Program Manager's Update

Advisory Committee on the Electronic Records Archives

April 7- 8, 2010

Agenda

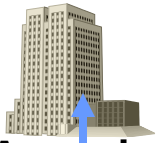
- ERA Overview
- Increment 3 Scope of Work and Status
- Remainder of the Contract

ERA Overview


Agencies


White House


Committees


Agencies

Base Instance

EOP Instance

Congressional Instance

Records Center Instance

Intra-System Communication

System Management

Network

Help Desk

Data Management

Preservation Framework

Public Access

Shared Services


Citizens

Current capability: solid fill

Future capability: hashed fill

Increment 3

- ✓ Based on System Evolution brief to Lifecycle Guidance Team January 2009
- ✓ Provide better foundation for “Business Workflow” and support for a re-engineering of NARA Business processes. Key to scaling life cycle record processing.
- ✓ It will allow for expansion to more agency support and faster Ingest processing.
- ✓ A combination of “system orchestration” and flexible workflow will form the basis for future support of preservation framework
- ✓ Implement original ERA vision of flexible, evolvable, scalable and open framework.

Increment 3

- Scope of work
  - ✓Software Enhancements (Releases 2.2.7, 2.2.8, 2.2.9, 2.2.10)
  - ✓EOP Support
  - ✓Congressional Records
  - ✓Online Public Access
  - ✓Preservation Framework Prototype
  - ✓ERA Base Architecture Evolution

Increment 3 Status

- ERA Base Software Enhancements
  - Enhancements to address authorized user defined changes and software defects not addressed at IOC
  - Necessary for Pilot 2
  - Four releases
    - ✓ 2.2.7 – deployed 4/18/09
      - Sequential Transfer shipment #'s
      - Packaging Tool Enhancements
      - Transfer Shipment Aggregate Report or TSAR
      - Don't store inner composites
    - ✓ 2.2.8 – deployed 7/31/09
      - Accounts vs # agencies
      - Formatted View and Print
      - Search updates for BO
      - RSI changes
    - ✓ 2.2.9 – deployed 12/14/09
      - Increase Transfer Limit
      - Ingest Verification
      - TR Updates
      - DTT Verification
    - ✓ 2.2.10 – deployed 3/6/10
      - Expanded browser support

Government Wide Expansion

- Initial Phase
  - ✓ Four collaborating agencies
- Enhanced Pilot Phase
  - ✓ 25 additional agencies
  - ✓ January 2010
- Voluntary phase with NARA acceptance July 2011
- Mandatory Phase Target date FY2012

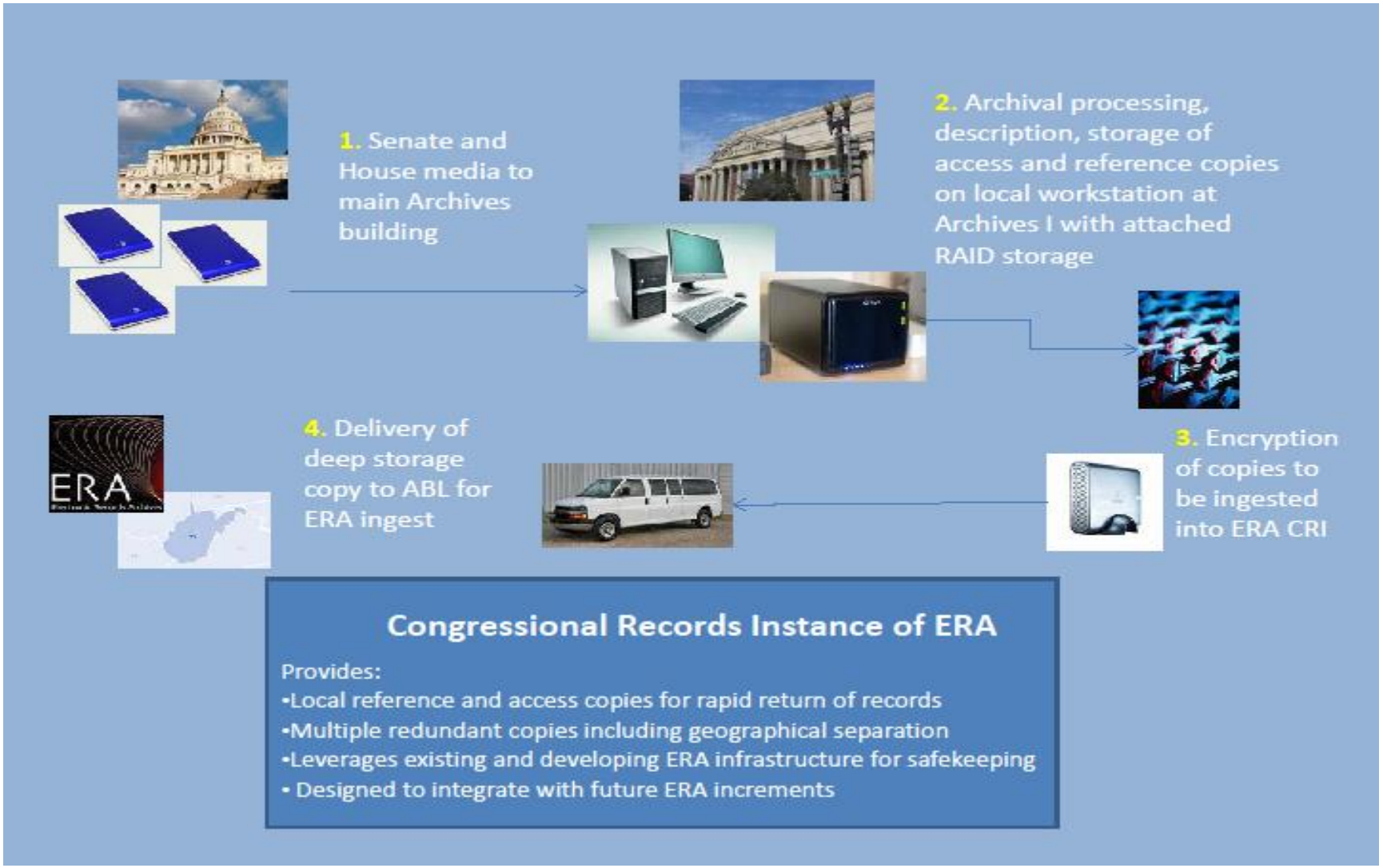
Increment 3 Status

- ERA EOP
  - Working with EOP on transport and ingest of restored emails related to settlement of CREW v. EOP et al; includes three sets of email (21 low days, 40 sample days, 33 settlement days).
  - ERA_2.1.11.0: Provides the ability to search and browse of the shared directories.
    - ✓ Deployed January 2010
  - ERA_2.1.11.1: Enhanced browse for shared directories (paging), provides an auditing tool for SASS Staff at Rocket Center and adds a SAS name for new data to arrive in June
    - ✓ Deployment target is June, with Handoff to NARA I&T in late May
  - ERA_2.1.12.0: Enhanced ability to search MerlinOne data,
    - ✓ Tentatively Includes upgrade to HCAP software version
    - ✓ Deployment date is being developed, target is fall 2010
  - ERA_2.1.13.0: Adds support for enhanced search of two additional data types (Timepiece and ECTS)
  - Approximately 54,000 queries have been made of system since operations began.

Increment 3 Status

- ERA Congressional Records Instance
  - Instance Functionality
    - ✓ Ingest, storage and retrieval for transfers of Congressional electronic records
    - ✓ Deployed to the production environment to leverage existing infrastructure
    - ✓ Provides simple but expandable storage of Congressional Records files shipped to Rocket Center on USB disk drives.
    - ✓ Congressional Records Files are stored on dedicated storage in the Archival Enclave.
    - ✓ Access to Congressional files is provided in the event NARA needs to retrieve Congressional files from Rocket Center. Access Workstations are used in Archival Enclave to copy Congressional files from archival storage onto encrypted removable media (appropriate media) to be shipped back to NARA.
    - ✓ Backup and restore capability for Congressional data.
  - Ready for Operations on 12/14/09
  - First Production Ingest Completed on 3/11/10
    - ✓ Also demonstrated the copy verification process along with testing of integrity of access protections in ERA Archival Storage

Congressional Records Instance



Increment 3 Status

- ERA Public Access Instance
  - ➢Instance Functionality
    - ✓ Search for Archival Research Catalog (ARC) descriptions of holdings and related digital surrogates.
    - ✓ Federate searches to, and access data from selected electronic records sets in the Access to Archival Databases (AAD) system.
    - ✓ A rich public user interface to search, browse and access NARA's archival holdings
    - ✓ Interface to the ARC system for frequent updates of descriptive data in the ERA Public Access Instance.

OPA Beta Environment Overview

- The OPA Beta is deployed in the non-production environment at Rocket Center.
  - ✓ An environment for designated users to evaluate initial capabilities of an eventual publicly-accessible access-only ERA Instance.

- The OPA Beta is comprised of three networked Linux-based servers and a firewall connecting OPA Beta to the CAT Management Enclave
  - ✓ The firewall is used to disconnect OPA Beta from the CAT Lab when CAT testing is in progress.
  - ✓ The OPA Beta Servers are connected to a dedicated burb in the DMZ application firewall, which provides access to users over the DREN and allows remote access to AAD and ARC Web systems from OPA Beta Servers.

- All NARA NET users who know the OPA Beta URL can access the OPA Beta

Increment 3 Status

- ERA Public Access Instance
  - ➢Hardware and Software installation at ABL completed.
    - ➢NARA Testing began March 1
  - ➢Beta final software release deployed 3/31/10.
    - ✓Waiting for system to complete indexing prior to resuming testing.
  - ➢Working connection issues at ABL.

Increment 3 Status

- ERA Preservation Framework Prototypes
  - Functionality
    - ✓ A flexible framework to enable the deployment of various software tools for the purpose of transforming electronic records from one format to another.
    - ✓ Enhanced capabilities to identify formats of ingested electronic records, and to persist their important archival and technical characteristics.
    - ✓ Initial electronic records preservation planning.

Preservation Prototype Demo Schedule

Date	Time	Focus	Status
12/10/09	2PM - 4PM	Search UI	Complete
12/22/09	2PM - 4PM	•EBCDIC to ASCII transformation. •Update ACE after transformation. •Verification and Validation capabilities •Step up to “I3-like” NARA desired ACE.	Complete.
1/7/10	2PM - 4PM	Output Summary Report Validation and Verification UI	Complete
1/21/10	2PM - 4PM	Search filtering out files already transformed. MS-WORD to PDF transformation	Complete
1/21/10	2PM - 4PM	Metadata extraction and automatic validation of significant properties for MS-WORD to PDF.	Complete
2/4/10	2PM - 4PM	MS-Excel to PDF transformation with metadata extraction and automated validation	Complete
~~2/18/10~~ 2/23/10	2PM - 4PM	MS-Excel to ODS transformation with metadata extraction and automated validation	Complete. Rescheduled to 2/23 to accommodate schedule conflict
3/4/10	2PM - 4PM	Completion of transformation framework. Additional ingest related operations	Complete
3/18/10	2PM - 4PM	Agreed upon rework	Complete
4/7/10	2PM - 4PM	Final demo	
4/8/10	9:30AM – 11AM	ACERA Demonstration	

Increment 3 Status

- ERA Base Architecture Evolution
  - ➢Functionality
    - ✓ Adopt XForms standard for Business Objects coupled with industry standard for code lists (Genericode)
      - Allows for lower software update costs, enabler for NARA wide EA data governance
    - ✓ Fully adopt industry standard based Business Process Management language and infrastructure (BPMN and BPEL)
      - Optimizes business workflow development, monitoring and maintenance.
    - ✓ Enhanced file format identification and metadata extraction
      - Positions ERA to ingest a wider range of file formats and better search the ingested assets
    - ✓ More robust ERA catalog design
      - Provides better search on business objects, more extensible for new data formats and asset types that are being discussed for Ingest
    - ✓ Alignment with NARA developed Reference Architecture
      - Supports the planned cost effective evolution of the system

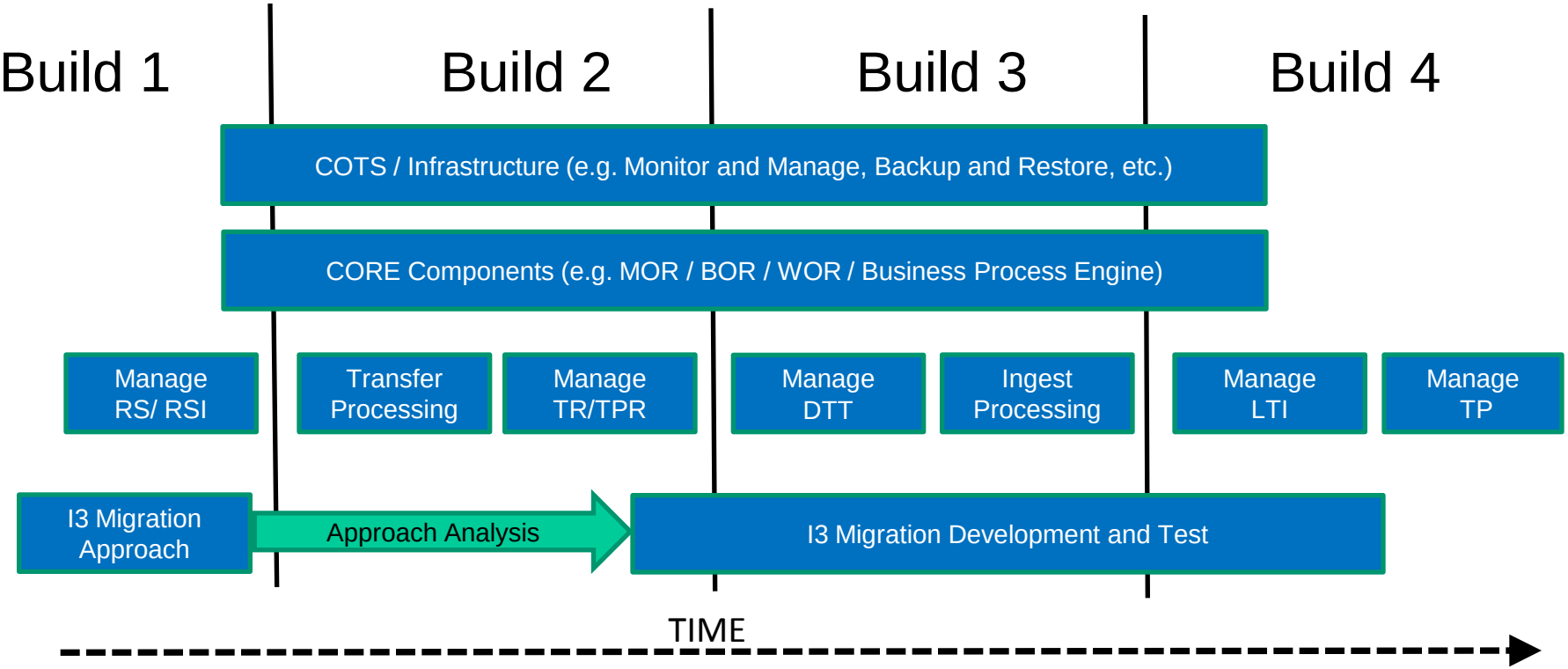
Increment 3 Status

- Base Architecture Evolution (cont):
  - ✓ Increased Software Code Development in three areas has impacted schedule duration
    - XForms Development and Data Model Development
    - Ingest Orchestration
    - Archive Search Capability
  - ✓ LMC working new corrective action plan.
  - ✓ Plan segregates remainder of work into four functionally oriented builds as optimal method for building and testing remaining work

ERA Base Architecture Evolution Development Approach

- Re-evaluated approach and determined segregating remainder of I3 into four functionally oriented builds is the optimal method for building and testing I3
  - This System Integration sequence brings all developed software, COTS and hardware products together in a logical, incremental approach based on functional capabilities
  - Each build provides a stable test platform for the integration and test of the next build
  - Reorganized functions based on the accessioning process lowers rework between requirements and architecture teams
  - Provides for a synchronized inspection package that consists of requirements, architecture and low-level design
- Each build represents a major set of capability that allows for incremental test and integration
  - Provides a better unified set of functions for NARA familiarity testing

ERA Base Architecture Evolution Functional Build-Up Strategy



Segregating remainder of I3 into four functionally oriented builds is the optimal method for building and testing I3

Increment 3 Status

- Base Architecture Evolution (cont):
  - ✓ Build 1 (CORE/Manage RS/RSI)
    - Software development completed on plan (3/26)
    - SWIT is underway and on track for 4/14 completion
  - ✓ Build 2 (Transfer Processing/Manage TP/TPR)
    - High Level Design completed on 3/16
    - Completed Low Level Design (3/19) and conducted Inspections (3/26-3/29)
  - ✓ Build 3 (Ingest Processing / Manage DTT)
    - On target for completion of High Level Design (4/9) and Low Level Design (4/12)
  - ✓ Build 4 (Manage LTI/TP)
    - On target for High Level Design Completion (5/5)

Acronyms

MOR: Metadata Object Repository
BOR: Business Object Repository
WOR: Working Object Repository
RS: Record Schedule
RSI: Record Schedule Item
TR: Transfer Request
TPR: Transfer Plan Repository
DTT: Data Type Template
LTI: Legal Transfer Instrument
TP: Transfer Plan

Near Term Priorities

- ERA Base Support and Point Releases.
- OPA Beta to General Availability.
- Preservation Prototype Support and Requirements Analysis.
- Back up and Restore alternatives analysis.
- EOP Data Type Augmentation.
- Classified System alternatives analysis.

Remainder of the Contract

- Implement Preservation Capabilities
  - ✓Implement Preservation Framework in Production.
  - ✓Develop Preservation capabilities
- Implement and Expand Public Access Capabilities
- Extend ERA Base Capabilities
  - ✓Capability & Support to Process eRecords
  - ✓Subsume legacy systems
- Deployment of ERA classified system