there are about three students per FTE.

The total number of individuals per unit ranges from 4 to 95, with an average of 19.36. The number
of student assistants skews the average, though. The maximum number of permanent staff ranges from 5 (other professional) to 15 (support staff) while the student assistant maximum is an incredible 63 at one institution! The median number of professional and support staff per unit ranges from one to three, while the median for student assistants is six.

Staffing for Manuscript Collection Activities
Without arrangement and description there would be very little information to put on the Web for researchers to use, so the survey next asked how many of the total staff reported above actually perform arrangement and description tasks, how much time they spend on these activities, and how much time they spend adding manuscript information to the Web. The 68 respondents reported that 762 individuals perform these activities, 59% of the total staff in the department.

Archivists
Fifty-six respondents reported a total of 147 archivists who spend even a small percentage of time on arrangement and description. The number per institution ranges from 1 to 13, with an average of 2.63 individuals. The archivists spend anywhere from 1% to 100% of their time on these activities, averaging 41.17%. The two archivists who spend 100% of their time on arrangement and description were listed as a Project Archivist and Contract Processing Archivist, positions created specifically to focus on such activities.

On average, archivists spend 11% of their time adding information about collections to the Web. Therefore, the average archivist spends about 16 hours of a 40-hour work week on arrangement and description activities, plus another 4 hours getting information about these collections on the Web. In addition, they spend time on a variety of other areas including reference, instruction, preservation, acquisitions, records management, exhibits, management, professional service, digital projects, donor relations, collection development, and outreach. As one respondent succinctly put it, “The usual.”

Department Head
Thirty-five respondents reported that the unit or department head spends some time on arrangement and description or on adding collection information to the Web. The maximum time for either activity is 50%. On average, slightly more time is spent on arrangement and description (12.32%) than on adding information to the Web (8.95%).

Librarians and Other Professionals
Thirty-five respondents reported a total of 69 librarians who spend time on manuscript collection arrangement and description or adding information to the Web. The range is 1 to 5 per institution, with an average of 1.87 librarians. They spend up to 100% of their time on arrangement and description, averaging 25%. As with department heads, librarians spend less time putting information on the Web, up to 65% of their time but only averaging 12%.

There are fewer other professionals involved in manuscript collection processing activities (20 institutions reported a total of 29 individuals), but the range per institution (1 to 5, average 1.45) and time commitment (arrangement and description average 27%, Web average 12%) mirror librarians.

Support Staff
Almost as many support staff as archivists spend time on arrangement and description and Web activities; 47 respondents reported a total of 139 individuals. The number of support staff per institution ranges from 1 to 15, with an average of 2.96. They spend up to 95% of their time on arrangement and description, averaging 44%, the highest for permanent staff. They also spend up to 75% of their time on putting information on the Web, averaging 11%.

Students and Other Staff
All but ten respondents report employing student assistants for manuscript collection arrangement and description work. Students account for the largest number of staff (317) and highest percentage of time devoted to these activities (64%). Although libraries depend heavily on students for arrangement and
description work, they do not depend on them for getting that same information on the Web. Student assistants spend up to 100% of their time in this capacity, but average only 9%.

Only 14 libraries reported employing other types of staff for manuscript activities. Typically, these are graduate assistants, student interns, grant-funded staff, or volunteers who work part-time. They spend about half their time on arrangement and description and less than 20% on Web work, on average.

**Size of Manuscript Collection**

Since institutions aren’t required to keep these statistics in a standardized unit of measurement, respondents reported the size of their collections in linear feet, cubic feet, linear meters, and items, with the majority using linear feet. The size of processed and unprocessed manuscript collections varied widely. The total size of processed collections ranges from 385 to 32,839 linear feet, with an average of 8142.78 linear feet of material. The total size of unprocessed collections ranges from 150 to 22,038 linear feet, with an average of 4499 linear feet of material.

**Levels of Description**

When asked about the level of description in their print or other traditional finding aids, most (48 or 73%) responded that they include collection-level description with other elements such as scope and contents note, biographical note, series descriptions, and folder lists. Only 5% answered that their finding aids contained less information than that. Just under a fourth of the respondents report an “other” level of description. While their explanations were quite diverse, several answered that the level of description varies from finding aid to finding aid.

All but three of 66 respondents consider a collection to be fully processed when there is a multi-level finding aid that includes folder-level description. Twenty-five of these (40%) also selected both “multi-level with series-level description” and “multi-level with collection-level description.” Nine others also selected one or the other. The respondent who answered “other” noted, “Never really fully processed. Currently considered such if multi-level description and box listed inventory.” Other comments indicated that the level of description varies from collection to collection.

More than half of the responding institutions (35 or 57%) are using database management software to keep or organize their collection-level data. The most commonly used software is Microsoft Access (15 institutions), followed by FileMaker Pro (7 institutions). Only 13 (21%) are using open-source software; two are using Archon and five the Archivists’ Toolkit. Smaller numbers of institutions use library or museum information management software. The highest percentage of respondents (62%) answered “other,” including WordPerfect, Excel, NoteTab, ExLibris Aleph, and Sirsi Dynix Workflows.

When asked if their institutions have adopted, at any level, the “More Product, Less Process” approach to arrangement and description as discussed in the 2005 *American Archivist* article by Mark Greene and Dennis Meissner, 74% reported that they had done so. The comments are very interesting, with various responses pointing out that Greene and Meissner did not invent the concept. Greene and Meissner themselves mention some of the earlier archivists and institutions who had advocated this and similar approaches. Although they were not the first to suggest that archivists can’t continue to process archival collections to a high level, their article, published in a widely-read and prominent archival journal, has resonated with the archival profession, more so than previous calls for this approach. The tone of their article, at times harsh, grabs the reader’s attention and has perhaps caused this surge in interest.

**Description and Content Standards**

Most of the respondents (43 or 70%) use the SAA standard *Describing Archives: A Content Standard (DACS)*, for describing materials. Of the eighteen (30%) who don’t, seven are Canadian and have their own standard for description. Of those who use DACS, 24 apply this standard to their legacy records.

Almost all respondents (60 or 92%) are creating MARC records for manuscript collections on some
level. A significant number (53 or 80%) are encoding finding aids in EAD, but not everyone is, yet. Those who are marking up their finding aids in EAD are doing so on a smaller scale than they create MARC records. For example, the average number of EAD files created is 530.58, versus 1560.18 MARC records.

Respondents’ comments on whether the time and effort to create EAD records equals the benefits are worth examination. While many responses were short, positive answers such as “absolutely” or “yes, definitely,” they were not all glowing recommendations. One of the negative responses from an institution that does use EAD was, “No. The payoff for the time and expense of creation is negligible.” Another explained, “Since our finding aids have been available on the Web for quite some time, first as plain HTML documents and then as EADs, I don’t think we’ve realized any particular benefit to changing the format, except perhaps that the finding aids look neater. Our researchers were finding our collections through search engines prior to the conversion.”

Others indicated they just weren’t sure yet. A few indicated that they didn’t really know if it was worth it but felt “this is a standard we want to follow.” Another even claimed, “There’s no way to easily measure the ‘benefits;’ however it would be irresponsible to not encode our finding aids.”

Among those who are using EAD, there seem to be divergent opinions about the ease of creating EAD finding aids. Several mentioned the ease with which their institution creates EAD finding aids, with one explaining, “The creation of new finding aids in EAD is no more complicated or time consuming than those created in any other format.” Others indicated that the time and effort is substantial, noting that “the special knowledge required for creating EAD finding aids and making their presence on our Web site effective has been an impediment to us backing the effort fully.” This is consistent with the findings of Elizabeth Yakel and Jihyun Kim, who listed “complexity of technology” as one reason for the lack of diffusion of EAD in the archival community. Another interesting comment from the survey discussed how one institution takes advantage of a template for creating EAD “that does not require any added effort.” But they fail “to see what, if any, benefit is derived from the EAD metadata. “

Those who do not use EAD were asked if they perceive any external or internal pressure to implement its use; the results were mixed. Even some of those who have implemented were compelled to comment. One respondent claimed that some staff members “can not see any advantage in using EAD over standard static HTML pages. We feel there are no justifications for increasing processing and description time and costs for minimal advantage… I believe that EAD is a labor-intensive throwback to library cataloging methods of the past.” In addition to the increased description time and cost—whether real or perceived—there is another negative connotation to EAD. Some respondents detect the feeling that implementing the use of EAD is embracing change for change’s sake, or, as one person put it, giving in to “a subtle ‘keeping [up] with the Joneses’ kind of internal pressure to adapt whatever is cutting edge and new.”

Web Presence

All but one of the responding libraries have at least some information about individual manuscript collections on their Web sites. The types of information differ, but most include collection title, a brief description of the collection, inclusive dates, extent, biographical or administrative history, and a unique collection identifier for some collections. Although they might include these elements, they don’t always include them for every collection; 54% report that the information varies by collection.

Manuscript Collection Information on the Web

When asked how many manuscript collections are represented on their Web sites, respondents’ answers ranged from 1 to 11,000 collections. Two institutions proudly proclaimed that all of their collections were represented online. The average number was 831.59, somewhere between the average number of EAD files and MARC records created. A common approach among the responding institutions is to add the col-
lection information to their Web site as new collections are processed or there is some level of intellectual control over the collections, though some have different plans and procedures in place for legacy collections. The level of intellectual control varied among the responses. Some only add information after a collection has been fully processed; others add basic information about a collection, regardless of level of arrangement and description and completion of finding aid. Several institutions assess the “importance” of a collection and place information about it on their Web site accordingly.

Finding Aids on the Web
The vast majority of the responding libraries (60 or 94%) have finding aids—regardless of format or presentation—on their Web sites. The number ranges from 1 to 6000 with 655 on average. The criteria for getting them there are similar to the criteria for getting any kind of information about manuscript collections on the Web. Some libraries have established policies and procedures, while others report that staff simply adds them when they have time.

Finding aids are delivered online in a variety of formats. The largest number of respondents (35 or 58%) is delivering them in HTML with EAD encoding. Others are presenting them as HTML from a word-processor document and as PDFs. Fourteen libraries have finding aids online that were created in at least two different ways.

Many of the respondents’ comments to questions throughout the survey mention the conversion of legacy finding aids. Depending on the institution, this conversion process could entail a great deal of work. Forty-eight libraries (77%) convert legacy finding aids to new styles for Web presentation. Although 71% of the respondents claim that all of their online finding aids reflect the same style, 13 of the institutions that convert legacy finding aids report their online finding aids don’t reflecting the same style. One might assume that the same institutions that convert legacy finding aids would want all of their finding aids to look the same, but this survey did not ask further questions about this.

Hosting/Harvesting Finding Aids
Forty of the responding institutions (63%) participate in EAD harvesting or consortial programs. Besides the large state and regional consortia such as Online Archive of California (OAC) and ARCHEION, a number of respondents contribute their finding aids to RLG/OCLC’s ArchiveGrid. Of those who do not participate in any such program, half are interested in doing so in the future.

Only a little more than half of the respondents (58%) replied that they have some mechanism that allows users to search across fields within the online finding aid; lack of an easy way to take advantage of the tagged elements in EAD is a common complaint. Institutions are using a variety of programs and special software to make this possible, including Orpheus, DLXS, Aleph, DB/TextWorks, Tamino, and PRIMO.

Usage Tracking
Only nine libraries track the use of in-house finding aids. Some of the tracking comes from user-registration records. One person commented, “Well, we don't track use of finding aids, we track use of collections.” It’s often easier to track use of online finding aids. Several of the 28 who do track their use mentioned a specific tool for this, such as Urchin. Others simply described the frequency with which they or a technical support person does this, and the answers describe varying levels of information they capture.

One respondent, whose institution can gather this information, explained why they don’t track it: “Since our finding aids have been harvested by any number of search engines and other projects, however, I doubt there is a realistic way to gather much useful information about how and by whom they are used.” When asked about the differences in frequency of use between online and in-house finding aids, 81% answered that online finding aids are “used” more. Since the survey did not define “use,” these numbers can only be interpreted in a general sense, as “use” might mean “accessed” and/or actually referenced or used in a reference request. Several responded that they no longer have anything other than online
finding aids, while others indicate that patrons use the online finding aids on computers in their reading rooms; some institutions provide access to both in the reading room.

**Training for Online Manuscript Activities**

Not surprisingly, most manuscripts staff are simply learning as they go. Clearly, some are learning more than others. When asked what kind of training or education opportunities staff use, 98% checked “on-the-job training.” Other popular answers included peers, professional association-sponsored workshops, library school, electronic discussion lists, conferences, and professional journals and readings.

**Organization for Online Manuscript Activities**

When asked if their library had reorganized to provide manuscripts information online, the majority (52 or 84%) said they had simply incorporated these activities into their existing areas without making any name changes to the unit, department, or library. Only two libraries (3%) actually changed the name of their department to reflect these changes. A large number (35 or 57%) answered that job descriptions were adjusted to include these activities even if the organization did not make any formal changes. Twenty-seven (44%) also indicated that their staff members receive assistance from other departments to complete these tasks.

Respondents’ comments about online manuscript activities that are distributed across the institution indicate that manuscripts staff are working with people in other units or departments, which may or may not be part of the same administrative structure, to get information onto the Web. These include departments or units whose functions include cataloging, metadata services, systems, IT, and digital services. One respondent said it this way:

Primary responsibility for the organization and description of manuscript collections resides in Special Collections & University Archives, and we routinely create and move online essentially all html finding aids. We coordinate closely with a Central Technical Services (CTS) Department when record-building intersects with the general library OPAC, and we coordinate closely with the Digital Library Services (DLS) unit when undertaking CONTENTdm initiatives. All of these activities can be properly thought of as ‘manuscript activities.’ Depending on the specific project, CTS, DLS, and/or Special Collections staff may be involved in scanning and providing metadata. It is a much more fluid world than it used to be!

**Challenges of Getting Manuscript Collection Information Online**

When asked to provide three challenges faced in getting manuscript information on the Web, most respondents gave what were expected and unsurprising answers. One person’s replies sum up these answers: “staff: not enough;” “time: not enough;” and “money: not enough.” Many of the responses were similarly succinct. Others, however, provided more specific information about just what suffers from this lack of resources. Without sufficient resources to meet modern demands, respondents find that they sometimes have to choose between processing collections, new and old, and making information available on the Web. With only one or two people available to work on these projects, both may suffer. They are forced to make a choice: process collections or make something available on the Web. Which is more important, processing collections or providing access to them? There is no clear answer to this question.

Archivists face a recurring problem in making realistic and useful decisions regarding their legacy finding aids; answers to this survey show this to be a big problem for some respondents. There are several potential challenges in dealing with these legacy finding aids. One challenge is that they do not follow the current content standards. Should they go up anyway so that at least there is information about the collections available to researchers? Not only do respondents report that their legacy finding aids don’t adhere to standards like DACS, but they also report that their legacy finding aids provide item-level information that makes it difficult to convert to modern
standards and styles. Some institutions have dealt with this problem through grant-funded initiatives.

In addition to the traditional lack of resource issues libraries have always faced (not enough time and staff to process collections), respondents also report that technological problems are challenging their efforts to get materials on the Web. Many of the responses indicate that staff lack the technological expertise to make the most of what a Web presence has to offer, and they lack the technical support from other departments within their institutions to make it happen.

Another challenge is the added reference work resulting from more access via the Web. One survey answer explained, “As we put up more finding aids, we are seeing increased use of the collections. Our electronic reference service particularly reflects this use. A challenge is to keep our reference service at a level to meet this added demand, from both in person and electronic researchers.”

**Benefits**
Certainly the most beneficial result of putting information about manuscript collections online is increased access to collections. Not only does this get the information out there, but it makes the collections easier for researchers to use. Respondents agree that another by-product of this increased access is increased reference activities. Several comments, however, indicate that patrons come to the reference room armed with more information about collections, i.e., box numbers, and therefore are easier to help.

**Conclusion**
The respondents are all managing to get at least some information about their manuscript collections onto the Web. Most of the comments indicated that they want to get more there, but are unable to do so for a variety of reasons. A select few have all their manuscript collections represented on the Web in some way, either as a MARC record, a brief blurb in HTML, or an EAD finding aid.

Almost all respondents are creating MARC records for their collections; fewer are creating EAD finding aids. These simple statistics, however, are deceptive because as one looks further one finds that libraries aren’t necessarily creating large numbers of finding aids in EAD, or any other format, or putting information about large numbers of collections on the Web. The overall numbers for creating MARC records for collections are more impressive. Why is it easier to create and use MARC records than to create and use EAD? This has been explored by others, such as Yakel and Kim, and the results of this survey appear consistent with their findings. For various reasons there is a range of attitudes and opinions relating to the ease, usefulness, and value of EAD.

The survey discovered a lot of information about who is doing the arrangement and description tasks necessary before the information is published on the Web. Librarians and archivists—those who are trained and educated professionals—squeeze these duties in between a multitude of other responsibilities. They are not the only ones in these institutions who perform arrangement and description tasks, but they do spend larger percentages of their time than anyone else in actually getting information on the Web. Although 74% of respondents report they have started to implement the “More Product, Less Process” approach to arrangement and description as suggested by Greene and Meissner, as a whole libraries are still reporting large unprocessed backlogs. Not surprisingly, 85% of the respondents report they only consider a collection fully processed when there is a finding aid with folder-level description.

The challenges in placing manuscript collection information on the Web were not surprising. The prevailing challenges relate to a lack of available resources, as well as the technological abilities in getting the information there. If libraries do not see an increase in resources, then how will they manage to get everything done? This lack of resources may not ever change. What might change, however, is the uneven level of technological ability and support across institutions that allows some to place more information on the Web than others. What does unite all of us is our belief that access to our collections is critical, and increased Web presence provides that access.
Notes

1. Ricky Erway and Jennifer Schaffner, “Shifting Gears: Gearing Up to Get into the Flow.”

   http://www.archivists.org/glossary/


5. One comment regarding this was seen in response to the entry “Notes from Spring MARAC meeting: MPLP, Friend or Foe?” in the popular blog ArchivesNext. On May 9, 2008, a commenter named Paul wrote that Greene and Meissner are “innovative thinkers that aren’t afraid to rattle cages.” Here is one example of this: “An unfortunate tendency on the part of processing archivists is to use the preparation of these text notes as an excuse to demonstrate their own knowledge (of both collection and historical context) and writing ability. Perhaps this is an attempt to demonstrate professionalism but, if so, it is a misguided one that further reduces processing productivity.” Greene and Meissner, 247.

