

STEER 2025 Session Notes



STEER UNCONFERENCE

STATE AND TERRITORIAL EXCELLENCE IN ELECTRONIC RECORDS

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



Sponsored by



May 14, 2025

Conference Organizers

STEER Working Group

Tara Maharjan - New Jersey (Co-Chair)

Josh Hager - North Carolina (Co-Chair)

Grant Robertson - South Carolina

Christian Skipper - Maryland

CoSA Representatives

Joy Banks - Executive Director

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.

Table of Contents

Opening Session: Continuing SERI's Story.....	4
Evolution of SERI.....	4
SERI's New Strategic Plan.....	4
CoSA Transitions.....	5
Track 1: Policy.....	6
Session 1: Advocacy for Electronic Records Programs.....	6
Translating your message for stakeholders:.....	6
Electronic records management policy.....	6
Connecticut's drafting of Electronic Records Policy:.....	6
Digital preservation tools.....	7
Internal Advocacy (Administration and Management).....	7
External Advocacy.....	8
Advocating for Digital Transfers.....	8
Session 2: Starting an Electronic Records Program.....	9
How do you build a program from scratch?.....	9
Training Stakeholders.....	9
BACKER.....	9
Overcoming financial barriers.....	10
Overcoming staffing barriers.....	10
General Thoughts.....	10
Lessons from Nick.....	11
What might be a good project to start with?.....	12
Complexity of records.....	12
Legal Arguments.....	12
Track 2: Conversations on Captures.....	12
Session 1: Social Media, Websites, and Data.....	13
Understanding Where State Archives are with collecting social media, websites, and data as public records.....	13
ArchiveIT and Archives Social.....	14
Appraisal of these items.....	14
Concerns.....	15
Action Item.....	15
Session 2: Spotlight on Cloud Repositories.....	16
Oregon Overview.....	16
Overview in other states.....	17
Track 3: New Frontiers.....	19
Session 1: The Next Phase of our Story: Machine Learning and AI in Electronic Records...	19
Different types of AI functionality.....	19

Detecting restricted information in records.....	19
Concerns with AI tools.....	20
Agentic AI.....	20
Resources.....	20
Training / Education.....	21
What do people want from SERI on AI?.....	21
Session 2: The Promises and Perils of Privacy Protection in Electronic Records.....	22
Privacy officer overview in Maryland.....	22
General Privacy Questions.....	22
Privacy Policy toolkit available?.....	23
CJISSECPOL.....	23
Databases as Records.....	24
Closing Session.....	26

Opening Session: Continuing SERI's Story

Subject Matter Experts:

- Joy Banks, CoSA Executive Director
- Allen Ramsey, Connecticut State Archives and CoSA Vice President
- Cathrine Giles, Kentucky Department for Libraries and Archives

Evolution of SERI

SERI was founded in 2011, but states have been looking at electronic records since the 1970s. It's become an essential part of CoSA's work over the past decade and a half. It sponsors Electronic Records Day in the fall and STEER each spring.

SERI rebranded from "State Electronic Records Institute" to "State Electronic Records Institute" in 2024 to acknowledge that SERI is a core function.

SERI's New Strategic Plan

SERI's Strategic Plan, 2024-2026:

https://growthzonecmsprodeastus.azureedge.net/sites/2163/2025/01/CoSA-SERI-Strategic-Plan-2024-2026_FINAL.pdf

Goal 1: Facilitate care and custodianship of public records

- Standards-based guidance
- Training
- Surveying and evaluation
- Direct assistance
- Communities of practice

Goal 2: Confirm the authority of archives and records management

- Advocacy and awareness
- Build the [CoSA Resource Center](#), which relies on submissions from members
 - The Resource Center is under construction after our recent website migration and motivated by our new SERI Strategic Plan

Goal 3: Universal access to public records

- Accessibility, especially for underserved populations
- Cultivating a culture of service

Goal 4: Sustainability of SERI

- Partnerships and funding
- Governance and leadership
- Staffing support

CoSA Transitions

- CoSA has transitioned from committees to working groups to eschew long-term commitments and allow State Archives staff to help as much as they can.
- Need to re-evaluate the DPCMM in our future - want to keep on top of making sure we are understanding State Archives' need.
- Recent BACKER publications: <https://statearchivists.org/resource-center/#resources-3>
- CoSA and SERI are more integrated in a more holistic approach. You don't need to silo impact into electronic and non-electronic.
- We've started some conversation with this BACKER publication:
Best Practices for Electronic Records Custody and Control:
<https://councilofstatearchivists.box.com/s/ngrc1s0lbgekdytkmthvrf7klk869079>

Track 1: Policy

Session 1: Advocacy for Electronic Records Programs

Subject Matter Expert: Allen Ramsey, Connecticut State Archives

Translating your message for stakeholders:

1. Know your audience, especially in the IT realm
 - a. Some light prep work is useful, but open communication is absolutely critical
 - i. Example: Discussing records retention schedules with IT as a method of ensuring document security and integrity
2. Establish common definitions and language at the beginning especially between agency personnel
 - a. Example: What is an “archived” website?
 - b. At higher level, you don’t get into definitions as much

Electronic records management policy

- Internal work first, then working with external stakeholders on what was missing
- Connecticut:
https://portal.ct.gov/csl/departments/public-records/state-program/policies-forms?language=en_US
 - Connecticut’s policy built off of digitization policy. Agencies are now allowed to digitize and destroy records based off the current standards.
- North Carolina: <https://archives.ncdcr.gov/government/digital-records>

Connecticut’s drafting of Electronic Records Policy:

1. Staff drafted policy over two years.
2. Feedback sought from two divisions, archives and records management
3. Once both units agreed on policy, the head of the agency examined the policy and addressed her concerns. She needed to have the knowledge to speak to commissioners about the plan.
4. They then worked with records management liaisons (employees in other agencies that have other job titles). They are clerks, commissioners, etc. that also have a liaison hat. These agencies follow guidance even if there’s no legal authority from the Archives.
 - a. The courts were insightful. They were already doing similar work and their workflows matched up to the suggested policy.
 - b. Town clerks and other municipal staff also provided feedback on the policy.
 - c. The model was emulated from the 2014 process where an in-person meeting occurred to gather feedback, but was conducted virtually.

5. After feedback, modifications were made before final publication.

Digital preservation tools

- Connecticut Digital Archive has the largest instance of Islandora, which has obvious pros for preservation and access but also some drawbacks like bot attacks and difficult large-scale transitions.

Internal Advocacy (Administration and Management)

When working with higher-level staff, especially elected officials, make presentations more concise as you'll only have a few minutes of their time. Have an elevator pitch ready. It's an interview question for staff - how do you convince a town clerk not to discard materials? There are strategies to this kind of work and it's key to know your audience.

Challenge of internal advocacy at management levels higher than the state archivist. How do you very succinctly convey the significance of our work? Parent organization is more focused on primary education and visibility of the museum. How can you build something to get people interested in our work?

- Visuals are helpful. A local police department made a visual where someone took a screenshot of all the numbers from scans from a retiree's desk. How do you find files in that mess? The new employee had to open every file and rename it to something that made sense.
- Show the immensity of the work. Demonstrate the size of records, like an email account with 10,000 messages, and ask how to comply with law and best practices.
- Bring out old technology like floppy disks and ask how to protect those records.
- Use presentations to the public to showcase work.

Use disaster planning and compliance as the "stick" to balance the "carrot" of doing the work for the right reasons to preserve history. Train other agencies and higher officials on the needs for compliance and what could go wrong if our work isn't supported.

If you can't advocate directly, how do you make inroads with those people that can?

- Manage up and try to foster those connections long-term
- Create one-page summaries of essential statistics on electronic records programs
- Integrate regular reporting and make sure those decision makers have key information

External Advocacy

Outreach programs to get interest in archives programs can lead to good publicity and help with advocacy, but they can also draw unwelcome attention when folks misconstrue what we say or take offense to benign comments.

You might get a chance to meet with legislative staff members rather than elected officials. Take advantage of those opportunities to get that person more knowledgeable about your work. Have folks over for tours or other events.

When bills are under consideration, work with the legislature about how you can offer help to make something more feasible and narrow down the focus of projects and budgets.

Work with external organizations that are fans of the organization to advocate on our behalf with the legislature or other policymakers. Consider “artsdays” to partner with non-profit organizations. Connecticut built a mobile outreach van that mirrors their main facility. History Day is another kind of event that can get folks interested.

Encourage folks to conduct research in born-digital collections. Incorporate them into resource lists and into the catalog. Create a one-stop shop for where to find your digital resources.

Advocating for Digital Transfers

How do you convince agencies to transfer electronic records when space isn't a major concern?

- Connecticut is trying to put infrastructure in place to make the transfer process more straightforward. There are some steps for automated transfers, but that has its own challenges. Some agencies want to make records available on a portal.
- Pennsylvania is creating a digital records center and leveraging compliance as a reason to transfer records. They are working with a vendor (LibNova) to develop agency instances to allow for transfers directly into the digital archives system. The system incorporates the records management schedule and has disposition codes applies to digital containers.
- Cost WILL become a problem. Do you want to pay the ever-growing cost for storage and maintenance? Do you want to respond to public records requests?

Advocating for proper transfer: Sometimes agencies push back on why they can't just email files.

- In Kentucky, it is that simple (for some). We have a dedicated email set up for folks to send in electronic records. The people who usually transfer this way are the folks who are in charge of meeting minutes, audits, etc. I inherited the email, which has been going for about 20 years. The problem I've been finding is that we're struggling with buy-in for new records officers. We have a solid group of people who regularly send in meeting minutes though! It was originally a dedicated email for state pubs from what I've been told and then transitioned into more general electronic records.

- It was originally a dedicated email for state pubs from what I've been told and then transitioned into more general electronic records
- Have a dedicated policy in place and provide that to agencies to explain the process.
- Use the tools from vendors to streamline the process, if possible.

Session 2: Starting an Electronic Records Program

Subject Matter Expert: Nick Conizzo, former BACKER administrator

How do you build a program from scratch?

- Connecticut: Had a program in 2010 based off a grant with some other states. Archivists had to build their knowledge of the topics. Supervisors were willing to get training for staff to do something, which was better than nothing. Was also critical to build a policy framework and make the policy iterative—it's never perfect first. Needed action plans for processing, accessioning, etc. You need to have that plan in place before you take anything into your custody.

Training Stakeholders

What are some of the most effective ways of training stakeholders on e-records management and compliance that folks have used (eg, videos, "office hours", manuals, one-pagers, in-person training?)

- Connecticut built a virtual program based off of CoSA's webinar series. They have webinars on distinct topics, including electronic records management and the role of the archives. Most webinars have over 100 state employees. They're also recorded for registrants. It's been effective to reach people who they wouldn't have found if they also had in-person training. The older training sessions were five hours long. They want to do a survey on what could be next steps.
- North Carolina: Offers a variety of programs and focuses on virtual workshops

BACKER

BACKER information: <https://statearchivists.org/resource-center/#resources-3>

- Grant in 2021 from IMLS to go through July 2025 (ended early)
- Investigate the next step, based off Access and Prepare
- What happens when you have electronic records in your collections?
- BACKER worked with six institutions (states and territories). In some cases, it involved weekly meetings and other intensive involvement. In other places, it was just a policy

overview. The reports are newly shared and include model workflows, best practice documents, and are currently writing recommendations for what comes next.

- CoSA ran the Digital Preservation Capability and Maturity Model surveys (DPCMM) and future planning for State Electronic Records program (SERP) to self-assess capacity for electronic records and figure out where you want to go. It also considers NDSA's levels of digital preservation to make these factors in our government ecosystem.
- You don't need to wait to start. It's not about achieving perfection (there's no such thing). It's getting to where your institution needs to go.
- New Jersey: You don't have to be scared about workflows. Every situation will bring up new concerns and challenges. They are living documents. It's important just to start.

Overcoming financial barriers

- New Jersey uses BitCurator for a lot of their work. They only pay for Preservica and an FTP tool approved by the state.
- Nevada's model workflow:
<https://councilofstatearchivists.app.box.com/file/1831568174156?s=6tt4ojr1zu9feqlx5utk5e0va9tpy44u>
- You can find open-source tools for many of the necessary components to digital preservation work, although the options are limited.
- Other tools to consider: DART (<https://aptrust.org/documentation-page/dart-digital-archivists-resource-tool/>) and DROID (<https://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/>)

Overcoming staffing barriers

- Build in knowledge of electronic records for all staff
- Try to get a dedicated position for electronic records if you can
- Consider applying for a grant-funded position to get the ball off the ground
- Archives can't do this work alone - the best operations are enterprise-wise.
- IT has a larger budget. Try to work with their staff if possible and get them engaged in these larger conversations.
- Craft partnerships with external partners to streamline the work. For example, organizations can assist with workflow decisions.
- You can't get something if you don't ask - so try asking!

General Thoughts

- Change is the only constant - be prepared to assess and refine your work. This is never going to be completely solved
 - There will always be new formats, new tools,

- If the record is defined but the volume increases prolifically, what adjustments should be made (e.g., surveillance footage - there is so much more of it now, so how can we mitigate the preservation?)
- You will never get to the same level of stability with electronic records that you had with your analog records
- The program you start now will look different than the one that started in 1995 - that's ok, and exciting. Do the best you can with what you have.
- Start the planning now - take what you must, but really have the planning done as much as possible
- Think about how you integrate your electronic records into your existing collections for all the same reasons you would do that for analog - chain of command, legal, transfer protocol, etc. Make the user experience as similar for access as what already exists. Make sure that the physical and digital systems for transfer align
 - You have the goals + standards - don't reinvent the wheel
- Electronic records and their ephemeral nature may necessitate changes in retention and disposition expectations. Transfer shouldn't necessarily wait until the end of a term, but might need to transfer every few months. The records are also constantly in flux. Determining how to capture information and when is a critical component of an e-recs program.
- Make the time for your own professional development - try to engage the administration so they have at least a basic understanding of the work to help make sure that you have the resources and support that you need

[State Archiving in the Digital Era: A Playbook for Preservation of Electronic Records \(2018\)](#)

CoSA is going to start a Tools & Tech series to go over the ins and outs of various programs across all facets of electronic records.

Lessons from Nick

- Start with a small project and consider the scale. The number of electronic records and systems are growing.
- We're a long way away from PDFs on a server. We're discussing databases with many different structures. We need to consider those structures to determine what we need to preserve.
- You can only go as far as the governance that you have. If things stop, there's a breakdown in coming up with classifications or figuring out what to do with those classifications. There needs to be a person executing strategies to make it work.
- Consider a holistic perspective but use one successful project to springboard your work. You can also demonstrate your work to IT, agency partners, etc.

What might be a good project to start with?

- A discreet constitutional office with a term, staff, and digital component. Those also often have constitutional mandates directing their acquisition and preservation
- Multimedia collections - many agencies want to offload those holdings
- Discreet program or event (like an anniversary, disaster, emergency response) that had a length of time, beginning and end
- Government is small and word will travel fast when you help solve someone's records problem

Complexity of records

Records are so complex right now. There are still "archival formats" for electronic records. Should we be trying to get the agencies/records creators to create their records in our "archival formats" or work with what they have?

- As archivists, we're trying to preserve the original in its context as best as we possibly can.
- We need to preserve the context, relationships with data (e.g, health services documents)
- Prioritize the preservation of the original as best you can - however it is. Document the systems and their structures as best as possible
- Context is critical, metadata is so much more - metadata is the system and critical for understanding the file
- Encourage the records makers to touch the records as little as possible

Legal Arguments

Any advice for talking with lawyers that aren't willing to allow the transfer of original files? our lawyers have staff transfer EVERYTHING to pdf and wipe all data

- Many states in public records laws say "records include XYZ" which often include the metadata and relational links - if you strip a record of that, then you are technically in violation
- If this is happening, be sure that you are documenting the action of those who are violating or going against best practices. This will also be extremely valuable to your successors in the archives
- As archivists, it's not really our job to tell them *how* to create their records
- The capture and collection of provenance is proper and necessary
- Get the technical documentation whenever you can - make anything around the creation of a database as historical so that you have those explanations for the future

If you are starting a program, and you need support, we are a community of practice and support - you are not alone, let us help you on the journey

Track 2: Conversations on Captures

Session 1: Social Media, Websites, and Data

Subject Matter Expert: Krista Sorenson, Minnesota State Archives

Description: Discuss the ongoing challenges associated with capturing social media, websites, and data as public records. This conversation will specifically address how the changes in the social media ecosystem have affected workflows and strategies for ensuring no archival public records are lost in the confusion.

Understanding Where State Archives are with collecting social media, websites, and data as public records

- Getting agencies to understand that we do not need every version of a website. Is something else considered the public record, such as a PDF version that can be the official record over a social media post? What do we do about comments on social media?
- Content over format. Some states are operating that if the information was something that was collected in paper form, then it should also be collected in electronic form even if it is shared on social media.
 - Connecticut State Library, but from what I understand, their guidance is to focus on the *content* of a record, not the *format*. That means if something is considered a permanent record on paper, then the same type of content shared through social media, email, or any other format would also be considered permanent and would need to be retained accordingly. We also make it the responsibility of the agencies to pull this data.
- Agencies in Texas are scheduling records retention schedules for social media. The archive is starting a pilot project to take social media
 - Texas did presentations at BPE (2022 and 2024) about a social media project they have been working on to process social media and then render the information
 - <https://bpexchange.wordpress.com/wp-content/uploads/2024/07/2024-bpe-thomas-slides.pptx>
 - <https://bpexchange.wordpress.com/wp-content/uploads/2024/01/2022-bpe-thomas-slides.pdf>
 - Youtube, Instagram, Facebook and Twitter - Creates a WARC file to encapsulate the data for rendering in hopes that vendors will want to help with rendering
 - <https://github.com/brianmatthewthomas/socialMediaprocessors>
 - bthomas@tsl.texas.gov if you want to ask more questions

- Kentucky is revising and updating its policy - is social media a public record because we do not own Facebook or other platforms. Updating their policy from 2015. New version of their policy is coming out soon. Big tip they give is to also have it written as a record copy elsewhere if it is important. Keep it in Word, social media should be a duplicate. Agencies need to download the social media themselves.
- In Alabama, the State and Local Government Records Commissions have determined that agency websites & official social media pages are permanent records. The Alabama Department of Archives & History captures "snapshots" of state agency websites & social media using Archive-It, which we consider to meet the obligations set by the Records Commission. We do not receive many local records, as these are judged to be best preserved in the community of origin, but we similarly encourage local govts to preserve snapshots of their website using the Wayback Machine. Also, I will add that we do not consider documents uploaded to websites to be an acceptable substitute for transmitting those records to the Archives.

ArchiveIT and Archives Social

- Using ArchiveIT in Wisconsin to create a one time snapshot of a website. Not relying on it to be the complete record, but to have some record
- People spoke about some of the challenges ArchiveIT has had in the last year with the attacks on the Internet Archive
- Archive-It has been working well to capture official websites, but it does not do well with social media sites
- NC is using ArchiveSocial for social media, and have just recently stopped using ArchiveIT for social media
- Websites changing URLs - needing to go back and change all the URLs in ArchiveIT. It takes a lot of time to make sure these URLs are up-to-date.
- Archives of Michigan uses Archive-It for state and local government websites and ArchiveSocial for select government social media accounts. (Only select accounts because of the expense to use for ALL of state government's social media.)
- Cynthia Ghering (gheringc1@michigan.gov) and Erin Gallagher (erin.gallagher@dncr.nc.gov) both offered to speak to people about their experiences with ArchiveSocial

Appraisal of these items

- Others are wondering if social media can be appraised away because the information is shared in a different way. A lot of agencies post the same exact information to each platform. Is there value in collecting the same info but on different platforms? Obviously, if collecting comments as well, then that would make it more unique. But would collecting from one platform be sufficient?
 - Texas is appraising the accounts as a whole, deduping is done in the appraisal and a note will be added.

Concerns

- Are we capturing too much? Are cost to maintain these materials over time a factor?
The maintenance of an additional external service was another factor in consideration.
One state noted that they are known to have issues with legislature changing its mind on what is allowed
- There are concerns about ownership of platforms both for social media but also ArchiveIT
- Social Media privacy
 - People not realizing that their comments are now public records
 - Are people thinking about take down policies? Do states have a social media policy so they can make it clear that these are public records? Pin it to the top so people are aware.
 - <https://www.warrencountyohio.gov/RecordsCenter/doc/GenInfo/Social.pdf>
 - We tell people to put this in their social media bio or pinned comment to warn the public that their comments/interactions can be records
"Communication through agency-related social media is considered a public record under KRS 171.410(1) and will be managed under the terms of KRS 171.410 through 171.740 and 725 KAR Chapter 1"
 - Wisconsin Take Down:
<https://wisconsinhistory.org/records/Article/CS17222#TakedownPolicy>

Action Item

- CoSA should consider a whitepaper or best practices document on social media policies (getting people to understand that their comments may be public records) and also take down notices.

Session 2: Spotlight on Cloud Repositories

Subject Matter Expert: Kris Stenson, Oregon State Archives

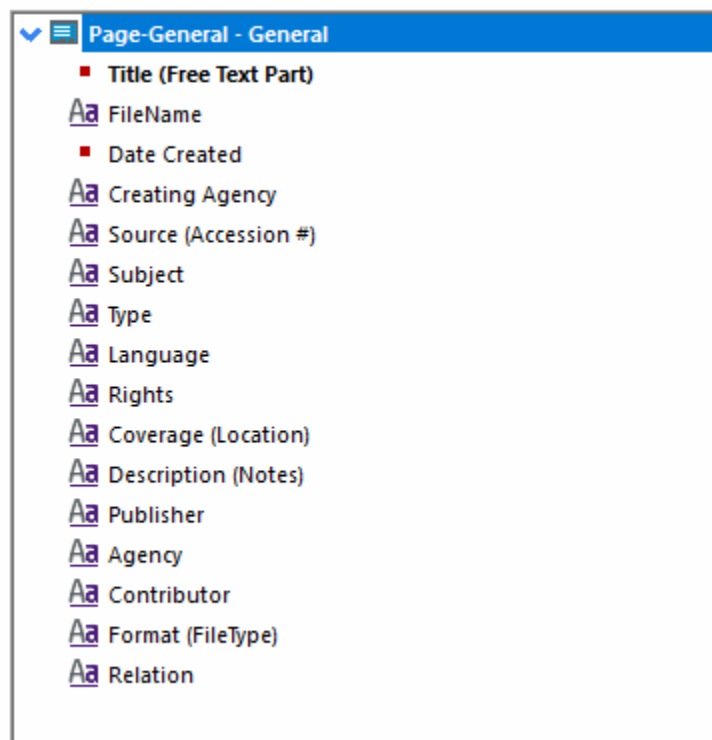
Description: In Part II of Capture Conversations, the discussion turns to cloud repositories. We'll discuss SharePoint, Google Drive, Dropbox, and more general issues of capturing records from cloud-based storage systems.

Oregon Overview

- Oregon: Has a confederated enterprise. Central IT covers a lot of agencies and dictates softwares, but it's not universal. M365 has been the dominant platform. Statewide leadership has been trying to limit anything not M365.
 - Sharepoint/OneDrive/Teams: utilize sharefile host (cloudbased) for file transfer.
 - Other platforms (GovQA)
 - Working on a large statewide retention project. Agency by agency. Intent is to manage records not to be exported. Utilize retention policies in suite instead.
 - Everything has default litigation hold – users can delete things on their user end but it doesn't affect server copy.
 - Implement manual checks on networks.
 - Authority and enforcement are very different: if an agency is using a nonconforming repository, hard to enforce.
 - Some agencies are independent to a point who are permitted to use different repositories
 - Boards and commissions: in the past issues state accounts to use repositories but they weren't used. Now working on setting up google suite to have better control of what they access
 - Offboarding workflow steps: check off point where the manager needs to identify active records.
 - Automatic retention only applies to active accounts. Deactivated accounts must be manually purged.
 - Generally speaking, don't want to transfer records out. (Exceptions: constitutional offices)
 - Microsoft doesn't export conveniently. It exports to several nested folders.
 - Utilizing data analytics tool to help process collections (not AI)
 - Get it out as it exists. Then use this to identify content and likely duplicates, pre-flag what might contain sensitive information. (helpful to reference archivist)
 - Still questions about long term access
 - Only post-pandemic they have been able to partner more with IT for non-archives applications and exert influence in policy-making
 - Changing approach (to IT) to collaborative partner who wanted to help made a huge difference.
 - They work primarily with IT branch that works directly with software

- Sharepoint data: Not sure about what metadata to keep. Won't be able to mandate all fields though. DublinCore schema
- Thinking about how to approach formats
- metadata elements that we're capturing on electronic docs:

Displayed fields



https://sos.oregon.gov/archives/Pages/state_admin_schedules.aspx

067 Oregon Blue Book Video Recordings, 2023 - [ongoing] 2GB

- (a) Retain video recordings used in the Digital Blue Book permanently, transfer to State Archives
- (b) Retain all other video files until no longer needed, destroy

Overview in other states

- West Virginia: uses Google. Long-term digital preservation files on externals.
 - State IT has control of accounts and what happens to them after separation
 - State IT has control of all email and personal drives
- Texas universities: uses Google
- Illinois: LIBNOVA. Hot storage is limited, cold storage has a larger limit.
- Maine: LIBNOVA.

- Alabama: uses AWS for cold storage of its digitized collections via a state contract. Paper and electronic records are treated the same way for purposes of retention because of format neutrality. Verifying and auditing electronic records destruction is the challenge.
- Texas: uses Preservica with govCloud and govCloud glacier. Using Box for quasi-ftp transfer but not true storage. Anything publicly accessible has to be in govCloud.
- Kentucky: has started outreach to remind state agencies that we're here and want their stuff. Working on guidance document for file formats when talking about long term or permanent records.
 - Offboarding:
 - <https://kdla.ky.gov/records/recmgmtguidance/Documents/Managing%20Electronic%20Records%20Offboarding%20Guidelines.pdf>
- Has anyone used StateRAMP?
 - Texas TexRAMP
- Ohio: If it could be considered a record (regardless of format) and you could have disposed of it but didn't, you are liable. State does provide guidance about this. Cloud cleanouts: advice/guide/train agency staff on how to do this.

Track 3: New Frontiers

Session 1: The Next Phase of our Story: Machine Learning and AI in Electronic Records

Subject Matter Expert: Grant Robertson, South Carolina Department of Archives and History

Different expressions of AI functionality

- Transcription (i.e. OCR)
 - Handwriting
 - Audio
- Sorting tasks
 - Pre-processing of accessioned materials
- Anomaly identification
 - PII
 - Hidden relationships
- Chatbots
 - Virtual reference archivists
 - Potential pitfalls

Detecting restricted information in records

- Depends on which tool you're using
 - Keep in mind that the tool is likely processing information remotely - sends your query to a server that processes the data elsewhere, likely by the vendor you're contracting with
 - Users should discuss the legal ramifications with their legal counsel - some states or organizations prohibit the use of their data for AI learning models, or being sold to third parties regardless of depersonalization
 - Some tools are being baked into operating systems or standard products, so we can't entirely disassociate ourselves from their use; it would be better to work with the vendors who provide our IT systems to ensure guardrails are in place
 - Google Gemini
 - Microsoft Copilot

Concerns with AI tools

- Sales to third parties
- “Hallucinations”
- Harvesting public data for use in private large language models
- Labor issues – expectations around increases of “productivity” / “efficiency”
- Concerns suggest that one route forward is the development of internal LLMs
- There are opportunities for greater collaborations with State IT divisions, but also the potential for obstacles in the case of tough working relationships

Agentic AI

- Connecticut asked if anyone is using this. "Agentic AI is focused on decisions as opposed to creating the actual new content, and doesn't solely rely on human prompts nor require human oversight."
 - Nobody is that we've heard
 - Attendee notes one example: Use of machine-learning in appraisal
 - Deciding which records should be deleted based on established rules
 - *Not* necessarily suggesting appraisal decisions, e.g. topics/categories/series - that would be GenAI
- Library of Virginia *may* be trying to train an LLM in the processing of Governor Kaine's email, comparing the results with human processing
 - This may just be Named Entity Recognition (NER) or Natural Language Processing (NLP)
 - The difference between this and GenAI is that the former is based on preexisting rules, while the latter is more adaptive and can produce its own new categories as it goes
 - A 2018 review of this project:
<http://www.lva.virginia.gov/news/broadside/2018-Fall.pdf>
- AI Use in General
 - SC has been experimenting with LLM transcription tools, including Adobe Premiere's newest features
 - None of it is “great”--but it is a starting point, and provides better access than having to listen to hours of audio

Resources

- <https://github.com/carascap/>
- <https://github.com/libratom/libratom>

- <https://github.com/BitCurator>
- [IMLS Project on Domestic Traffic Ads](#)
- [Projects by Professor Lise Jaillant](#)
- General terms to guide deeper investigation:
 - Digital Humanities
 - GraphRAG
 - Computer Vision

Training / Education

- Limited resources at the moment
- Maryland guidance on AI
 - <https://doit.maryland.gov/policies/Documents/Policies/Interim-GenAI-Guidance.pdf>

What do people want from SERI on AI?

- How to incorporate AI tools safely into privacy policies
- Jargon/language on AI technology
- TN: “what AI is and what it is not, types of AI and examples applicable to archives and state records”
 - NC agrees, sees a valuable distinction between records in archives custody vs records in agency custody; use vs. training
- KY: “descriptions/explanations of real use-cases in states”
- [not noted in session]: when NOT to use AI

Session 2: The Promises and Perils of Privacy Protection in Electronic Records

Subject Matter Expert: Christian Skipper, Maryland Department of General Services

- Christian Intro - Director of MD Records Mgmt Division
 - RM and Archives are separate in MD, but we work closely together

Privacy officer overview in Maryland

- MD requires a Privacy Officer at every state agency
- Was appointed as Privacy Officer, as I was the records manager already
 - The main direction is to protect the privacy of Maryland citizens
 - Work closely with the State's Privacy Officer to do that
 - Named as Privacy Officer in early 2022
 - Had to create a data inventory as one of my first tasks
- The program for data inventory has been largely successful and created out of whole cloth..
 - Each agency was mandated to create a complete data inventory along with risks assigned with each
 - The way that we got that win was breaking out all of the structural subdivisions of the agency into a document
 - Scheduling 1:1 meetings with the top leader of that division/unit
 - Shared with them a relevant spreadsheet to share all of their data they could think of "off the top of their head", then coming back two weeks later with more targeted questions
 - "What PII may be present in this data set?"
 - What IS "PII"?
 - Connections with 3rd party vendors
 - It was a priority of the agency head, so we had a lot of support
 - Annual review
 - Every december, we check in with each unit to see how things changed, adding and removing datasets as necessary

General Privacy Questions

- Surprise with how many datasets we collected were not noted in active retention schedules
- Popular misconception that datasets are not subject to retention schedules etc
 - Leads to an opportunity to educate folks about the legal risk of indefinite retention of records

- State officials have set up guard rails around PII retention, and now we can use that as leverage
- Data sets – Mostly talking about a microsoft access database or a sql table
 - At DOT, highway construction data was being documented
 - We have to be aware of the data we're collection, like SSNs or EINs, and what the requirements are for how it is disposed
- MN: For active records—what about things that have a disposition to the archive—how is the PII in that record handled?
 - Christian: I am coming at this from just being concerned with our specific agency's data, that they created, for active records, but a portion of that will be transferred in the future
 - Important Caveat: Agencies produce draft retention schedules for their agency, and those are submitted to the archive for approval
 - In terms of transfers to the archive, it allows us to appraise those situations early on.
 - MD Privacy Protection laws

Privacy Policy toolkit available?

- Christian: I have a reference document of what PII is in plain language that I can share with vendors
 - State of Maryland Privacy Law that identifies PII:
[https://govt.westlaw.com/mdc/Document/N2E2E4580C56311EAB1B5B491D06C2573?viewType=FullText&originationContext=documenttoc&transitionType=CategoryPageItem&contextData=\(sc.Default\)](https://govt.westlaw.com/mdc/Document/N2E2E4580C56311EAB1B5B491D06C2573?viewType=FullText&originationContext=documenttoc&transitionType=CategoryPageItem&contextData=(sc.Default))

CJISSECPOL

- Question: Is anyone working with records that are subject to the FBI's Criminal Justice Information Services (CJIS) Security Policy (CJISSECPOL)? If so, how are you meeting CJISSECPOL requirements for those records?
 - Christian does not have any data on his data inventory that references this policy.
 - Respondent: There are a lot of restrictions on what you can do with that data—how are you protecting that data? –CT is trying to procure this data now, but there are only 2 or 3 state archives that have addressed this problem...the systems that can handle the restrictions they have, are very restrictive—almost impossible to meet
- How has it been used to minimize data collection
- Retention Implications

- Compiling this data has been great for identifying retention gaps/failures/mismatches
 - Sometimes internal policies are thought of as being “laws” or “regulations”, but there is no legal schedule for it.

Databases as Records

- Fields that flag privacy material specifically
- With a database, it may collect a single piece of information about an individual, but when it is combined, internally, or with multiple databases, once you have combined multiple pieces of PII, they become protected
 - We currently flag it in the inventory, but not the db itself—that’s an area of improvement
 - We are currently in an area of ambiguity around whether the DB is a convenience copy or a record, the data itself could be a public record
 - Geospatial data layers is my particular interest here—from having relational geo-spatial data would be amazing to have 100 years from now, but flattened versions of those files
- It’s somewhat doable for PII in text documents, but the AV files are a lot more challenging
- WI: here the retention schedules should identify when the PII exists, and if it does, we handle it. For eRecords, we make it only accessible onsite, in the reading room
 - Always a moving target
 - WI: Right, we’re building out a new DAMS, and then we ask, does this also apply to our private collections and various other manuscript collections
 - CS: Records are public, unless there’s a specific law that says otherwise
 - NC: for eRecords, if it is transferred to us, it has to be permanent
 - Once it’s in our custody, it has to follow the statutes that exist—the other side is that we have private collections, but not state agency records, so people can dictate if this is available and open to the public—but we always do our best to ensure that protected PII is redacted
 - CS: So we also run the state records center—the non-permanent record holding for state agencies (strictly physical records)--any access to those is governed by the record creator.
 - WI: As we digitize more of those older collections, these were not even cases that were considered when the collection was created, or the schedules were created.

- GR: EPADD has a feature that allows for obfuscation of text that could be really good for allowing access to unprocessed collections
- CS: FPE brings up in the chat that we can just use “current laws that are set in place”. The SERI community is also a great resource for this in terms of predicting what future restrictions may be—having a connection with all of the other states archives regulations and proactively flag certain information in the case that it does in the future
 - The other thing I’m thinking about is how the aggregation or combination of data DOES change the classification of data. If a 3rd party takes all of those data pieces and puts them together, then there is a risk that they have now created restricted records.

Closing Session

What are some of the takeaways that you learned from the sessions?

- Good is better than perfect.
- The cost of doing nothing is almost always greater than the cost of doing something
- It's possible to have a great relationship with IT

Track 3 - New frontiers

Session 1

- Cross pollination between the two tracts
- Ideas for things we can develop
 - Needing to develop a standard language in the field - SERI whitepaper, outlining basic terms
 - Training
 - Tracking of who is using AI and how
 - Agentic AI - AI with agency to make and act on decisions
- Different types of AI - sorting, ChatBots (virtual reference archivists)
- Pitfalls of introducing the technologies
- Detection of restricted information in records - may be useful but also introduce potential legal risks. Right now, there still needs to be some human interaction. If you're allowing the use of your data to *train* tools, it may allow sensitive data to enter into the public realm unintentionally
- Limited data right now on who may be using these tools

Session 2

- Getting sensitive data identified at the record creator side so that it reduces the risk of inadvertently revealing PII or other sensitive data at the point of access
- Stay in touch with legal council - rely on them to be aware of current laws, but also be in touch with SERI/CoSA so that you are also aware with developing laws and what might be a risk for you in the future
- Large sets of data placed online may expose PII that we wouldn't want to be exposed

Track 2 - Capture Conversations

Session 1 - social media, websites, and data

- Content over format - maybe start advising that social media is the secondary method of sharing policy so that the official record is more stable
- When using tools, be aware of the risk
- To do
 - Take down notice policy document from CoSA?
 - Social media usage
 - May 29 Tech & Tools

Session2 - Cloud

- Good is better than perfect
- Balance your requirements so that barriers still welcome transfer of records
- It is possible to have a good relationship with state IT, but it takes work

Track 1

Session 1 - Advocacy

- Know your audience - know their goal, how can your work fit their goal
- Establish common definitions and language at the beginning
- Workflow - do internal policies then work with external stakeholders
- Proceed with confidence - you are the expert even if you don't have it all figured out right now (they don't need to know that!)
- Use a carrot & a stick
- Be aware of legislation coming through sessions - educate yourself about bills and what they may be missing

Session 2 - Getting started

- Don't let finances be a barrier
- Don't let staffing be a barrier
- Create training tools
- Find a discreet project that you can successfully complete
- Don't reinvent the wheel - expand what you're already doing to include