



# What the National Archives has been up to for the past Two Years

Leslie Johnston, NARA

Designing Storage Architectures for Digital Collections Meeting

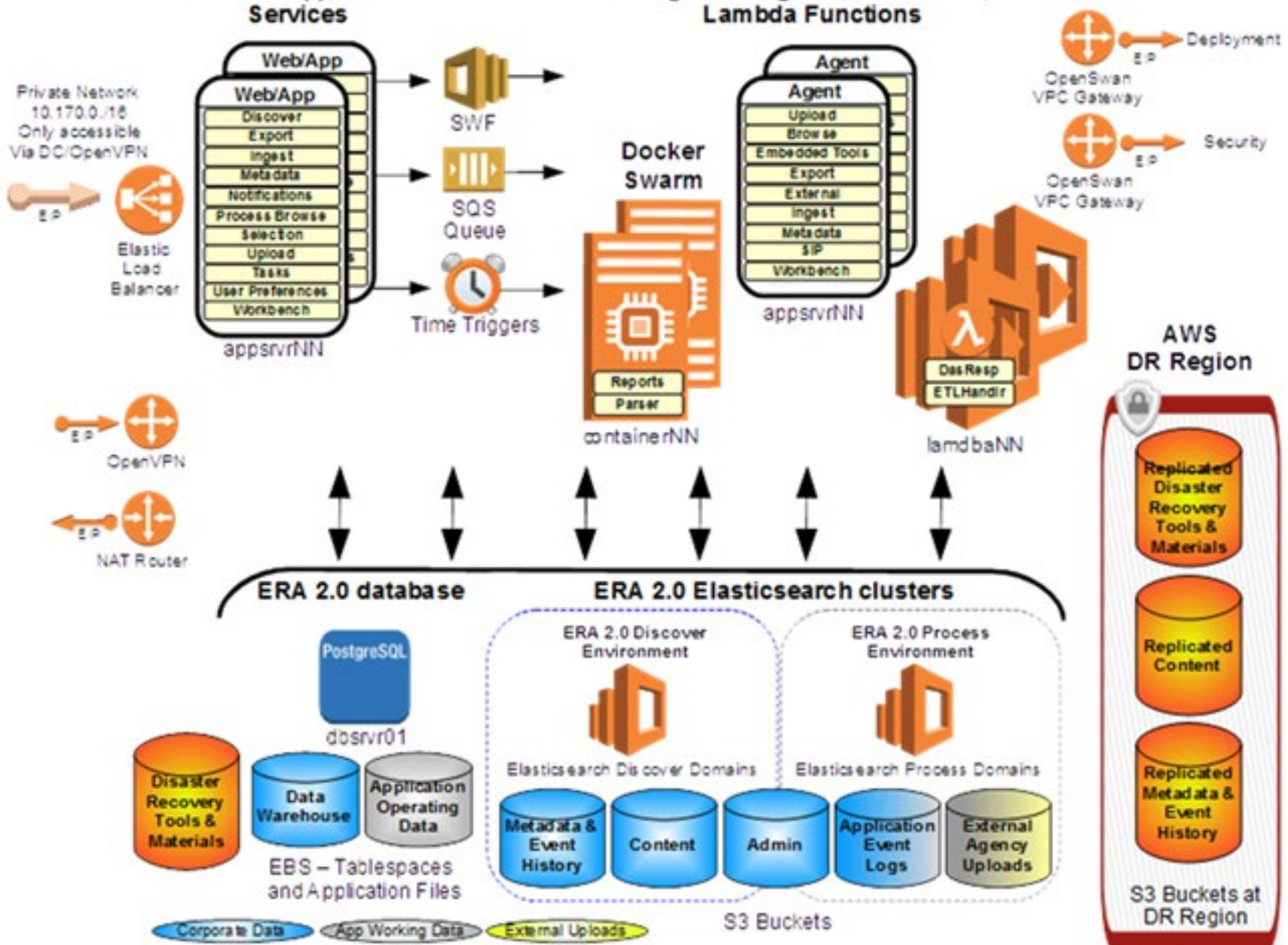
March 2022

# ERA 2.0 Development has Continued

- ERA 2.0 is the AWS GovCloud-based electronic processing environment and preservation repository at NARA, currently storing 16.6 million files in 100 TB. Across all the ERA systems, we have 1.93 billion files in 1 PB of storage.
- Recent initiatives:
  - A major migration of federal records that had not been ingested into the old ERA system. Several other migrations to the cloud are underway or planned.
  - Improving local Snowball and distributed Direct Connect workflows for transferring and moving files for ingest in facilities across the country. Several digitization units are now loading surrogates.
  - PIV authentication
  - The integration of additional embedded tools and new tools in the processing VMs.
  - Documentation

## Web and Application Services

## Background Agents, Containers, & Lambda Functions



# Cloud-to-Cloud

- Agencies need to move from their cloud environment to the NARA environment without the added risk of numerous file copying actions in and out of local and cloud environments
- The initial implementation is in Amazon GovCloud. Many agencies are asking for Azure, given the extensive use of Office 365 and SharePoint-based electronic records management systems.
- The first large-scale external test is underway with the Census Bureau to transfer the 2020 Census files.
- The first large internal test is for the 1950 Census, copying the digitized records from a different S3 bucket into the repository.

# Business Objects and Record Schedules

- The Business Object Management (BOM) 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 and migrate all existing schedule and transfer data will be completed in 2022.

# Linked Data

- After receiving international community feedback on its proof-of-concept, NARA is preparing to release its Digital Preservation Framework—a risk assessment matrix and electronic record preservations plans—as Linked Data, with tight integration into WikiData for Digital Preservation.

# Program Assessment

- Completed its second ISO 16363 self-assessment using the Primary Trustworthy Digital Repository Authorisation Body (PTAB) instrument.

Metrics Status	2019	2021
Meets	32	52
Partially Meets	64	54
Does Not Meet	13	3



# Thank you!

Leslie Johnston  
Director of Digital Preservation  
U.S. National Archives (NARA)  
[Leslie.Johnston@nara.gov](mailto:Leslie.Johnston@nara.gov)