

ERA 2.0: The National Archives New Framework For Electronic Records Preservation

What is ERA?

- The United States National Archives and Records Administration (NARA) Electronic Records Archives, or ERA, encompasses a set of the operations for scheduling, transferring, reviewing, describing, transforming, and storing electronic records received from U.S. government agencies.
- It comprises multiple systems handling different types of electronic records: unclassified and classified Federal, Presidential, and Legislative, each with its own regulatory management and access controls, alongside the tools needed to support functions across the archival life cycle.

The Current ERA System

- ERA was developed through a large-scale project in the early 2000s, which involved business owners from across the agency who identified functional requirements.
 - The project was scoped with a single system in mind to deliver all aspects of functionality from records scheduling by agencies through public delivery.
- The current system supports records scheduling submission and approval, records transfer by agencies, a single processing workflow for federal records, and the storage and management of records.
- It is used by dozens of NARA staff and by records managers at over two hundred agencies that schedule and transfer records to NARA.
- The project successfully leveraged work from the community, including the integration of open source tools such as PRONOM, DROID, and JHOVE.

The Current ERA System

- There were several lessons learned from the implementation.
 - The scale of the project (over 1400 complex functional and regulatory requirements) made this difficult to achieve in the timeframe originally envisioned.
 - When the system went into production in 2008 it did not include all the planned-for functionality, including subsuming other customized NARA production systems or providing the full range of necessary preservation functions.
 - There were challenges in maintaining communication about the status of the development to several hundred staff and colleagues at federal agencies. This required more aggressive communication strategies than were used.
 - There was also a large scale change management effort which was not solely technical but also organizational, as the roles of some staff would change with the launch of the system.
 - Soon after going live, ERA development was put on hold to review the state of the system and plan for future work. It was determined in 2012 that the system required extensive modernization and refactoring to improve the user experience and add needed functionality.

ERA 2.0 Project Goals

- The high-level goal for the current project is to improve the efficiency of the process moving from records scheduling to public access. In enabling new levels of efficiency and automation.
- ERA 2.0 cannot be a single system solution for all operations; instead there must be a modular suite of tools – commercial, open source, and developed in-house - that are interchangeable with new tools in response to technology changes, new business needs, and to handle new record formats that will appear in the future.
- The system must be interoperable with other tools and systems at NARA, exchanging metadata and data through APIs.

ERA 2.0 Implementation

- The project follows an incremental “Agile with Discipline” approach, where a set of User Stories are documented instead of a massive list of fixed requirements, and development takes place in four-week sprints.
- The User Stories provide implementation guidance and acceptance criteria for the functional requirements, and each four-week sprint both closes stories and surfaces requirements for new stories.
- The process leverages work done to date, but is more transparent than in the past, and provides for a much greater level of participation for NARA staff to guide the development.
- The project began in fall 2014, and is being implemented in conjunction with IBM.

ERA 2.0 Modules: DPE

- The Digital Processing Environment (DPE) supports the ingest process, and provides the software tools and metadata editing capabilities necessary for electronic records verification and processing functions.
- The initial implementation includes approximately fifteen tools, both embedded thin client as well as thick client tools accessed through virtual machines provided for each processing archivist. Initial tools include those for format characterization, bulk file reformatting, image manipulation, common business productivity software, PII recognition, and redaction.
 - The list currently includes, but is not limited to, DROID, ImageMagick, Adobe Photoshop, MS Office, and several internally developed tools.

ERA 2.0 Modules: DOR

- The Digital Object Repository (DOR) will provide all holdings management and preservation functions, including recording of fixities, object versioning, auditing, and reporting.
- When additional processing is needed, such as bulk preservation actions or creation of new public use versions of records to meet researcher requests, files are copied into the DPE environment for the work to take place, then versioned back into the DOR.
- DOR maintains the audit trail for all actions performed automatically or manually in its data warehouse.
- DOR will include advanced search functions for staff of both metadata and record content to support highly granular discovery requirements to review records for Personally Identifiable Information (PII), classification status, and to find responsive records for Special Access or Freedom of Information Act (FOIA) requests, and access requests mandated by law to support litigation or to supply access to government information that is not yet publicly available.

ERA 2.0 Modules: BOM

- The Business Object Management (BOM) in the current system includes only a single workflow for federal records that excludes other record types.
- ERA 2.0 will comprise multiple workflows for Federal, Presidential, Legislative and Judicial records, digitized analog records, and donated materials, with a mechanism to easily update workflows or instantiate new workflows as requirements change.

ERA 2.0 Digital Preservation Functions

- Digital Preservation User Stories were written against the Open Archival Information System standard (ISO 14721) and the Trustworthy Digital Repositories standard (ISO 16363), and validated against TRAC and DRAMBORA as proposed assessment tools. Examples of User Stories include, but are not limited to:
 - As an Administrator, I want to be able to monitor the digital object inventory in the DOR and maintain reports on the inventory.
 - As an Administrator, I want to receive regular reports on all incidents of data corruption or loss and steps taken to repair/replace corrupt or lost data.
 - As an Administrator, I want to receive regular reports on any events or maintenance that occur within and on the system-level operations that might impact the digital objects in the DOR (i.e. major system upgrades, electrical malfunctions, unexpected system downtime, etc.) and steps taken to repair/restore system.
 - As an Administrator, I want to have reports on the hardware and software and locations that are supporting the back-up functionality necessary for preservation and that the system is sufficient for the DOR services to be maintained at a low-risk of data loss.
 - As an Administrator, I want to ensure that not one person has write access to all copies.
 - As an Administrator, I want to be able to demonstrate that any/all multiple copies of digital objects are synchronized and maintain reports on a periodic basis.
 - As an Administrator, I want to be able to ensure that software and hardware systems are capable of maintaining and recording provenance information, and that these systems are reviewable in the future.
 - As an Archivist, I want to be able to demonstrate the provenance of digital objects through downloadable reports documenting their traceability from receipt and each interaction they have been subject to.
 - As an Archivist, I want monitoring and notification when digital objects approach obsolescence.
 - As an Archivist, I want to view the selection of the digital materials for deletion in order to verify and review the selection prior to deletion.
 - As an Archivist, I want to ingest transformations of digital object(s) stored in DOR in order to store reference and preservation copies of digital objects.

ERA 2.0 Implementation

- Each module – DPE, DOR, and BOM – is an independent codebase; the goal is to implement modules that can be updated or replaced without massive dependent code changes in the other aspects of the system.
- The development, Pilot, and to-be production system are all in the Amazon Web Services cloud.
- The implementation is a custom Java application but leverages open source tools in the stack and will, at least for unclassified records, be cloud-based.
- Government agencies are increasingly managing their electronic records in the cloud, and NARA would like to be able to take advantage of an interpretation of the “Data at Rest” concept where records stored by agencies in the cloud do not need to be repeatedly moved to come under NARA control, be processed, or made accessible.
- Processing archivists will be working entirely in the cloud, utilizing embedded tools for bulk actions such as format characterization and reformatting, and working in virtualized workbenches directly on the files in the cloud to take advantage of elastic compute capabilities.

Current ERA 2.0 Status

- In 2015 we launched a Pilot Digital Processing Environment with an updated workflow and, a new modular processing environment with a suite of utilities and tools, and a Pilot Digital Object Repository for managed preservation services and storage.
- The Pilot - now in its fourth release - is being tested by over 100 NARA staff members to ensure that this next generation of ERA fits into the new federal frameworks for electronic records management and transfer to NARA.
- Development is continuing; the planned production launch is currently scheduled for early 2018. Iterative development will continue after the initial launch, as will the development of an accompanying separate environment for Classified records.
- The ultimate goal is to greatly expand the functionality of ERA to include workflows for all record types and digitized records, greatly expand the range of tools needed to process records and create public use versions, and to implement a robust preservation environment, all in a framework that makes it easier to update the system to meet as yet unknown needs.