
REPORT ON ARCHIVES 2013 ANNUAL MEETING OF THE SOCIETY OF AMERICAN ARCHIVISTS (SAA)

**August 14-17, 2013
Jody L. DeRidder**

The Society of American Archivists (SAA) is the primary national organization for archivists, and its annual meeting is their premier gathering in the U. S. Increasingly SAA is extending its focus into the digital realm: digital curation and preservation, digital libraries, digital humanities, digital content management, and born-digital materials. Providing online access to special collections and archives, management of digital content and electronic records, analyzing/extracting and migrating incoming digital materials, and improving usability of online content: all these topics dominated the sessions. All but the plenary sessions were joint presentations, with each presenter given about 10-30 minutes each, so a tremendous amount of information was packed into a very short time. I moderated a panel, on which I presented, and also ran the annual meeting for the Metadata and Digital Objects Roundtable (for which I am now senior co-chair of the steering committee). The sessions I attended included the plenaries, Princeton's description of improvements to finding aid access, the Electronic Records and Manuscript Repositories section meetings, presentations by Donnelly Walton, Aaron Trehub and Tracey Berezansky, a description of applications of archival standards to the item level, the ArchivesSpace update, and of course the sessions in which I was involved.

TABLE OF CONTENTS

Plenaries.....	2
New Approaches and Software.....	2
It's All About the Items: Digital Objects and Aggregations in Archival Description.....	2
Disruptive Components: Reimagining Archival Access Systems.....	3
ArchivesSpace: A Next-Generation Archives Management System.....	4
Other Reports From the Field.....	4
Exploring the User Experience With Digital Primary Sources.....	4
Journeys of Reconciliation: Institutions Studying Their Relationships to Slavery.....	6
Virtual Libraries and Digital Preservation in Alabama: the Role of Archives and Special Collections.....	6
Section and Roundtable Meetings.....	7
Metadata and Digital Objects Roundtable.....	7

Electronic Records Section	8
Manuscript Repositories Section.....	8
Conclusions	8

PLENARIES

The first plenary focused on introducing award winners and fellows, and then on efforts to rebuild New Orleans. The only part that I found notable was in the remarks by David Ferriero, Archivist of the United States. He spoke of how funding for NHPRC (National Historical Publications and Records Commission) had been zeroed out, and then partially restored with commitments from other funds. He is further concerned by the possibility of a total government shutdown if a federal budget is not passed.

The second plenary focused on the need to encourage students, and find ways to develop entry-level positions that do not require experience. The job market is very poor, due to cutbacks in funding.

NEW APPROACHES AND SOFTWARE

IT'S ALL ABOUT THE ITEMS: DIGITAL OBJECTS AND AGGREGATIONS IN ARCHIVAL DESCRIPTION

This presentation, chaired by Bill Landis (Yale) focused on attempts to utilize archival standards and descriptive principles in describing digital content. Results were mixed, and raised many questions.

Kelsey Shepherd (Amherst College) used Descriptive Cataloging of Rare Materials ([MSS](#) - Manuscripts) and ([G](#) - Graphic) content standards, but then added the creator into the titles, adopted Princeton's guidelines for dates, and [DLF Aquifer's guidelines](#) for encoding it into MODS. Issues raised included differentiation between items, authorship, provenance, and defining the boundaries of an item. Results can be viewed [online](#).

Kat Timms (Libraries and Archives Canada) described born-digital content using [Rules for Archival Description](#) but also did not find that they covered all their descriptive needs.

Greg Bak (University of Manitoba) said they struggled with what is the item and what is the aggregation. He viewed all aggregations (on every level of granularity) as relationships that needed to be expressed in the description. Other relationships to be so expressed included subjects, genre, artificial collections and folksonomies (tags). Even archival provenance had different meanings depending upon the point of view: traditional, serial (multiple provenances), parallel provenance, and societal provenance. In archiving emails, looking at

the sender, recipient, and subject lines, he concluded that “the line between content and metadata is arbitrary.” In looking at the use of content in social media, he found metadata to be a foreign concept to those realms, in which the data itself is the value. He concluded that “metadata is a natural concept for librarians and a foreign concept for archivists.”

DISRUPTIVE COMPONENTS: REIMAGINING ARCHIVAL ACCESS SYSTEMS

This presentation chaired by Daniel Santamaria of Princeton, with commentator Barbara Aikens of the Smithsonian Institution Archives of American Art, focused on making finding aids more useful. The room had 266 seats, but 45 more people crowded in.

To improve awareness, links to appropriate finding aids were added to Wikipedia on pages related to the topical content or the provenance. Maureen Callahan demonstrated new functionality using the [George F. Kennan Papers](#) as an example. The browse-able hierarchy of the finding aid is on the left; one component at a time is shown in the window, containing links to components within it. At each level, the patron can post a comment (They’re using the [IntenseDebate](#) plugin for a comment system), ask a question (email), or request a box (a shopping-cart interface). Requestors must register once per year for access, and when requesting access to boxes are to specify the date they wish to come in. They cleaned up the finding aids to support keyword searching between date ranges.

In the general search across collections, facets for limiting include repository, subject, genre, language, collection, and whether the content is available online. Various sort options are supported. In the content online, the user may browse thumbnails, large images, or download PDFs.

Since people like to browse, the tabs across the top support browse by topic (taken from scope notes), names, collections, and locations. “Related Materials” is populated by using the EAC-CPF ([Encoded Archival Context: Corporate Bodies, Persons and Families Tag Library](#)) records. By adding “.xml” to a link, one can view the encoded finding aid (example: <http://findingaids.princeton.edu/collections/MC180.xml>); they also have preliminary RDF available by adding “.rdf” to the link).

To support this level of functionality, they had to clean up existing finding aids and develop policies for how new ones are created. For example, they used attributes to normalize all the unitdate values to support sorting:

```
<unitdate normal="1917-05-01">1917 May 1</unitdate>
```

and the altrender attribute on titles to strip leading articles:

```
<unittitle altrender="Walrus and the Carpenter">"The Walrus and the Carpenter"</unittitle>
```

Linked data is supported as well. The authfilenumber attributes in subject fields link the Library of Congress subject headings:

```
<subject rules="legacy" encodinganalog="650" source="lcs" authfilenumber="http://id.loc.gov/authorities/subjects/sh85104431">Political satire, American.</subject>
```

and names are linked to the Virtual International Authority File entries when possible:

```
<corpname rules="legacy" encodinganalog="110" source="lcnaf" role="col" authfilenumber="http://viaf.org/viaf/129838879">Seeley G. Mudd Manuscript Library.</corpname>
```

They are currently adding Schema.org encoding to the display to increase Google's indexing of each component, in an effort to make every component separately findable on the web.

After export from Archivists Toolkit, all finding aids are run through a series of six XSLT stylesheets and then validated against their own schema (a relaxed version for legacy finding aids). Although their wiki is not public, they will gladly share their schemas and transformation stylesheets upon request.

ARCHIVESPACE: A NEXT-GENERATION ARCHIVES MANAGEMENT SYSTEM

The ArchivesSpace presentation (with about 500 attendees) unfortunately focused on the process of development, not on the expected software functionality. Software development is by the firm Hudson Molonglo (in the UK, represented by James Bullen). The team is impressive, including Katherine Kott (manager), Mark Matienzo (Yale), Brad Westbrook (Lyrasis), Kyle Rimkus and Chris Prom (UIUC), Brian Tingle and Adrian Turner (UC), and more. The software is written using: JRuby, Sinatra, Rails, JQuery, Twitter Bootstrap, JSON, MySQL, Apache Derby, Solr/Lucene, RSpec, Selenium, SimpleCov (for testing), Git (version control) and Ant.

This software is to incorporate functionality of both [Archon](#) and [Archivists Toolkit](#) (replacing both), and will have support for display and access (including components, though digital assets can only be linked) as well as export of EADs or METS. The alpha version can currently be [viewed online](#); to test the implementation, you can [log in](#) as admin/admin [here](#). Feedback is welcome, and more information is [available](#).

Membership is based on JSTOR institution-size calculations; membership information including is [online](#); regular membership (after software is released: version 1 is due out in September 2013) for a large institution is \$5,000 per year. There are currently 56 charter members. Training workshops and video tutorials will be forthcoming.

OTHER REPORTS FROM THE FIELD

EXPLORING THE USER EXPERIENCE WITH DIGITAL PRIMARY SOURCES

This session drew about 60 more attendees than we had seats, for a total of about 250 people. In the introduction, I described the problem space: we are providing more and more primary sources online, without clearly identifying the needs of our audiences. Each of the presenters addressed a different approach to addressing this issue.

Sherri Berger (UCOP) described her findings in identifying the differing user groups for the separate interfaces of Calisphere and the Online Archive of California. Rachel Hu (California Digital Library) worked in the virtual environment to obtain feedback from users (both contextual and generative). The tools used included online survey and card sort, Google Draw (to have participants design their preferred interface) and ReadyTalk phone/web conferencing software. Her findings included the following:

- Themed collections, found via Google search, were very relevant to users' needs
- The images were the primary draw
- Object level description was important
- More description with thumbnails was desired.

The priority of information value was as follows:

- Image
- Title
- Description
- Institutional affiliation
- Copyright and use information (they wanted to download!)
- Author/photographer
- Historical background

Other things participants found important were:

- Instructions for how to cite
- High resolution images
- Better zoom capabilities
- Links to related images
- Links to related content outside the database

In the interface design feedback, it seems a pattern is emerging for a return to the "long view" in which the image is at the top, followed by the metadata, which is better for mobile devices.

She was surprised to find that none of their patrons printed content.

I reported on the findings of our qualitative study on how UA faculty researchers use online primary sources. I spoke to the preferred metadata, the dependence upon search refinement, the desperate need for organization of research content, and noted the questions we need to address:

- How do they find out about new content?
- Once they find a database, how can they tell what's in it?
- Where are the clear and comprehensive instructions for how to search in each database?
- Where are the browse interfaces they need?
- How can they obtain the content they need from the interface?

- How can they be better supported in organizing the content they collect, for research?

Donghee Sinn (Albany) talked about her citation analysis of how historians located and selected digital primary sources. Primary method of location was via web searches. Interface problems encountered included:

- deficiencies in systems' general usability
- poor quality of digital images
- insufficiency and lack of diversity
- help unavailable
- text in image not searchable
- too sophisticated (less useful for novice users)

Roger Schonfeld (Ithaca S+R) spoke of the immense impact Google Books is having on historians, and asked how archives can provide this level of discovery and access. Researchers conduct far less analysis in the archives now, utilizing online material instead. He said that the amount of information available is overwhelming to scholars, and we need to find ways to improve search and retrieval. Services are also needed to assist scholars in gaining intellectual control over citations, research notes, digital captures, and more.

JOURNEYS OF RECONCILIATION: INSTITUTIONS STUDYING THEIR RELATIONSHIPS TO SLAVERY

This presentation focused on the official approaches taken at Brown University, The College of William and Mary, and the University of Alabama, to acknowledge and address the relationships between their institutions and slavery. About 75 attendees were notably quiet and subdued during the presentation, which was recorded. The presenters were careful in their presentations to only state known facts, and not add interpretations or personal feelings to the discussion. The atmosphere was a little tense, but the presenters handled it well, especially Donnelly, who faced questions with great diplomacy. I was impressed by the depth of information she made available, and would like to suggest the development of a libguide on the subject, referencing appropriate collections and landmarks on campus.

VIRTUAL LIBRARIES AND DIGITAL PRESERVATION IN ALABAMA: THE ROLE OF ARCHIVES AND SPECIAL COLLECTIONS

Aaron Trehub (Auburn) chaired a presentation of the challenges and successes in building and maintaining a shared virtual library, AlabamaMosaic in a relatively poor state. Midge Coates (Auburn) spoke of the progression of the delivery system from CONTENTdm-based to VuFind. Tracey Berezansky (ADAH) spoke of the difficulties faced with lack of funding

and adequate personnel. Dana Chandler (Tuskegee) waxed poetic about the wonderful resources brought to light by this venture, and Susanna Leberman (Huntsville-Madison County Public Library) spoke to the ability of even small institutions to participate in LOCKSS via ADPNet. Aaron answered several questions on ADPNet funding structure. Approximately 45 people attended.

SECTION AND ROUNDTABLE MEETINGS

METADATA AND DIGITAL OBJECTS ROUNDTABLE

The session began with four short presentations:

- Heather Gilbert (College of Charleston) reported on their replacement of CONTENTdm with a Drupal/Fedora/Hydra-Blacklight system. Surprised to find that Fedora doesn't come with an ingest system, they used Rutgers's [OpenWMS](#) which was only in beta form. As difficult as it was to build the infrastructure, their biggest challenge was metadata normalization, with minimal resources. She recommended the use of [TeamBox](#) for a web-based project management solution (handles up to 5 people per project), Google Drive for sharing files, [Trello](#) for shared to-do lists, and color-coded Google spreadsheets to chart progress.
- Courtney Mumma (product manager for [Archivematica](#)) says the system accepts simple Dublin Core metadata (plus optional other metadata) with either simple or compound objects, and creates PREMIS-compliant METS records for ingestion into your choice of archiving systems.
- Jordan Patty (George Mason) talked about the different ways they've created item level records linking to resources in Archivists Toolkit so they can export EADs that link to these items.
- Jane Zhang (Catholic University of America) discussed issues in differing granularity of representation of archival materials.

After a brief business meeting, the group split into small discussion groups on the following topics:

- How do we best engage our members?
- How can MDOR best provide guidance on standards, techniques & tools, software, and best practices?
- How best should MDOR utilize different social media platforms? (Twitter, LinkedIn, listserv, website, etc.)
- Survey results show that digital content management issues impact almost every section in SAA. What is MDOR's niche? Or do we need to have one?
- How should we manage overlap with other sections and roundtables?

Results of the discussions were shared with the group.

ELECTRONIC RECORDS SECTION

The Electronic Records Section meeting started with brief updates, and then split into discussion groups on the following topics:

- Evolution of email records management
- Triage for digital preservation: decision making in light of limited resources
- How do archivists overcome technobabble and learn how to communicate with IT specialists in the library?
- Setting up that e-records workstation: let's compare and discuss e-records workflows and tools for manuscript repositories
- Based on the uptake of the "[Demystifying Born Digital Reports](#)" (OCLC), what needs doing next? What would help you get your born-digital content under control? What barriers do you face? (This was led by Ricky Erway of OCLC)
- Support for the E-records Management Awareness Project

MANUSCRIPT REPOSITORIES SECTION

The Manuscripts Repositories Section meeting focused on the "[Jump In Initiative](#)" based on the OCLC reports (referenced above), encouraging archivists to begin managing born-digital content. The panel presented on the results of their efforts thus far. [Reports](#) from participants are available on their website.

CONCLUSIONS

I suggest we adopt ArchivesSpace when it is available and has supported migration tools from Archivists Toolkit (which will no longer be supported).

We need to address how to handle digital content coming into archives and special collections.

We should investigate the feasibility of utilizing Archivematica for preparing our content for preservation, with the caveat that does not accommodate updates to metadata, and requires Dublin Core for all uploaded content.

We need to provide improvements to Acumen:

- Scope and content description on the entry page
- Add downloadable PDF content with included citations
- Browse interfaces
- Help pages
- Incorporate Schema.org encoding for both finding aids and digital items, to increase web access
- Consider incorporating linked data for subjects, names, geographical locations
- Support search refinement and faceting

Outreach librarians should provide assistance and guidance on how to organize materials selected for research.