

Electronic Records Section Meeting Notes

SAA 2014 Annual Meeting

Thursday, August 14 • 3:30pm - 5:00pm

Agenda:

[Tim Pyatt - SAA Council report and recent news](#)

[Ricky Erway - OCLC research report](#)

[Sharmila Bhatia - SAA Annual Meeting 2015 Program Committee](#)

[Erin O'Meara - iPres conference and Digi Pres Drop-In Center](#)

[Nat Wilson - Section business and accomplishments in 2014](#)

[Unconference](#)

[Breakout group topics](#)

[Breakout group notes](#)

[Do things being born digital allow us to collect items we would not have previously?](#)

[\(Dan Noonan\)](#)

[What are the archives and records management issues associated with big data?](#)

[\(Glen McAninch\)](#)

[Descriptive metadata created by the archivist: how sustainable is this method for the future of electronic records? \(Nat Wilson\)](#)

[PII identification and remediation \(John Rees\)](#)

[Policies re: E-records \(Cynthia Ghering\)](#)

[NARA's Minimum Metadata Requirements \(issues of getting metadata from record creators\) \(Meg Phillips and Seth Shaw\)](#)

[ERS blog discussion and brainstorm \(Martin Gengenbach\)](#)

Tim Pyatt - SAA Council report and recent news

- see minutes from recent Council meetings, available here: http://www2.archivists.org/groups/480/group_minutes and information from August meeting, available here: <http://www2.archivists.org/groups/saa-council/august-11-12-2014-council-meeting-agenda>, http://www2.archivists.org/sites/all/files/0814-1-I-A-Agenda_080514.pdf
- Council adopted Best Practices for Internships as a Component of Graduate Archival Education: <http://www2.archivists.org/standards/best-practices-for-internships-as-a-component-of-graduate-archival-education>
- Council discussed Standards Committee's Best Practices for Volunteers in Archives: <http://www2.archivists.org/sites/all/files/0814-1-II-B-Stds-BestPract-Volunteers.pdf>
- Council discussed Advocacy Agenda Issue Brief: Health Information Portability and Accountability Act (HIPAA): <http://www2.archivists.org/sites/all/files/0814-1-III-A-IssueBrief-HIPAA.pdf>
- Council discussed Considerations Associated with Future Joint Annual Meetings (please read and comment on these): <http://www2.archivists.org/sites/all/files/0814-1-IV-A-FutureJointMtgs.pdf>

- Council approved a Joint Working Group on Issues and Awareness with CoSA and NARA
- Council discussed Archives and Archivists Listserv -- now governed by SAA Code of Conduct -- new procedures will be evaluated as we move forward:
<http://www2.archivists.org/sites/all/files/0814-1-IV-D-A&AList.pdf>

Ricky Erway - OCLC research report

- series of reports in "Demystifying" series:
<http://oclc.org/research/publications/library/2012/2012-06r.html>
 - latest report: "Agreement Elements for Outsourcing Transfer of Born Digital Content"
- other reports, including reports on data curation and the evolving scholarly record, available here: <http://oclc.org/research/publications/reports.html>
 - data curation: report on interinstitutional collaboration around data curation; working on report called building blocks for digital curation; working on funding strategies for data curation; research on re-use and attitudes about data curation
 - work on the evolving scholarly record: what is the scholarly record now? (things that precede and follow the official publication)

Sharmila Bhatia - SAA Annual Meeting 2015 Program Committee

- info about 2015 meeting: <http://www2.archivists.org/am2015>
- next year's conference will be in a convention center
- experimentation with sessions for next year:
 - there is no theme, but if there were a theme it would be "let's try something a little different"
 - format chosen for a session proposal needs to be justified: why is it the right one for your audience?
 - all formats welcome
 - sessions will be shorter (60 minutes or possibly 75 minutes)
 - no longer using component group endorsements (this is an experiment)
 - new feature for this meeting is pop-up sessions (submit pop-up ideas in May or June, will be more informal)
 - deadline for session proposals is 10/8

Erin O'Meara - iPres conference and Digi Pres Drop-In Center

- iPres is in Melbourne from 10/6-10/10: <http://ipres2014.org/>
- Erin is on the program committee
- see schedule here:
<http://ipres2014.org/sites/default/files/upload/provisional.programme.pdf>
 - they are planning a systems showcase: have invited all of the major digital repository offerings to do a feature set demo with q&a (includes Tessella, Archivematica, The Keep, Duraspace)
 - also planned: a panel about whether we are succeeding in digital preservation efforts (hoping for lively debate)

- Digital Preservation Drop-In Center at SAA 2014
 - thank you to volunteers and component groups involved (WAR, SNAP, ERS)
 - if you have feedback or interest, get in touch with Erin

Nat Wilson - Section business and accomplishments in 2014

- thank you to departing steering committee members and webmaster
- welcome to new committee members and web master:
 - <http://www2.archivists.org/groups/electronic-records-section/ers-2014-election-information>
- discussion of accomplishments from 2014:
 - ERS blog planning
 - purpose of the blog is to share expertise and information
 - one of the unconference breakout sessions will discuss the blog
- official transfer of Chair position to Martin Gengenbach

Unconference

Breakout group topics

https://docs.google.com/spreadsheets/d/1HDdx24jeFEeQ-FIHwzOIPM0ag4YxzcRIM48te_gNPWk/edit#gid=1496345943

Name	Discussion Topic	Brief Description of Topic
Dan Noonan	Do things being born digital allow us to collect items we would not have previously?	Records that are now created digitally that we may not have collected in the past due to their sheer volume, syllabi, research data, enterprise data in data warehouses for example, we may now have the capacity to collect. Because we can should we? Do we want to? What value do they add to the historic record?
Glen McAninch	What are the archives and records management issues associated with big data?	The group discussion could cover many issues such as: storage, linked data elements; privacy; preservation and processing challenges, visualization and analytic tools; metadata challenges; scalability; and management (including record management). The types of records to discussed could include: web sites and systems; spatial and other databases; social media; and large collections of unstructured data with little associated metadata. A definition of big data may need to be formulated or reformulated. Questions the group could ask are: What resources or projects do you know of or are willing to discuss? What are the challenges you face? What would you like to work on? How do you appraisal of big data? What can we do about the challenges? What can digital preservationists learn from or teach big data creators?
Nat Wilson	Descriptive metadata created by the archivist: how sustainable is this	Should archivists of large digital collections rely more on computer and user generated metadata to facilitate access? For instance key term searching, facial recognition software, user created tags and user created transcription. What would this

	method for the future of electronic records?	mean for archivists and researchers?
John Rees	PII identification and remediation	Health sciences archivists working in the paper world have a relatively easy time of identifying/restricting PII/PHI content. As we move to born digital collecting we are especially needy for tools and techniques that will allow us easily identify/restrict/release these same electronic data.
Cynthia Ghering	Policies re: E-records	Success (and failures) in efforts to establish institutional policies addressing the creation, access, and preservation of born-digital business and historical records.
Meg Phillips	NARA's Minimum Metadata Requirements (issues of getting metadata from record creators)	
Martin Gengenbach	ERS blog discussion and brainstorm	

Breakout group notes

Do things being born digital allow us to collect items we would not have previously? (Dan Noonan)

- the need to document policies and decisions about what you're going to collect
- forever is a very long time
 - we have to collect with the notion that at some point we will have to reappraise; that there's always someone making reappraisal decisions
- email: valuable to collect b/c is documentation of conversations that we would never have collected in the past
- need to ramp up training and appraisal instruction at departmental level, so that when we do get collections pitched to us, we can be ready to pre-appraise and evaluate
- with more complex digital objects (research data, web objects, learning management systems), we will have to collect more sophisticated info in order to be able to preserve the complex objects and make them available; don't just take it b/c you *can* take it
- space, time will still be an issue
- Erik Moore's thought: in the past, we did exclude things based on format (for example, we turned down materials if a paper collection was too voluminous to handle)--so this isn't a new issue

What are the archives and records management issues associated with big data? (Glen McAninch)

- breakout group discussion participants included representatives from universities, OCLC, federal and state government, Microsoft, etc.
- what is big data? group didn't reach any single great conclusion about this

- discussed issues such as appraisal, access, privacy, designated communities
- the group would like to plan a blog post series about big data and break it down into various issues:
 - what is big data?
 - an interview/case study, in the vein of LC's The Signal blog

Descriptive metadata created by the archivist: how sustainable is this method for the future of electronic records? (Nat Wilson)

- computer-generated metadata
 - OCR, extracting info from documents for keyword searching
- topic modelling; can create subjects based on frequency of words and sub-subjects
 - need for types of programs and descriptions of tools that do this* (would make a good blog post)
- tools that generate reports about the types and quantities of files you have in a collection and ways to visualize that data, too
- FTK and forensics software
 - some of these tools are still very difficult to implement or extremely expensive
- using Tableau for visualizing electronic records
- perhaps a hybrid of computer-generated metadata and a little bit of human-generated metadata
- perhaps a single metadata record that would refer to an entire database
- free-text-searching and the consequences of relying heavily on that for search; terms have changed dramatically over time, so the search terms might not apply (for example, WWW for "web")
- natural language processing
- archivist-created metadata should be just administrative metadata?
- in one of the archives represented in the discussion, metadata isn't for the users, it's for the archivists: in their archives, the archivists are expected to do all of the research--their own descriptive metadata is imperative for them
- recommendations from the audience about tools for metadata and data mining?:
 - DiRT Directory: <http://dirtdirectory.org/>

PII identification and remediation (John Rees)

- tools discussed:
 - NARA's PII sniffer
 - FTK (but full version is proprietary)
- enormous amount of false positives from tools that look for PII and how time consuming it is to use them; but, is it time-consuming compared to going page by page, the old fashioned way?
 - risk analysis; figure out points in time when you need to look for PII (when you're putting an item online, etc.) -- whether and to what depth PII identification and remediation needs to be done
 - where in the archival process you might do it: run scan when you first take image or ingest materials, to identify where the problem areas might be; but if you're

looking for semantic identification, you might need to know the collections well to know what to search for to find PII

- is there space for donor education, to put the onus on donors to find PII
- virtual reading rooms and PII
 - library of VA and governor kaine's records
 - virtual reading room terms of access

Policies re: E-records (Cynthia Ghering)

- variety of backgrounds represented in the breakout group: museums, religious institutions, higher ed
- chicken and the egg conundrum: make the policy or the procedures/process first?
 - need to do both; tackle the low-hanging fruit, the thing with universal appeal in your institution (this is photographs for several organizations represented in the group)
- need to have an executive sponsor to help your process move through and make sure people understand your policy
- Cynthia will submit notes from the breakout group for a blog entry

NARA's Minimum Metadata Requirements (issues of getting metadata from record creators) (Meg Phillips and Seth Shaw)

- what metadata will you require from donors?
- often, metadata is so hard to extract from a system, so archivists will take whatever metadata we can get from the donor
- being archivists, we often deal with things in aggregate--ask donors to describe in aggregate
- metadata is context, in place of physical examination from diplomatics, and context is absolutely essential to the record

ERS blog discussion and brainstorm (Martin Gengenbach)

- motivations for blog:
 - idea behind the blog is the need to encourage communication among ERS membership
 - how to find a unique voice for the blog?
 - blog could maintain open lines of communication and offer ways to encourage member contributions
- how to find a unique voice/identity for ERS?
 - MDOR overlap
 - ERS has struggled with activity between the annual meetings
 - blog as a way to self-define
 - digital archives just = archives now; everyone is working with digital
 - how will current digital archivists become more specialized and take on new job titles in the coming years?
 - what can ERS members offer broader SAA?
 - service provider to component groups?

- undergird digital records/archives foundation for SAA, act as liaisons?
 - Is ERS still the group of official experts in e-records?
 - How do the e-records topics apply to humanities, science, cultural heritage?
 - need to know the audience (balance between new practitioners and advanced)
 - SNAP as a model of component group identity and collaboration--they have a common, well-defined cause: <http://snaproundtable.wordpress.com/>
 - idea of crowdsourcing the blog and ERS activities/support, or perhaps section-source as opposed to crowdsource
 - can we provide a forum for other groups who don't have a blog
- blog content ideas:
 - Mark Matienzo started a digital archives RSS at one point a few years ago; if there is still interest in RSS
 - balance aggregation and automation with new, original content
 - lessons learned at Smithsonian: pulling and pushing content and guest posts bumps up the readership
 - maybe distribute a SurveyMonkey to see what topics people want to hear about?
 - the existing digi pres q&a is helpful for the recipes that people want, the questions that people have
 - pinging, article recommendations
 - establishing types of information/categories of posts
 - report on other conferences
 - one model: dh + lib: <http://acrl.ala.org/dh/>
 - another model: RM Roundtable blog: <http://saarmrt.wordpress.com/>
 - hesitant to partner with other blogs - don't want to steal ideas, how to enhance existing content and feature existing bloggers and thinkers?:
 - invite them to be interviewed on blog or for Skype interviews, post about it, commission them to write something
 - first thursday: make post that links out to things the section will find interesting
 - monthly "Jump In" initiative update with people from different sections
 - member of the week
 - erecs pioneers
 - have a question posts
- logistical questions:
 - who will run, own, and sustain the blog?
 - announce upcoming blog topics so people can sign up?
 - have a schedule for original content?
 - use ERS listserv to announce blog posts? (maybe automated posting)
 - use social media to promote? (Facebook, Twitter, Tumblr, via Hootsuite/Buffer, etc.)
 - provide guidelines for writers