STEER UNCONFERENCE

STATE AND TERRITORIAL EXCELLENCE IN ELECTRONIC RECORDS

Hosted by the Council of State Archivists as part of the State Electronic Records Initiative

CoSA
Council of State Archivists
DOCUMENTING GOVERNMENT • PROMOTING HISTORY • SECURING RIGHTS

Sponsored by

Preservica

May 18, 2023

Conference Organizers

SERI Advocacy and Outreach Subcommittee
Cathrine Giles - Kentucky (Co-Chair)
Josh Hager - North Carolina (Co-Chair)
Kelly Horton - Arizona
Patsy Mitchell - Tennessee
Jeannine Roe - Indiana
Christian Skipper - Maryland

SERI Steering Committee Representatives
Allen Ramsey - Connecticut (Co-Chair)
Nick Connizzo - CoSA (Co-Chair)

CoSA Representatives
Joy Banks - Executive Director
Anne Ackerson
Lisa Johnston
Rebecca Julson

We wish to thank all attendees for a fantastic discussion! We look forward to further collaborations among the state and territorial archives concerning electronic records management.

Records Management Track Notes
- Discussion 1 (12:50 - 2:00 pm) - General Summary
  - Dashboards
  - Compliance Tracking
  - Retention
- Discussion 2 (2:15 - 3:10 pm) - General Summary
  - Storage
  - Transfers
  - Scan & Destroy

Arrangement and Description Track Notes
- Discussion 1 (12:50 - 2:00 pm) - General Summary
  - Workflows
    - Ingest
    - What happens after ingest:
    - Staffing, training, and collaboration:
- Discussion 2 (2:15 - 3:10 pm) - General Summary
  - Confidentiality

Digitization and Access Track Notes
- Discussion 1 (12:50 - 2:00 pm) - General Summary
    - Workflows
    - Metadata
    - Indexing
    - Other topics
- Discussion 2 (2:15 - 3:10 pm) - General Summary
  - AI & Metadata
  - Getting Started & Staffing Issues
  - Access Systems

Email Management Notes
- Discussion (3:15 - 4:10 pm) - General Summary

Final Thoughts and Wrap-Up Discussion

Records Management Track Notes

Discussion 1 (12:50 - 2:00 pm) - General Summary

Participants: 48

Dashboards

- 1 county records person in attendance to learn best practices
- Illinois has not had to address these as records. Their State Records Act defines what is considered records, governmental entities have admin rules and codes which work in conjunction with retention schedules. No one has come out and asked to deal with dashboards
    - IL's perspective would be they could be considered "reference" unless they are creating new data.
    - Is it just the DATA that is a record?
- Indiana is considering how (or if) the way of presenting that data is what should be preserved. Can example data (at least) be preserved to show look/feel demonstrative of its function when it was alive.
- Anything important enough for dashboard-type retention would be a longer time frame in Illinois (more than 2 yrs).
- Indiana considers that if a state agency creates a dashboard, it's a record (not reference material). It may not be archival or permanent, however.
- For Covid data, at one point Maryland had been taking daily screenshots; they also tried crawling unsuccessfully. Realized all data is on the MD Data Portal. Have not moved on pulling the data down from the portal.

- Indiana is looking at access as a main feature. Once the dashboards are no longer available online through the agency, how can patrons utilize the data?
    - MD - is there cultural value?
    - IN agrees, yes there is cultural value. Why were things displayed or created in a certain way or not?
- MD County Commissions are made available through a portal that has a dashboard look and feel but is a website: https://go.boarddocs.com/mabe/carps/Board.nsf/Public. Boarddocs can't be crawled with Archive-it. Agencies post here and though they also are supposed to submit those records to the archives directly, but often don't. Because they can't be pulled down by the archives, they are inaccessible. Possibly being created ONLY is this html webpage format.
    - Emily shared https://washingtoncounty.civicweb.net/Portal/
    - IN did county survey about how they are storing and serving out certain kinds of records and many of them are using a certain portal.
- MD GovDelivery - unable to save records out of GovDelivery. IN does create an archive of our newsletters on the GovDelivery site. If we did try to collect, we'd either collect the bulletins emailed out or take screenshots. No way of downloading from GovDelivery and retain the formatting. MD links PDF versions because their newsletters have images.
    - IN suggested PrintFriendly as a possible solution: https://www.printfriendly.com/

Compliance Tracking

- Wanted to hear what other agencies are doing
- MI and CA do not have a lot of enforcement implemented
- "Don't want to be the records police"
- MI has a content mgmt system, CA does not and instead has a more decentralized network of agencies that do their own things
- Mechanisms include: surveys
- Side discussion on the potential impacts of AI for cleanup and classification and identification, new research uses. What could be the overall impact for the profession?
- Content management systems and retention schedules, including different approaches and considerations.

Retention

- Retention Schedules
    - Maine has general and agency schedules
        - Law rarely points to retention of specific records

        - Just revised general schedules in 2022 to capture common records
    - Warren County in Ohio
        - Sunshine law state
        - State law mandated counties and townships to create their own schedule about a decade ago
        - An organization that works with county archives produces mock schedules to assist other agencies
            - CARMA
            - Interagency group that collaborate with the State Archivist.
            - Free meetings for interested attendees
            - Listserv
    - Standardization
        - In North Carolina, there are standard format language at the State level that local governments can pull from so they match that guidance
            - This has worked well for administrative record types that are pretty standard, less so for program-specific or statute-driven record series
        - Arizona sets general retention schedules for every level of government
            - These are very numerous
            - Supplemented with custom retention schedules designed to capture records specific to certain agencies
    - Challenges with electronic records at the county/local level
        - Arizona has had a hard time providing guidance to local offices with limited resources
        - CARMA in Ohio has been providing general guidance to local agencies on data retention, but resources and IT support are still a challenge
            - Ohio has a data board, so ARM professionals are working with that group to begin allocating resources to data retention programs
        - In North Carolina, university system wants archives to provide clearer guidance that electronic records have to adhere to the same laws as paper records, not just have blanket disposition or indefinite retention driven by IT
        - NAGARA has basic server clean-up webinars on their website (free for members, circa $40 for non-members)
        - Maryland has encouraged the development of schedules and timely disposition of electronic records by reminding agencies of the legal and financial burdens of maintaining those files indefinitely
            - FOIA laws and equivalents are costly if unnecessary records are retained.
    - Process of moving from custom to general schedules
        - In Arizona, statute enables a central agency to set records policy throughout the state

                - Records Officers are required by law but limited consequences for breaking this requirement, so the state has focused on training and building relationships with records officers
                - Moving away from custom schedules has taken years, but largely came down to the state producing general schedules based on studying commonalities across custom schedules and agencies largely being happy to comply - not many holdouts so far.
            - Most agencies want help and guidance.
- Maintaining institutional knowledge
    - In Indiana and Maryland, agencies are required to have a records coordinators nominated by the agency and approved by the State government
        - This doesn't necessarily resolve the institutional knowledge issue, but does encourage consistency by having a dedicated point of contact
    - In Ohio, sending an annual email out to department heads, directors, and elected officials asking for a records management update helped in this effort as it ensured updated contact information in the face of turnover

Discussion 2 (2:15 - 3:10 pm) - General Summary

Storage

- Digital storage transfer
    - North Carolina is shifting from on-prem to cloud storage
        - Not a lot of experience with this at the moment
        - Storage is handled through State IT department and IT demanded shift to cloud
        - Still in the process of determining functional requirements
            - Planning to have it look like the current system, but there may be necessary changes (especially with security)
            - One unresolved issue is out-of-state storage due to cloud server locations
    - Preservica
        - Handles online digital preservation
        - It's important to address security permissions at the application level, but ensuring that cloud solution meets IT requirements is important
- Cloud Storage Disposition
    - Preservica manages this through an audit trail that shows, via metadata and reporting, when a record was uploaded, converted between file types, viewed, and/or disposed of.
        - The record itself is destroyed, just disposition is tracked.
- Long-Term Digital Storage

    - In Nevada, transitioning to building a digital infrastructure right now
        - Specifically exploring on-prem solution and how agencies might transfer records
- Large File Formats
    - Preservica has handled a lot of audio/visual material, and they encourage not only large storage space but also heavy computing power due to access requirements and search capabilities when it comes to bodycam footage
        - Similar issue with large engineering design formats, like geospatial data layers
    - In North Carolina, geospatial data is transferred to the Archives
        - This is large-scale, and is consequently transferred on a rolling basis

Transfers

- Automating transfer workflows
    - MI thinking about this, looking for other examples of others who have done the same. O365 state with centralized IT. Have lifecycle management, but not linked to retention schedules.
- Preserve365 from Preservica
    - Mike Quinn shared that they have two reports from their EAG (early access group) looking at best practices on transfers in O365. The two reports focus on two groups including end users/record managers and archivists. It identifies their needs and challenges.
        - Special Report: Records Governance & Preservation in Microsoft 365: https://drive.google.com/file/d/1076qm9q_vfnKNxtJH8IEslwasQSFgNl8/view?usp=sharing
        - Special Report: Archival Transfer for M365 Content: https://drive.google.com/file/d/1X9VAli8UFI7htHyVN_S24EwMNxWwOVk-/view?usp=sharing
- Preserve365 built on O365 records management capabilities.
    - Shared drive issues
    - Education component in management of shared drive contents by end users
    - Discussion of using cost/risk to show IT the importance of shared drive records management.
    - Implementation of O365 records management and tools
    - MI hasn't fully implemented the capabilities. They have 600 schedules with 900 records series.
    - They are working on lifecycle management with timeframes applied for data stored in various O365 buckets (Teams chats, etc). They want records custodians to move important records before the deadlines for automatic storage kick in. Started this year, big educational curve, helps encourage agencies to not use these bucks as dumping ground for records.
    - Mike Q observed that not many users implementing records management labels in O365 as expected. Some of it is a licensing level issue.

    - Discussion of how life cycle management can be a good way to get people thinking about electronic records management, but still no easy way to transfer the records once identified/tagged.
- Challenge of different platforms
    - Some states like AZ working with many different platforms (ex: schools using Google, some agencies using O365 and some local office strictly paper). Wide range of needs, formats, budgets. No automated transfer yet.
- Regulatory changes
    - Discussion of positive and negative impacts of regulatory changes
- Possible tools to automate the transfer process, workflows.
    - APPX has an agency portal for request of transfer, but for now it is linked to paper records transfers.
    - Discussion of status of Hawaii's plans to build an automated transfer workflow in-house
    - Discussion of Oregon's electronic records center
    - MI used Microfocus Content Manager (used to be HP Trim). Two data sets for paper and erecs. They charge agencies for the services. Works pretty well.
    - Preservica hopes to build on research around the challenges to erecs transfers (automated) from agency to Archives)

Scan & Destroy

- MD mentioned they were on a fact finding mission.
- IN shared current policies:
  https://www.in.gov/iara/files/policy-20-01-erecords-retentionanddisposition.pdf
  https://www.in.gov/iara/files/policy-20-02-erecords-technicalstandards.pdf
- IN also shared a recent bill:
  https://iga.in.gov/legislative/2023/bills/house/1591#digest-heading
- Maine has no scanning policies but they do have guidance documents; also have a requirement for archival documents to be retained in paper or microform; scanning is only for ease of access. Looking to NARA guidance.
    - AZ does the same. Non-permanent can be scanned to standards and tossed. Being pushed to move toward scanning of permanent records. Looking to create standards for this. Would be more comfortable if they could confirm that their electronic system was capable of storing those records appropriately.
    - A participant mentioned this ties into storage. They have $ to acquire a digital preservation system, but it's difficult to keep track of what every agency is using for storage or to ensure the records will be available in the future.
        - IN shared ERMS guidance for agencies:
          https://www.in.gov/iara/files/2022_Electronic_Recordkeeping_System_Checklist.pdf

        - CA Trustworthy System Regulations: https://www.sos.ca.gov/administration/regulations/current-regulations/technology/trustworthy-electronic-document-or-record-preservation
- OH code does allow archives to have electronic copy as original or official record aside from certain record types. Not much guidance has been put out. Digitization and born-digital records grew enormously during pandemic. “Just because we can digitize, doesn’t mean we should digitize everything.” More and more repositories are going toward digital retention. Need to make sure they are using preservation quality systems and that are secure.
    - Ohio shared ERC guidance created over last 10 years that touches on a variety of topics like preservation software, email preservation, retention, and other records management, located at https://ohioerc.org/. Ohio also has a YouTube series
    - WA also has robust resources:
    - https://www2.sos.wa.gov/archives/recordsmanagement/how-to-scan-image-records-and-go-paperless.aspx
- MD asked, “For those that have policies or offer guidance, did handwritten vs typed documents make a difference or were a concern in your decisions of what is allowed to be S&D’ed?”
    - OH having discussions regarding hybrid paper/electronic records and challenges with their management, because they can change or be added to and are therefore permanent but “living” documents.
    - CA’s emphasis has been on the output and storage handling of the scans and electronic records, more so than the source format.
- CA asked about destruction documentation.
    - AZ has certificates of destruction and require agencies to file a minimum of one per year. Planning to develop user training so docs get filled out more consistently. Will not accept destruction forms until a Records Officer is assigned to the agency.
- CA asked about how certificates are being managed.
    - AZ prints out the paper and manages physically (print and store). Retention has dropped from permanent to 50 years. Would like to set up an online submission portal in the future.
    - MD disposal tracking/archive: https://msa.maryland.gov/msa/intromsa/html/record_mgmt1/records-disposal.html
    - And guidance and forms https://msa.maryland.gov/msa/intromsa/html/record_mgmt1/records-disposal.html
    - OH maintains official copies of the form for auditing or records request purposes.

Arrangement and Description Track Notes

Discussion 1 (12:50 - 2:00 pm) - General Summary

Workflows

Ingest

- Connecticut received federal funding for two projects
    - 1) Direct agency transfers to digital repository (BITS)
    - 2) Looking for suggestions to creating a secure repository
- Other states report varying procedures for digital ingest
- NY reports that SharePoint will alter metadata for Office formats upon ingest
- MO has a robust ingest tool through InfoLinks to upload records, assign them to record groups, and accession them into custody. It's used right now for routine state publications. It has size limitations.
- Library of Congress is no longer supporting Bagger, so new solutions are needed for ensuring file integrity in those states who use it.
    - New Brunswick reports using DROID - National Archives in UK - free and open-sourced, one version has Java baked into it
    - DROID is built into Preservica
        - https://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/

- New Brunswick is trying to emulate Library and Archives Canada's approach to pre-ingest workflow: Preserving the bits : Library and Archives Canada's Pre-Ingest workflow - Digital Preservation Coalition (dpconline.org) where digital pres unit takes initial pass to flag issues then assigns to archivist for processing

What happens after ingest:

-AZ reports that researchers use the Records Management blog to find new items, but are using the 856 field to link out in their catalog
-TN tends to process born-digital records when they're in hybrid collections, where they list items or folders in finding aids
-NC talked about the varying levels of processing - every transfer gets some processing through Bit Curator, but some collections get more detailed processing based on appraisal or the state of the collection to make it accessible
-NJ asked if folks had processing manuals - most institutions do not have one for born-digital records
-New Brunswick shared this resource for what finding aids might look like for born-digital collections: Examples of Born Digital Description in Finding Aids – This is a community-driven initiative to document current descriptive practices around born digital archival materials in 2021-2022. This is a snapshot of practices and is not intended to change or update over time. (wordpress.com)

Action Item: Create a high-level best practices guide for digital description that's broken down into phases or certain areas of the process. Go from the higher-level to create format-specific guidance (i.e. Preservica, AXAEM, etc.).

Staffing, training, and collaboration:

-Most institutions have only one electronic records specialist or no single specialist and everyone helps in part-time ways
-Finding solutions for staffing are difficult
-NJ talked about introducing staff in different departments assisting in born-digital projects
-NC asked about how to find guidance since it's so diffuse. CA reported taking two courses through Digital Curation and Arrangement and Description of Digital Records through SAA: https://mysaa.archivists.org/nc__event?id=a0l5a00000F0udJAAR
-Resources are either too technical or not technical enough. **Potential idea for Tools and Resources to evaluate tools based on the level of technical expertise needed to use tools.**
-How do you learn digital skills once you're hired? A lot of reading of notes. NJ reports working with CoSA's Nick Conizzo to help with workflows as well as creating surveys. You can contact Nick at nconnizzo@statearchivists.org.
-DigiPres implementation guidelines: https://www.dpconline.org/digipres/implement-digipres

-Nick (CoSA) talked about BACKER and getting involved with it: https://www.statearchivists.org/electronic-records/state-electronic-records-initiative/backer
-Best Practices Exchange: https://bpexchange.wordpress.com/
-AL shared a case study on their experience starting electronic records program: https://files.archivists.org/pubs/GovtRecordsCaseStudies/CASE-3-CollaborativeProactiveApproach.pdf
-Preservica Community hub: https://community.preservica.com/

Discussion 2 (2:15 - 3:10 pm) - General Summary

Confidentiality

- TN asked if we ask agencies to identify series or records that may have confidentiality concerns. Some states do, but most agree that it's not particularly effective yet.
- TN and NC discussed data dumps and how "non-confidential" series can contain irrelevant files that do post confidentiality concerns
- Question posed: Should all digital records be restricted until reviewed item by item?
- MT asked if we should ask patrons to sign letters agreeing not to share PII if encountered
- NC asked if it's even possible to catch all PII and confidential materials before providing access
- "The problem you will encounter with any "permission-based" restrictions is that web crawlers certainly won't respect you and if you expose your data to the web you should expect that it will be copied and reproduced elsewhere."
- VA reports asking patrons to sign a policy on using materials with confidential info: https://www.lva.virginia.gov/about/policies/Archives_and_Map_Research_Rooms.pdf
- A participant mentioned tagging documents with metadata to denote exemptions for access using built-in functionality in SharePoint, but he reported difficulties in getting agency buy-in
- Some institutions ask all staff to mark emails as public/internal/restricted/confidential, but the archives doesn't necessarily trust the distinctions, as some people overclassify their records
- Is due diligence enough for ensuring confidentiality? It's an open question.
- Michigan's statewide IT requires everyone to flag emails as public/internal/restricted/confidential, but people overuse the confidential tag.
- AZ discussed agency usage of "attorney-client privilege" to prevent transfers and access to records.
- AL discussed a memorandum of understanding with agencies.
- NC and VA discussed building relationships with agencies over time, both to encourage regular transfers and to assure them that we can maintain confidentiality where needed.
- **Potential for communities of governors and legislative archivists to collaborate**
- Restrictions in large collections

- NY reported challenges in applying software to find restrictions due to difficulties in finding all restricted information as well as hurdles in formats like PST calendar files
- A participant mentioned that there are tools for restrictions, but they aren't yet robust enough. He reiterated asking agencies to identify confidentiality in transfers.
- RATOM tool for evaluating emails: https://ratom.web.unc.edu/about/
- It's an incredibly difficult question of legalities and ethics when addressing whether born-digital records should receive an automatic restriction upon ingest

Digitization and Access Track Notes

Discussion 1 (12:50 - 2:00 pm) - General Summary

Workflows

- Share our workflows
- Sharing the tools that we're using
    - Bitcurator or other tools
    - Preservica
- Is this a 2-track topic: basics of just knowing what we have and then more extensive set of steps
- On demand scans? Yes or no?

Metadata

- Standards (or lack thereof) - what are the local workarounds that institutions are using?
- Be consistent in any misuse of fields for future migration + mapping
- Metadata language, specifically harmful language - difficulty of identifying underrepresented communities
- Developing boilerplate language to address the sensitive/harmful language warning while allowing the materials to remain as originals

- Explaining any access limitations
- Who is creating the metadata? - most seem to do everything
- Metadata standards. CA uses Preservica. Hangup: feels like every type of material is going to need a different set of metadata. NY has just started using Preservica
    - Maine: Different legislative sessions their own agent/creator/author in archives space. Some data doesn't fit into metadata schema. Have to create or knowingly misuse fields
    - New York: also archives space. Focused on selecting different types of labels. What is the EAD tag appropriate to the data? Consistent misuse was key to successful migration. Try to make metadata work with field that fits best.
    - No set standards
    - A big challenge is how you're structuring metadata within a system. Bigger learning curve in how you create and structure metadata terms.
- Metadata language: how to reflect changes in harmful language
    - Maine: boilerplate statement but still maintain links. Open Context Initiative working with indigenous tribes to add tags to material that identifies sensitive topics. Access not affected for the most part. Only if from a donor or not strictly a state government.
    - New York: addressing issues with description itself. Education used compilation of images
- If you are creating metadata, who is doing that? A separate team?
    - CA: a team but they are on other teams
    - One participant: one digital archivist and two contract staff. Share duties with cataloging and metadata entry
    - New York: staff dedicated

Indexing

- Missouri + Utah discussed challenges
- Systems
    - CONTENTdm
    - Microfocus
    - AXAEM
- Support for described features
- Managing expectations from users given technology limitations
- Missouri State Archives
    - uses FromThePage
    - traditional indexing with staff and volunteers (FamilySearch indexing)
    - currently in transition from an old records management and archival system
    - programs used: Tower Trim and Micro Focus
    - looking for something that will do the records management and the digital preservation
    - looked at AXAEM but Micro Focus is more robust (US Navy and Oregon State Archives use Micro Focus)

    - goal - to get automatic indexing with born-digital records
    - System they are moving from has a lot of issues related to migration
    - moving from a global license to a per-seat license (150 concurrent users licensed now)
    - customer base - 3,000
    - ContentDM - they are better than they used to be - using it successfully for photo collection
    - started self-hosted and moved to vendor-hosted for the same reasons
    - ContentDM not helpful for publications - should be able to search but it cannot
        - publications like session laws and revised statutes
        - search for a word and this word appears in these session laws
        - does not highlight the session laws
- Utah State Archives
    - involved with indexing for a long time - started with traditional methods and currently have database online
    - Main platform - AXAEM
    - managing expectations (i.e. different than FamilySearch) because not everything is searchable by name
        - internal and external users assume they can search by many types of facets
    - thinking of using a newer feature of FromThePage to add additional metadata
    - used ContentDM - tried to use their PDF features for indexing on digital accession logs but wanted to get them into the system searchable (was not working)
    - ability to split them into pages that you load
    - thought it would do the searchable text (not yet)
    - started self-hosted and got 3 versions behind because their IT did not want to support it so they are not vendor-hosted
    - IIIF support - has been helpful
    - connected to their name index database - can have multiple names per record (feature not available in ContentDM)
    - in the midst of a year-long transition and update the programming everywhere
        - could only link to a JPEG image on their local server
        - reprogram it to use IIIF
    - exclusively uses compound objects - everything optimized for that (manuscripts - works well)
    - run OCR if it is type written, otherwise, no
    - court index books online (since 2015)
        - Ability to look through the books online - saves people a step

Other topics

- Conscious/Reparative Description - NC is in final stages of developing staff guidelines
    - Is there a perfect answer? No

    - Be sure to define what you're doing:
        - Additional description
        - Corrective description
        - Who will be involved and how will they be compensated?
    - This can be an overwhelming process. Once you see it, it's everywhere. Be prepared for burnout + internalizing the emotional labor of this work.
    - Set the standard for the state
    - Show the priority - get it on your work plan
        - Start with a survey, case study - establish the baseline
        - Then work to push out to other areas
    - Work up to workshops, trainings, resource lists - take the time to educate everyone who will be a part of the work. Be aware of where you'll need to educate outside the core group that's working on
    - Connecticut State Archives: our first steps as part of a grant project which we developed a Terminology and Usage Disclaimer and Policy, https://libguides.ctstatelibrary.org/archives/uncoveringnewhaven
    - Thinking through state agency names that had problematic titles
- Migration - NC is shifting from CONTENTdm to Quartex over the next 5 months with a target of end of May 2023. Official launch end of June 2023
    - Collaborating with the state library
    - Used a set of standards for the testing, but reality of what's promised and what's delivered is always an issue
    - Wished for more prep time for metadata and asset cleanup
    - Moved from homegrown to ArchivesSpace - forced move due to infrastructure. Sometimes the forced deadline is a good thing so they don't get drawn out over an indefinite period of time. It will never be perfect.
    - Coming to terms with the fact: this will never be done. Just make sure you're doing the best and getting the access component
    - CT is doing a "slow crawl" from Cdm to digital repository Islandora
- Metadata - quilting saying applies "Finished is better than perfect"

Discussion 2 (2:15 - 3:10 pm) - General Summary

AI & Metadata

- Use of Whisper AI for transcription
- Automated Handwriting transcription
    - Census - Ancestry/LDS + software
    - Older v newer spellings creating challenges
- MIT transcription software development for Ancestry
- Using AI to find + redact PII
- Adobe
- YouTube for transcription - not so great

- Accents + accuracy
- ABBY - a tangent for another time
- From The Page - outsourcing transcription
- AI internal + external advocacy + conversations

Getting Started & Staffing Issues

- For getting started:
    - Some archives are really just getting started with eRecs + preservation programs
    - NJ is sending a survey to state agencies to develop scope for structuring + building the program - always a good idea to do a temp check with other agencies (still within scope, broaden/narro)
    - Relationship building with other agencies is really important - expand beyond the paper record network. Reach out to the agencies before you want the records (end of term, change of party, change of administration)
    - Having a thoughtful metadata plan will help in the development of your electronic records planning
- For staffing
    - You'll need to involve other people/individuals - "you are not an island"
- Nevada: just getting started. Concerned about governor's records coming to state archives (social media, emails). Going with archives space to capture social media.
- New Jersey: just starting. Preservica purchased before electronic archivists started.
    - How to get IT to work with you?
    - Danielle.marchetti@sos.nj.gov | Danielle from New Jersey State Archives | To ask about survey
- Create relationships with agencies. Reach out to agencies before a problem.
- Don't be okay with letting agencies just keep their records.
- Make it easy and enticing.
- What search terms are non-archivists using and is the metadata capturing those terms? Have a plan for metadata.
    - "Building the plane when you're trying to fly the plane"
- Survey agencies to get scope of work
- NJ: Two people in digitization is not enough. Realizing how much other departments (collection management needs to be involved). How much are other departments going to be involved? Justify hiring more people.
- NV: Three people and a half time position.
    - Agencies lost interest in electronic records initiative

Access Systems

- ID - public facing access systems for indexing + digital collections
- Online indexes powered by AXAEM/Solr: https://archives.utah.gov/research/indexes/.
- There seems to be a strong emphasis on public access, making it as easy as possible for the user

- NY looking for a systems for what will work best
- Publishing electronic records created by public agencies
- AXAEM - what existing tool could we use now? Will it work?
- Multiple systems for different needs - is there one system for all needs? Creating access barriers for the users
- Public presentation of digital content
    - ArchivesSpace
- Indexes + tools for publishing - AXAEM came up again
- Email collections - challenges on the way + how to PREPARE
- **Indiana Archives and Records Administration**
    - public facing - collection access - based on indexing or digitized collections
        - great resource but now looking at implementing something that is better suited for born-digital collections
        - levels of access to support that different collection types, formats, larger volume of data
    - starting at square one - looking for way to provide access for digitized collections
    - interest - focused on public access - go onto website and interact with this as a public citizen
    - asking themselves what existing tool could they use now
    - can request different features in their system
    - combing through other states' systems and websites, surveying what other places are doing - testing search functions
- **Email**
    - for collecting agencies, focusing on elected or appointed officials' inboxes
    - maps to retention schedule and any other electronic records
    - easier in conversations to discuss email versus other types of electronic records
    - determining where do they start - policy and decision making records, format, supplying records, research use
    - they have the infrastructure to transfer ERs
    - born-digital - format - dependent on the situation
    - working with Indiana Department of Technology (their IT) - there have been challenges
    - everything M365
    - created a lot of documentation to provide to the in-agency coordinator
    - ***if someone sees a really good tool - contact ER program at Indiana***
- **New York State Archives**
    - no system in place at New York State Archives - they have a digital collection of their holdings
    - thought they would use that to publish electronic records transferred by state agencies
    - objects that are emphasized that look good for navigating but there is a disconnect in public presentation with description and access tools
    - ArchivesSpace - describe and manage holdings
        - they don't publish finding aids through it

        - doesn't seem to work well with series and how researchers are using it
    - used CollectiveAccess to publish the digital collections and Preservica for managing ERs
    - no neat way to integrate public interface with ideal finding aid interface
    - challenge - presenting a unified experience for our online users
    - big increase in states agencies to transfer born-digital collection-rethink separation of systems
    - indexes-looking for a tool that would help create and publish a large amount of indexing data (i.e. 1 very large series consists of more than 500,000 entries)
    - collect and publish through their Drupal website -product to specifically index archival records
- **Montana Historical Society**
    - hunting for a state program where they can upload digitized and born-digital materials
    - ArchivesSpace for small collection, government record series and manuscript collections in conjunction OrbisCascade (ArchivesWest)
    - Digital Collections - accessible online Montana History Portal (ContentDM) (with State Library)
    - OrbisCascade - good for finding aids of large collections
        - hoping to use ArchivesSpace as front end for that instead - in discussion for a long time
    - ArchivesSpace will eventually be front-end user access page
    - Looking at ArchiveMatica (in addition to Preservica) to be front end for oral histories, born-digital, and digitized collections
- **Utah State Archives**
    - uses AXAEM for their name indexes (special instance of AXAEM)
    - large volume of names in their index - birth and death certificates (structured database)
    - use Solr to provide a search interface
    - AXAEM started in Utah before it was opened up to other users (Index is a module in AXAEM)
    - Online indexes powered by AXAEM/Solr: https://archives.utah.gov/research/indexes/.
    - new version of system - allows staff to import a spreadsheet
        - customize fields and columns for each index
    - use FromThePage to use each index to connect to records online
    - FamilySearch shared one of their indexes in raw form-use Delimit application to work with indexes in that format

Email Management Notes

Discussion (3:15 - 4:10 pm) - General Summary

Nick Connizzo (nconnizzo@statearchivists.org)

- project manager for COSA - BACKER and PREPARE
- formerly project manager on PREPARE grant
    - https://www.statearchivists.org/electronic-records/state-electronic-records-initiative/cosa-prepare
- We don't control what the records producers do"
- Not everyone uses email as you expect

Discussion questions:

What are some challenges that you face regarding email?

- not everyone understand email retention at all or has thought about it
- no one helped them set it up to build in retention to their processes
- accumulation for years - thousands of emails - many of them could have been deleted according to retention
- if not explained, people aren't thinking about it
- redaction kept coming up in his last session
- even with ePADD - not immediately obvious how to use it or the delivery module
- how do we provide access in a way that makes sense?
- leverage deep learning modules /computer assisted techniques or algorithms to help us process
- lexicons are lacking
- "At present, we don't have e-mail. We are preparing for the possibility that we may get the current governor's e-mails as part of the transfer of other records. We have no idea what e-mail platform is being used. How to prepare?"
- Nick Connizzo recommend looking at the MoVE-IT project CoSA did which specifically addresses transferring electronic records: https://www.statearchivists.org/electronic-records/state-electronic-records-initiative/move-it
- The Library of Virginia used machine learning to process part of the email of Governor Tim Kaine's administration. They gave a presentation about the experience in 2019. https://www.youtube.com/watch?v=piB8ERowIts
    - need a tool like this to go through email
    - if he did it manually it would have taken 10 years to do
    - spot check - develop ways of due diligence - too much to do manually
    - teaching the machine and working together with it
        - especially with restricted material
        - learns quickly - saves a lot of time
    - very expensive (what they use)

MDworking in a MS Outlook framework
-adopted National Archives capstone approach to email
-didn't dispose of any non-permanent email - up through 2021

As an RM for an agency, what would you want to know? To facilitate for the first time.
-want to have the beginning and end clearly delineated
-know what specific users are involved
-include sent and received messages
-managing the scope ahead of the transfer

-a tendency for these transfers to be data dumps
    -things that aren't public records or have a shorter retention

Policies in place to weed, put it with the proper series?
-encourage archivists and RM to explaining to and educating professional that email was not a mechanism of attachment records retention
    -attachment was meant to be downloaded by the recipient or saved/organized by the sender
    -encourage not to just store attachments in email

South Carolina: One thing that I'm currently doing is going through the records of previous governor's series in our holdings to get a better idea of what previous archivists viewed as official records--it's not always 1:1, but it's helping guide not only selection, but also arrangement (pre-email governors' series, that is) - +1 from CT

Connecticut:
-experience with governor's records
-tried to match up his records with who they worked for in the governor's office
-organized by user and the governor's name (?)
-trying to understand what analog archivists are doing with the paper records

NC
-goal - researchers can search for information (paper or digital) and then how they access will be different but the searching would be through the same place or a description gap

SC
-looking at researchers' files from the 90s
-see how previous archivists handled it

Many governors' offices are using Intranet Quorum (IQ) to manage constituent correspondence as well -- something that needs to be accounted for since it's very different from O365 or G-Suite https://intranetquorum.com/

Interest in a webinar talking about SC + CT approach to email archiving

Nick Connizzo

- very few are willing to go through and categorize things
- when you are linking records in your email and sending them to people
    - not always functional later - strips the context
    - same with emailing attachments with comments
- Figure out your end points - what are you trying to do?
- How do you share, contextualize a collection of 1 million emails

NC

- I was at an IT conference yesterday and there was a presentation on the rights management challenges those direct links via email to OneDrive and SharePoint documents create for IT site managers. So that feature is creating challenges for a lot of folks
- asked to find if a certain railroad car (used to transport material) to and from a nuclear plant was in the email
- discovery - keyword searches
- looked up alternate spellings online - zero results for that search
- flip side - agencies - get requests - I want every email about [insert topic]

MD

- lots of broad requests - they are legally obligated to reply to them
- required to respond within a certain period of time
- managing expectations for searches that will take a long time

MD Record Request Process:

- paper trail created (informally) of searching for the information and who was searching
- letting requester know there will be costs associated (2 hours of free labor)
    - can request an extension from the requester
- close communication with legal counsel about what can be provided and what cannot, referring to specific citations

Nick Connizzo

- did natural language processing with PREPARE
- options - tech is immature for the archives community
- good appetite for model that accounted for peculiarities of state government

Questions to Consider:

- What is the current state of your institution's email management (program)?
- What solutions have you found successful?
- What other kinds of resources can CoSA/SERI provide?

Chat question: Does anyone have experience pulling emails out of gmail/google and losing some searchability? Maryland Gubernatorial emails being pulled out of Gmail and losing searchability

- Mbox exports - PREPARE did investigate, as part of file format conversion. Mbox converter software often strips headers
- Utah, CoSA use Gmail
- Emails from the 1990's saved as html documents - is there a way to convert back to recapture the email header information? No success

Well, and are folks placing "processed" email collections into their digital repositories? As individual emails? As PSTs? As MBOX? Are folks creating finding aids for individual accounts? Etc.

- PDF (CT)
- SC: if it came from IQ, they are mostly PDFs. I'm also keeping the PSTs but converting them to mbox as a prStatewide IT is currently retaining ours (I misspoke about them being in our digital repository), so this will be a future challenge for us as a preservation format. there are also folders with a lot of .emls, htmls and .txts from previous ingestions
- NC: Statewide IT is currently retaining ours (I misspoke about them being in our digital repository), so this will be a future challenge for us.
- Generally, some states have accessioned a number of "legacy" PSTs that they're holding onto. many, many states are just leaving things in their O365 environment
- M365 Guidance (NC): https://archives.ncdcr.gov/government/digital-records/digital-records-policies-and-guidelines/microsoft-365-best-practices-and-usage

The conversation about FOIA requests is important as some states/territories have not ventured into "collecting" email yet due to the inability to provide similar levels of access for FOIA requests that agencies can provide (e.g. through eDiscovery)

Storage:
PDF (CT)

Emails part of FOIA requests, provided via PDF
(separate PDF version of the email collected and provided)

Preservica - refining text search across email (something for the Preservica Community Hub)
Preservica has an email group set up too

Nick Connizzo
-MailBag - SUNY
-EA-PDF - E-PDF - migrating email /mailbox into a PDF
-presenting on it in the future at the Email Symposium

EABCC Email Symposium:
https://emailarchivesgrant.library.illinois.edu/email-archiving-symposium/
APPX Shoptalk: A practical approach to processing millions of emails:
https://www.statearchivists.org/events/event-description?CalendarEventKey=077a9955-5ecd-4cde-a312-0187fbcc89a5&Home=%2fprograms-education%2fcosa-webinars
Email preservation & access group (may require sign in)https://community.preservica.com/groups/email-preservation-and-access-110
SERI: https://www.statearchivists.org/electronic-records/state-electronic-records-initiative

Final Thoughts and Wrap-Up Discussion

Records Management

- dashboards
    - define how they might be used
    - is the dashboard itself a record
    - what format allows it to retain its functionality when it is a living document
    - do you collect the data?
    - keep the state data - views that were used?
    - how does the vendor play into things?
- working in a collaborative way with agencies

RM - Storage

- shifting from on premises to cloud storage
- challenges of disposition in cloud storage
    - challenges of large file format - storage and access
- complications with transfers from Office 365
- encouraging lifecycle management as a concept to prevent challenges later

Arrangement and Description

- no breakout rooms - one big group - 30 people
- ingest - how we're getting records from agencies
    - transfers, file integrity
- what happens once you have the records in your agency
- what is a processed collection?
- are they discoverable to people? Catalog records, finding aids, etc?
- need to establish a working group to define best practices for A and D for born-digital material.
    - maybe need to take on as a group
- training - where do you go to get the skills to do this work?
- confidentiality - legal issues, ethical concerns
    - can you even make things available?
    - do you have to look at every item to make that determination?

Digitization and Access - focused on workflows, AI, migration, access systems

- no metadata standards - what to use
- discussed challenges with indexing
- guidelines around reparative description
    - never a perfect answer - ongoing work
- "things you wish you knew pre-migration"
    - metadata and asset cleanup
    - finished is better than perfect

AI

- don't use YouTube for transcription

- leverage AI for use in programs
- Get agencies involved and email programs started
- best access systems for your agencies

High level guidelines for electronic records-

- criteria for level of expertise you'll need to use a resource in the Resource center
- Filtering so that you can access based on your level of understanding
- States + the tools they are using
- Grant is interested in helping with what we're working on

Grant Robertson in South Carolina created a spreadsheet of collection management systems and their features.
https://docs.google.com/spreadsheets/d/1cXOug3qM0pNNeD_wssiVEv9c0W1Y5I1VDTnSPTk7fb4/edit#gid=0

Future Collaborations

Inspired by the Unconference to collaborate? Please fill out this spreadsheet to connect with individuals who are seeking collaborations on presentations for CoSA. While not an official submission of a presentation, the SERI Education & Programming Subcommittee or the SERI Advocacy & Outreach Subcommittee may reach out to discuss details with you! Alternatively, you can reach out to the SERI Education & Programming Subcommittee when you're ready to present! You can find them at: **https://www.statearchivists.org/electronic-records/state-electronic-records-initiative/seri-subcommittees**