Identifying Requirements for an Electronic Records Processing Environment (aka moving from Waterfall to Agile in a [hopefully] not too painful process)

Leslie Johnston, NARA
2014 Digital Library Federation Forum
Optimized Ingest Framework (OIF)

A new model for managing electronic records for preservation and access

- Modular approach to NARA systems managing digital materials
- Departure from model of a single, monolithic system to manage the ingest, processing, preservation, and access to electronic records.
- Sustainable approach for necessary evolution of existing systems.
- Establishment of an environment to provide necessary processing flexibility and tools for a wide variety of digital files and metadata.
- More robust solution for digital preservation with reduced complexity.
- A stepping stone toward greater automation in managing digital files.
Optimized Ingest Framework (OIF)

Concepts developed in a series of 2013-2014 meetings:

- Need for a flexible processing environment with an expandable set of software tools to verify and process electronic records.
- Requirement to manage increasing volume and varieties of digital files.
- No current workflows for non-Federal digital materials (e.g., digital surrogate masters, Legislative, Donated, Supreme Court, etc.).
- Growing storage and access needs for digital surrogates.
- Need for a more efficient means to provide public access to ingested electronic records from agencies.
- Need to explore & develop capabilities and processes to address projected growth – Presidential Directive implications, Data-at-Rest solutions.
NARA Lifecycle Systems

DAS: Metadata
BOM: Record schedules, Transfers, Workflows

Digital Processing Environment (DPE)
Processing, transforms, and storage of digital materials
Packaging for other systems

Digital Object Repository (DOR)
Information package processing
Preservation storage and data management
Manage external data-at-rest

Online Public Access (OPA)
Public Access (cloud)

SIP - Submission Information Package
AIP - Archival Information Package
SEIP - Search Engine Information Package
DIP - Dissemination Information Package
So Far This Looks Pretty Much like Every Other Project...

But it won’t be like other projects at NARA

This will be the first fully Agile project for the organization.

And the first to be implemented for a cloud/on premise hybrid production environment.
Opportunities to Pilot a New Preservation Architecture

- Quickly test and refine assumptions made in the redefinition of processing workflows and selection of tools for processing.

- Develop data packaging specifications and APIs for the exchange of metadata and files between multiple systems.

- Design for both a cloud (unclassified) and on premise (classified) environment.

- Test the use of a range of file processing tools for both a cloud and on premise environment for processor, memory, and storage needs.

- Benchmark network transfers between on premise and cloud environments and between cloud environments.
Opportunities for the Organization

- Introduce the IT organization, business owners, and agency management to the Agile methodology.

- Demonstrate how requirements, risks, and tasks documented in a Waterfall style can be re-purposed for Agile processes.

- Introduce new processes and communication styles and tools and staff roles – which cannot help but cause some anxiety.

- Improve the efficiency of electronic records ingest and make records available to the public for use more quickly.
Let’s Discuss...

Leslie Johnston
NARA
leslie.johnston@nara.gov