

# Persistent Digital Archives and Library System



# About Us

---

- ◎ Partnership among seven state archives and libraries
  - Arizona (lead)
  - Alabama
  - Florida
  - New Mexico
  - New York
  - South Carolina
  - Wisconsin
- ◎ Government records and publications

# About us

---

- ◎ Richard Pearce-Moses, Principal Investigator  
Clayton State University  
rpearce-moses@clayton.edu
- ◎ Pete Watters, Principal Investigator  
Arizona State Library  
pwatters@lib.az.us
- ◎ Brian Schnackel, Lead Developer  
Arizona State Library  
bschnackel@lib.az.us

## About Us

---

- Principal funding by the Library of Congress, National Digital Information Infrastructure and Preservation Program (NDIIPP)
- Additional funding from the Institute for Museum and Library Services, Library Services and Technology Act from the Arizona State Library and Archives

# Fundamental Assumption

---

- ◎ Many agencies hold some – *not all* – archival records in an electronic recordkeeping system
  - The records may be digitized or born digital
  - The records have metadata used as access points

# Research Goals

---

## ◉ Develop a curatorial rationale

- Articulate general business rules and workflows for processing large sets of electronic records, publications if possible
- Get some insight into how much rules will vary from series to series
- Test whether archivists' work can be limited to rules written for aggregates, not items; if they can work with rules, not records

# Research Goals

---

- ◎ Test the use of middleware to implement those business rules in software as an integrated workflow to process collections of records and publications
  - Test Microsoft BizTalk in particular

# Research Goals

---

- To evaluate LOCKSS as the basis of a storage network
  - Designed for journals on publisher's website collected by many repositories, not records in recordkeeping systems managed by a single repository
  - Capacity and performance
  - Interagency interoperability



# Research Goals

---

- ◎ Build a community of shared practice that meets the needs of a wide range of repositories
  - For best practices ~ what works, what's practical
  - For resource sharing ~ avoid redundant work
  - For mutual support and development
  - For sustainability
  - Not just for the project, but for the future

# Additional Goals

---

- Remove barriers to preservation by keeping costs as low as possible
- Investigate the use of open source software

# Preliminary Results

---

# Fundamental Assumption

---

- ⦿ Majority of records in agencies' recordkeeping systems are not archival
- ⦿ Metadata is frequently limited
  - Quantity and quality is very uneven
  - Searching based on text strings
  - Not formed as headings used as access points

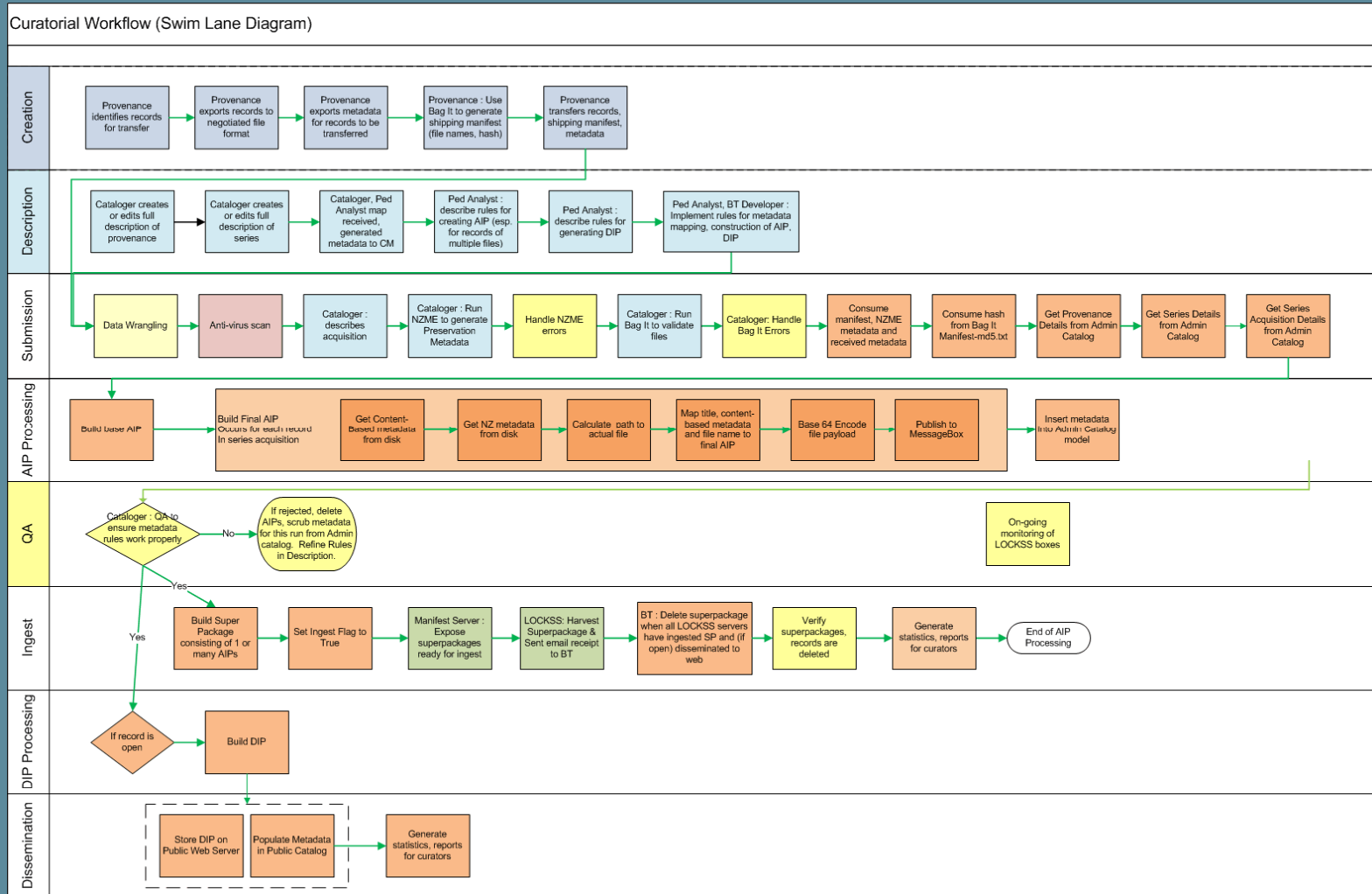
# 1A. Curatorial Rationale

## Generalized Workflow, Business Rules

---

- ⊙ Creation (in the agency of origin)
- ⊙ Description
- ⊙ Submission
  - Validate transfer (Bag It)
  - Data wrangling (Preprocessing to normalize records)
  - Generate preservation metadata (NZME)
  - Accessioning
- ⊙ AIP processing and Quality Control
- ⊙ Ingest into LOCKSS distributed digital preservation network
- ⊙ DIP processing
- ⊙ Dissemination
  - Web catalog
  - Finding aids

# Curatorial Rationale: Swimlane Diagram



## 1B. Curatorial Rationale

### Variation in workflow, business rules

---

- ⦿ Workflow requires minimal tweaks
- ⦿ Some records require specialized preprocessor to normalize the records' format (e.g., PST to XML)
- ⦿ Metadata mapping requires customization, but this is not particularly difficult in BizTalk

## 1C. Curatorial Rationale

### Archivists work with rules, not records

---

- ◎ Some archivists found this difficult; they wanted to work with records and relished the exceptions
- ◎ Rules work when the records and metadata match the model
  - Lack of metadata
  - Problematic metadata
  - Authority control a significant issue



## 2A. Test Middleware

---

- ⊙ Demonstrated that the rules and workflow can be automated
  - Not yet a “soup to nuts” integrated system
- ⊙ Code can be reused
  - For common records series (e.g., PSTs)
  - For similar records (e.g., PDFs w/ metadata in Excel, Access)
- ⊙ Successful normalization , enhancement of metadata
  - Tools help: New Zealand Metadata Editor provides basic preservation information, Bag It provides integrity information

## 2B. Test BizTalk

---

- ◉ Designed to work with many small XML files, not large digital objects
- ◉ Could not serve as an integrated workflow platform appropriate for curators
- ◉ Requires expertise in programming
- ◉ Handles metadata normalization and enhancement very well
- ◉ Coordinates many tasks, with dependency and error handling well

### 3. Test LOCKSS

---

- ⦿ Can work with records
  - Concerns for upper limits of Unix file systems probably not necessary.
  - Concern over capacity: we hope to test a 16TB system.
  - Providing rack space for other repositories complicated by firewalls and paranoia
- ⦿ Multiple, distributed copies in LOCKSS provides more than protection from disaster
  - Protects authenticity of the records

## 4. Collaborative Community

---

- ⊙ Agreed on
  - Metadata standards
  - Hardware and software infrastructure
  - Workflow and business rules
  - Provided for extensibility and flexibility
- ⊙ Ability to share code has enormous potential
- ⊙ Governance is tough
- ⊙ Make sure you have the right partners

# Other Lessons

---

- ◎ Both open source and commercial software have risks
  - Widely adopted software reduces risk
  - Support community reduces risk
  - Sustainability hard to predict
    - Good products disappear without support
    - Commercial products “evolve”

# For more information

---

## ◎ Pete Watters

- pwatters@lib.az.us
- Arizona State Library, Archives & Public Records

## ◎ Richard Pearce-Moses

- rpearcemoses@clayton.edu
- Clayton State University