



ERA 2.0: A Cloud-Based Preservation Repository at the National Archives

Leslie Johnston, NARA
Designing Storage Architectures for Digital Collections Meeting
September 17, 2018

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, and storing electronic records received from U.S. government agencies.
- It currently includes multiple systems handling different types of electronic records: unclassified and classified Federal, Presidential, and Legislative, each with its own regulatory management and access controls.



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.
- Processing archivists work entirely in the cloud in virtualized workbenches directly on the files in the cloud to take advantage of powerful compute capabilities.
- The initial implementation includes approximately fifteen tools, some embedded in the system and some integrated as thick client tools in the VMs used by each processing archivist. Initial tools include those for format characterization, bulk file reformatting, image manipulation, and common business productivity software.
 - The list currently includes, but is not limited to, DROID, Adobe Photoshop, Microsoft Office, and several internally developed tools.

ERA 2.0 Modules: DOR

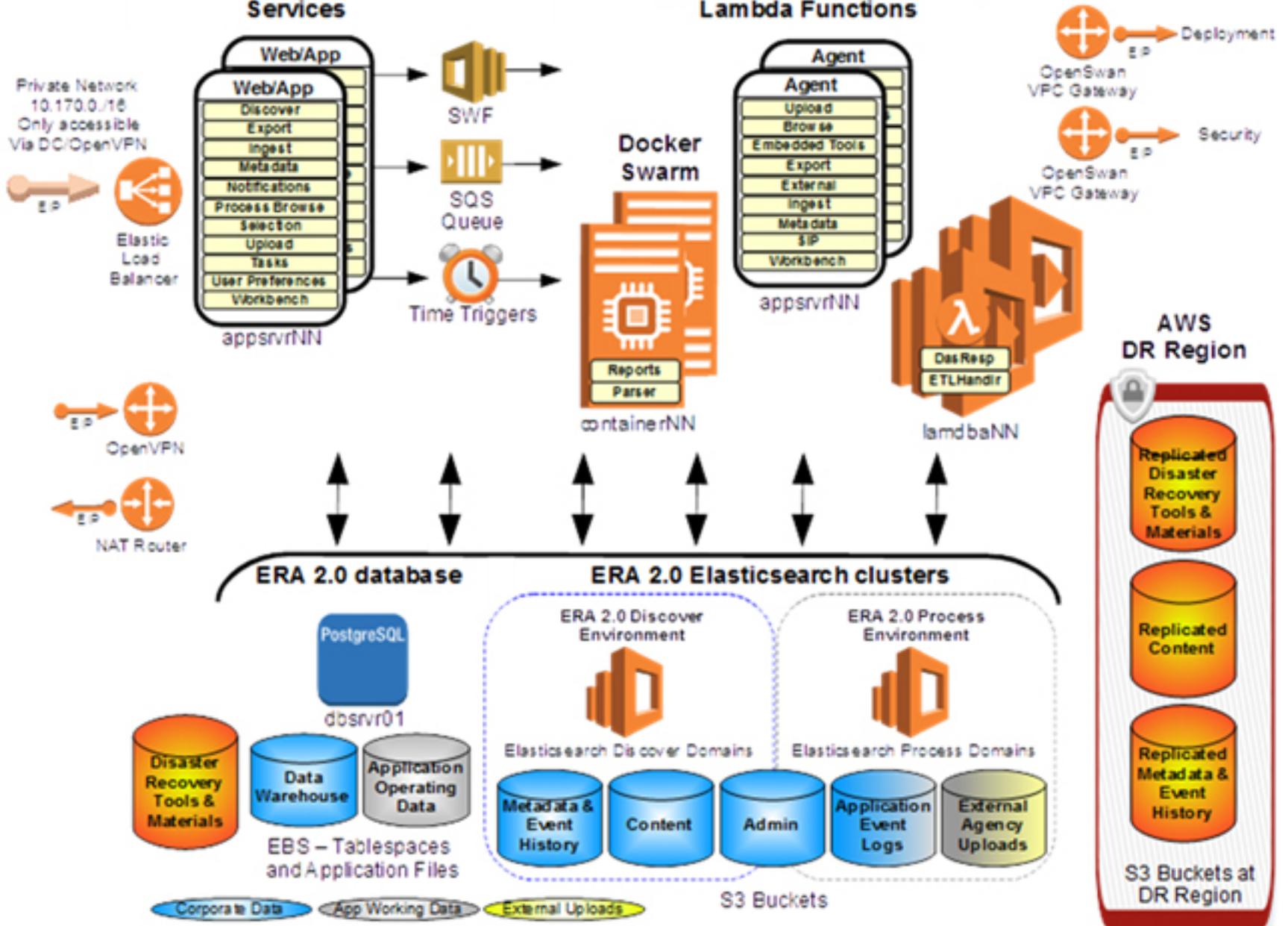
- The Digital Object Repository (DOR) provides all holdings management and preservation functions, including recording of fixities, object versioning, auditing, and reporting.
- When additional processing is needed, such as bulk preservation actions or creation of new public use versions of records to 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 has both metadata and item content full text advanced search functions for staff of both metadata and record content to support requirements to review records for Personally Identifiable Information (PII), classification status, and to find responsive records for Special Access, Freedom of Information Act (FOIA) requests, or Declassification tasks.

ERA 2.0 Modules: BOM

- The Business Object Management (BOM) in the current system includes only a single workflow for federal records.
- ERA 2.0 will have multiple workflows for Federal, Presidential, Legislative and Judicial records, digitized analog records, and donated materials, with a mechanism to easily update workflows or create new workflows as requirements change.
- The work to integrate a new workflow engine using Alfresco is beginning in October 2018 following a several months long Business Process Re-engineering project.

Web and Application Services

Background Agents, Containers, & Lambda Functions



What Else Has NARA Done This Year?

- Instituted a new framework for identifying file format risks and prioritizing preservation actions.
 - Presenting a paper on this at iPres 2018; the paper will be in the Proceedings; the paper and spreadsheet will be made available after the conference.
- Instituted a template for digital preservation and digitization SOPs and created an agency-wide SOP library.
- Started an ISO 16363 self-assessment using the Primary Trustworthy Digital Repository Authorization Body (PTAB) instrument.



Discussion

Leslie Johnston

Director of Digital Preservation

U.S. National Archives (NARA)

Leslie.Johnston@nara.gov