Building E-Discovery into your System/Process

★ ★ ★

Eric F. Stein
Global Information Services (A/GIS)
June 29, 2023

U.S. DEPARTMENT of STATE

1

★ ★ ★

What is eDiscovery?

- **e-discovery** noun
  e-dis·cov·ery ˈē-dis-ˈkə-və-rē
  : discovery of records and documents (as e-mails) kept in electronic form
  https://www.merriam-webster.com/legal/e-discovery

- **Electronic discovery** (also **ediscovery** or **e-discovery**) refers to discovery in legal proceedings such as litigation, government investigations, or Freedom of Information Act requests, where the information sought is in electronic format (often referred to as electronically stored information or ESI).[1] Electronic discovery is subject to rules of civil procedure and agreed-upon processes, often involving review for privilege and relevance before data are turned over to the requesting party.
  https://en.wikipedia.org/wiki/Electronic_discovery

U.S. DEPARTMENT of STATE   2

2

★ ★ ★

Why does it matter?

- Find needed information in a timely manner
- Make connections between existing data and information
- Assists in responding to requests for information
- Electronic records and data grows exponentially every second
- Technology is complex
- Often challenging to search electronic databases – why?
  - Do we know what we are looking for? Targeted searches and broader ones.
  - Unstructured data and information
  - Different kinds of data and information
  - System(s) was not designed to do anything other than store information
  - New sets of data and information are being captured (outside of what was originally intended)
  - Too busy to make changes or updates to system
  - Lack of resources to update systems or make them interoperable

U.S. DEPARTMENT of STATE   3

3

Three System and Process Scenarios

1. Planning for a new system (including just an exercise of planning for your ideal system even if no resources available)

2. Modifying an existing system (change to one system)

3. Modifying existing systems (connecting two or more systems to make them interoperable)

U.S. DEPARTMENT *of* STATE

4

4

Planning for a new system (including just an exercise of planning for your ideal system even if no resources available)

What do you need from an eDiscovery tool?
- Locate information
- Deduplication of information
- Transfer of information identified to another IT tool or system
- Redaction capability
- Ability to capture records or data from different sources
- Agency specific considerations
- Records and data considerations – types of files – documents, images, videos
  - Ability to search across everything
  - Just capture everything
- Consider a Request for Information (RFI)/market research – NexGen FOIA Showcase 2022
- Define very specific and clear requirements
  - Program and functional requirement
  - Technical Requirements

Also worth doing if you are just thinking about a new system – can help identify core issues or needed improvements

U.S. DEPARTMENT *of* STATE

5

5

Planning for a new system (including just an exercise of planning for your ideal system even if no resources available)

State Example – eRecords

- Appendix – 2019 eRecords ASAP Slides - The State of State's E-Records: Capture and Archive

- Office of Management and Budget (OMB) and National Archives and Records Administration (NARA) Mandates – 2016, 2019 M-19-21 Transition to Federal Records

  - 2016 – Manage Emails
  - 2019 – Manage Electronic Records Electronically
  - 2024 – M-23-07 – https://www.whitehouse.gov/wp-content/uploads/2022/12/M_23_07-M-Memo-Electronic-Records_final.pdf

- Challenge – Transition from paper-based print and file policy to electronic records management

U.S. DEPARTMENT *of* STATE

6

6

★ ★ ★

Planning for a new system (including just an exercise of planning for your ideal system even if no resources available)

State Example – eRecords

- eRecords Archive
  - Manages Department's state.gov email and permanent electronic records
  - Receives over 2 million unique records daily; currently has over 3 billion records
  - Gathering requirements
    - Driven by the OMB/NARA mandate to manage emails
    - Identified key intra-agency partners – program offices, IT office, budget office
    - Reviewed information captured in emails
    - Considered data and metadata standards during system development
    - Gathered functional requirements
    - Gathered technical requirements
    - Imagined how the records and data could be used immediately with other systems – eRecords and a FOIA tool (and other document production)
    - Wanted to be able to do centralized searches of this new system; deduplicate search results to avoid inefficient reviews of the same emails/records so more time can be used to review records for possible release
    - Considered machine learning and AI capabilities during system development – how could such tools be developed; data standards critical

U.S. DEPARTMENT *of* STATE   7

7

★ ★ ★

Modifying an existing system (change to one system)

Questions
- Why?
- What do you need it to do?
- What problem(s) are you trying to solve?
- Have you asked other offices within agency about solutions?
- Have you asked other agencies?
  - Contact Technology Committee - https://www.foia.gov/chief-foia-officers-council/committee/technology-committee
- Is your system capable of the changes that you are seeking?
- Do you have support for a new system?
- Do you have resources to support it?
- What data are you capturing?
- What data do you want to capture?
- Is capture enough? How else can this information/data be used?
- Are there partners who could benefit from this tool?
- Is your data structured how you need it?

U.S. DEPARTMENT *of* STATE   8

8

★ ★ ★

Modifying existing systems (connecting two or more systems to make them interoperable)

Considering points from previous slides on system requirements

Process considerations
- What is the workflow or process from end-to-end?
- How does the technology support the workflow?
- Should there be AI capabilities in both systems?
  - Possible issues or conflicts of different systems being connect
    - May not be possible
    - May create new problems
      - Both systems have AI so we will now have super AI, right?
    - Connecting systems may affect the technological abilities of another system
    - Unanticipated problems.
- Do you need two systems?
- Do you need to connect more than two systems together?
  - Has the end-to-end process been considered?
  - Does process need to be changed? Are the right people involved in discussions?

U.S. DEPARTMENT *of* STATE   9

9

★ ★ ★

Other Considerations

- Not all IT systems at agencies are connected complicating eDiscovery
  - Unlikely most agencies will have one central search function
  - Will need to consider the ability to search centralized and decentralized systems
    - Where can all this data be stored?
    - How can it be accessed? By whom?
    - Is it possible to make connections between these different data sets (records, information, etc.)?
- People drive process
  - Consider process before technology. What are you trying to do? Iterative approaches/testing
  - Design elements – users and customer experience
- New data/information and legacy/existing data/information
  - Make a plan for a specific day forward.
  - Make a separate plan for legacy/existing data/information
- Don't be afraid to ask questions – talk to a colleague, supervisor, Tech Committee
- Change takes time

U.S. DEPARTMENT of STATE

10

10

★ ★ ★

Discussion

- Contact information – Eric F. Stein, SteinEF@state.gov
- Thank you!

U.S. DEPARTMENT of STATE

11

11

★ ★ ★

Appendix – 2019 ASAP Slides on eRecords

U.S. DEPARTMENT of STATE

12

12

The State of State's E-Records: Capture and Archive

ASAP Presentation
July 24, 2019

UNCLASSIFIED   13

13

Session Overview

Session 3.06 –

Intermediate/Advanced Track –

The State of State's E-Records: Capture and Archive

The State Department invested in a centralized, cloud-based eRecords archive
that will be interoperable with the Department's FOIA processing solution.
Learn how the State Department has integrated powerful search tools and
artificial intelligence into its eRecords archive to streamline the FOIA process.

-Eric Stein, Dept. of State
-Tim Kootz, Dept. of State

UNCLASSIFIED   14

14

State Department Overview

• The State Department conducts diplomatic relations and sets foreign
  policy for the United States.

• 276 Diplomatic and Consular Posts dispersed across 191 countries

• 100,000 Personnel use our IT systems

• We send and receive approximately 6-8 million emails per day

• ~1.2 million emails per day after de-duplication.

UNCLASSIFIED   15

15

Drivers for Change

- Managing Government Records Directive (MGRD) – A presidential directive signed November 28, 2011.
  - Manage email electronically by December 31, 2016.
  - Federal agencies will manage all permanent electronic records in an electronic format December 31, 2019.
    - Success Criteria –
      - Policies
      - Systems
      - Access
      - Disposition

- The key takeaway – If it is born digitally, it needs to stay in a digital format through its lifecycle.

UNCLASSIFIED   16

16

Connecting Records Management and Access Programs

A robust records management program is a force multiplier for information access programs

Creation or Receipt of a Record	Maintenance of a Record		Disposition of a Record	
Create/Receive 01	Use 02	Store 03	Disposition 04	Preserve / Destroy

UNCLASSIFIED   17

17

eRecords: The Department's Centralized Repository

1.2M
Unique emails sent and received per day from around the world

1B
Emails stored in the eRecords archive

Cloud based
- Big data architecture
- Performance and uptime

Robust metadata enrichment
- Marking tool
- Active directory

Open Source tools are the backbone of eRecords
- Scalable
- Non-proprietary

UNCLASSIFIED   18

18

eRecords Roadmap

Further enhance eRecords to meet OMB M-19-21 2019 deadline:
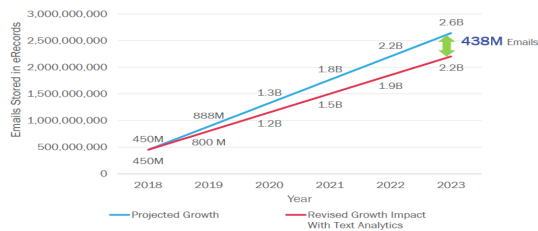
- Index and enrich other objects – All permanent records (not just email)
- Migrate legacy archive content into eRecords by end of FY19
- Simplify the electronic records transfer process – Automate capture as much as possible
- Seamless integration with FOIAXpress
- Continue to leverage machine learning and other forms of artificial intelligence
  - Records categorization
  - Classification recommendations
  - Technology assisted review
  - Near duplicate recognition

UNCLASSIFIED   19

19

Estimated Growth

19
888M
1.3B
1.8B
2.2B
2.6B
450M
800 M
1.2B
1.5B
1.9B
2.2B
438M Emails

Emails Stored in eRecords

Year
2018 2019 2020 2021 2022 2023

Projected Growth — Revised Growth Impact With Text Analytics

UNCLASSIFIED   20

20

Other Key Factors

- Marking Client – Unclassified and Classified Emails
- eRecords Case Dashboard
- Executing a Search – Current abilities
  - Email Rendering
  - Threading
  - Export (interoperability)
  - Artificial Intelligence (AI)
- Future capabilities
  - AI – Technology Assisted Review (TAR)
  - AI – Natural Language Processing
  - AI – Machine Learning

UNCLASSIFIED   21

21

6/25/2023

Contact Us

Eric F. Stein – SteinEF@state.gov

Timothy Kootz – KootzTJ@state.gov

Secondary Contact Information:
Records@State.gov

UNCLASSIFIED   22

22