ERA Preservation Prototype

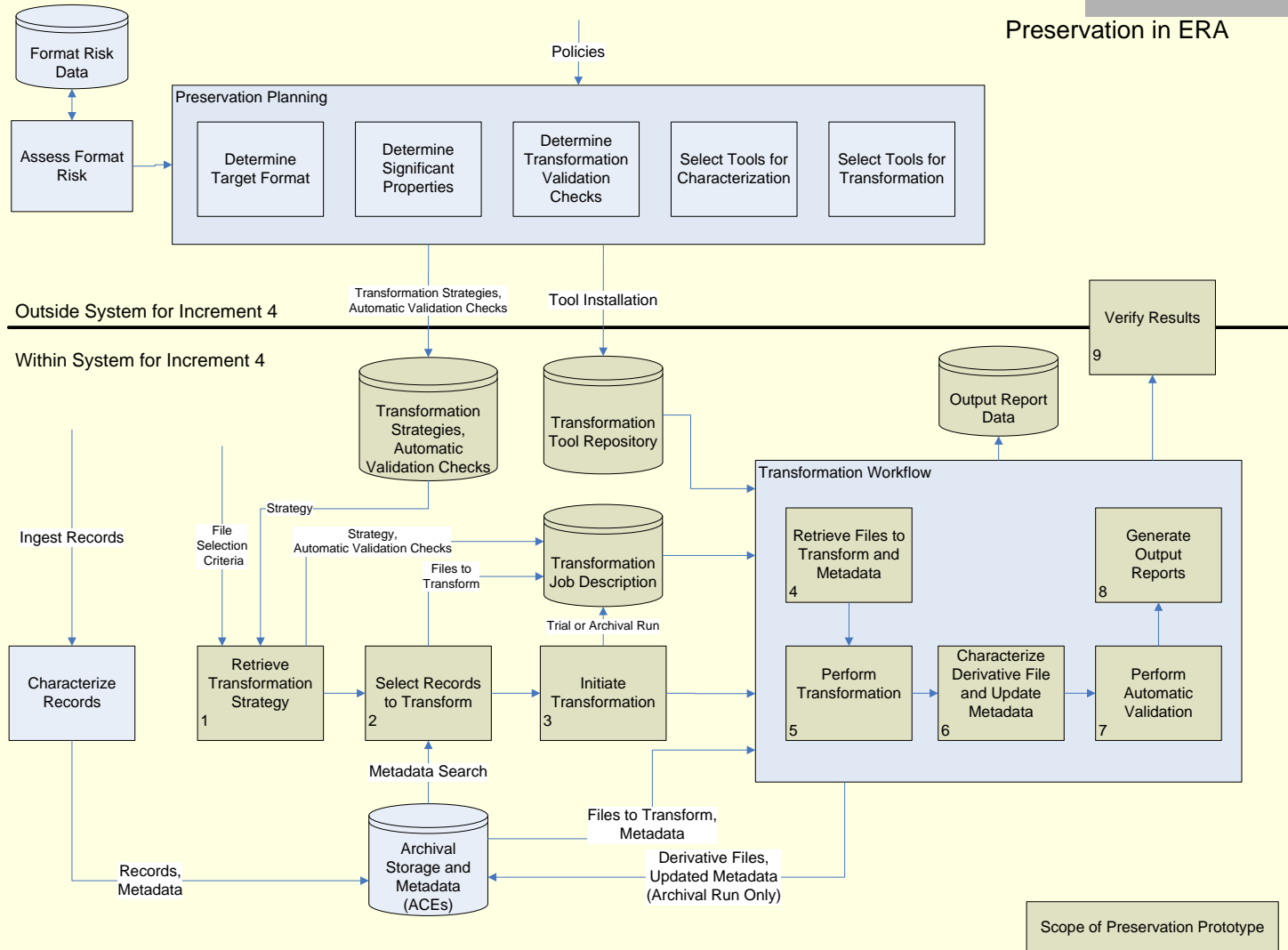Business Analysis Team

April 7, 2010

Prototype Goals

- Create a framework that will enable us to easily swap in and swap out transformation tools and approaches
- The framework will be integrated into one system (ERA) that can ingest, characterize, and transform records – can transform records for access/preservation purposes
- Ensure that the transformations we create are authentic
- Take advantage of the work of other institutions (TNA, Harvard, NZNL, NA of Australia, Planets)

Out of Scope for the Prototype

- We are not trying to determine the optimal strategy for a given digital format

- We are not trying to determine the optimal tool for each transformation

- We are not building the Increment 3 ingest, ACE, search, or file extract services in the prototype – these are being built by the System Evolution set of services

- We are not building in preservation planning or risk assessment

- The prototype will not be scalable – the framework will be

What's in and what's out?

Preservation in ERA

Format Risk Data

Policies

Preservation Planning
- Determine Target Format
- Determine Significant Properties
- Determine Transformation Validation Checks
- Select Tools for Characterization
- Select Tools for Transformation

Assess Format Risk

Transformation Strategies, Automatic Validation Checks

Tool Installation

Verify Results
9

Outside System for Increment 4

Within System for Increment 4

Transformation Strategies, Automatic Validation Checks

Transformation Tool Repository

Output Report Data

Ingest Records

Strategy

File Selection Criteria

Strategy, Automatic Validation Checks

Files to Transform

Transformation Job Description

Trial or Archival Run

Transformation Workflow
- Retrieve Files to Transform and Metadata — 4
- Generate Output Reports — 8
- Perform Transformation — 5
- Characterize Derivative File and Update Metadata — 6
- Perform Automatic Validation — 7

Characterize Records

Retrieve Transformation Strategy
1

Select Records to Transform
2

Initiate Transformation
3

Metadata Search

Records, Metadata

Archival Storage and Metadata (ACEs)

Files to Transform, Metadata

Derivative Files, Updated Metadata (Archival Run Only)

Scope of Preservation Prototype

What will be transformed?

Original Digital Format	Target Digital Format	Digital Record Type	Transformation Tool
EBCDIC	ASCII	Data Files	Version of APS
MS Word	PDF	Textual	JODConverter
MS Excel	ODS	Data files (spreadsheets)	JODConverter
MS Excel	PDF	Data files (spreadsheets)	JODConverter

How will we know we have succeeded?

- Can filter on records that we want to transform – records that do not already have a transformed version
- Can select an appropriate transformation strategy and apply it to any aggregate (series/transfer group)
- Can plug transformation tools into the framework and configure them correctly
- Can transform records to correct target format, based on transformation strategy
- Can conduct a trial run and archival run
- Can verify results of transformation via reports and visually
- Metadata of the derivatives will point to the correct originals and vice versa

Demo

Lesson Learned

- File identification – lack of signatures for some file types (EBCDIC)

- Metadata extraction – some of the things we want to know cannot be extracted with given tools

- Search – how to filter?

- Transformation issues – what constitutes a successful transformation?

- Process – frequent iterations and customer demos seem to have worked well

Next Steps

- NARA needs to hire records preservers to help us evaluate optimal tools and approaches

- Have LMC build the production version of the framework in Increment 4

- Add more formats and tools to the framework in Increments 4 and 5