


# PeDALS

## Persistent Digital Archives & Library System



Richard Pearce-Moses

Deputy Director for Technology & Information  
Resources

Arizona State Library, Archives and Public Records

# A Word from Our Sponsors

---

- Library of Congress  
National Digital Information and  
Infrastructure Preservation Program  
(NDIIPP)
  
- Institute for Museum and Library Services

# Project Partners

---

- Arizona
  - Florida
  - New York
  - South Carolina
  - Wisconsin
- 
- Kudos to the Washington State Archives

# Technical Goals

---

- ❑ To develop a curatorial rationale that can be implemented in software to support an automated, integrated workflow to process collections of digital publications and images
- ❑ To build “digital stacks” using LOCKSS as the basis of an inexpensive storage network that can preserve the authenticity and integrity of the materials.

# Additional Goals

---

- To build a community of shared practice that meets the needs of a wide range of repositories
  - For best practices
  - For resource sharing
- To remove cost barriers by keeping costs as low as possible

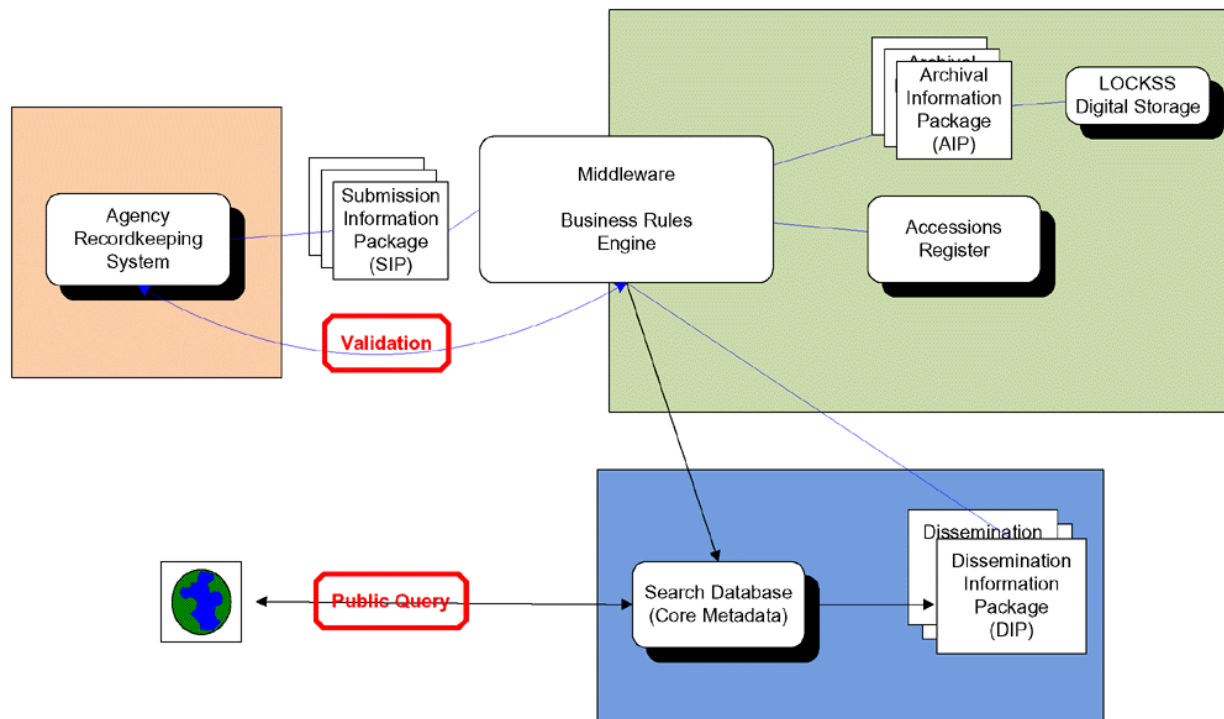
# Curatorial Rationale

---

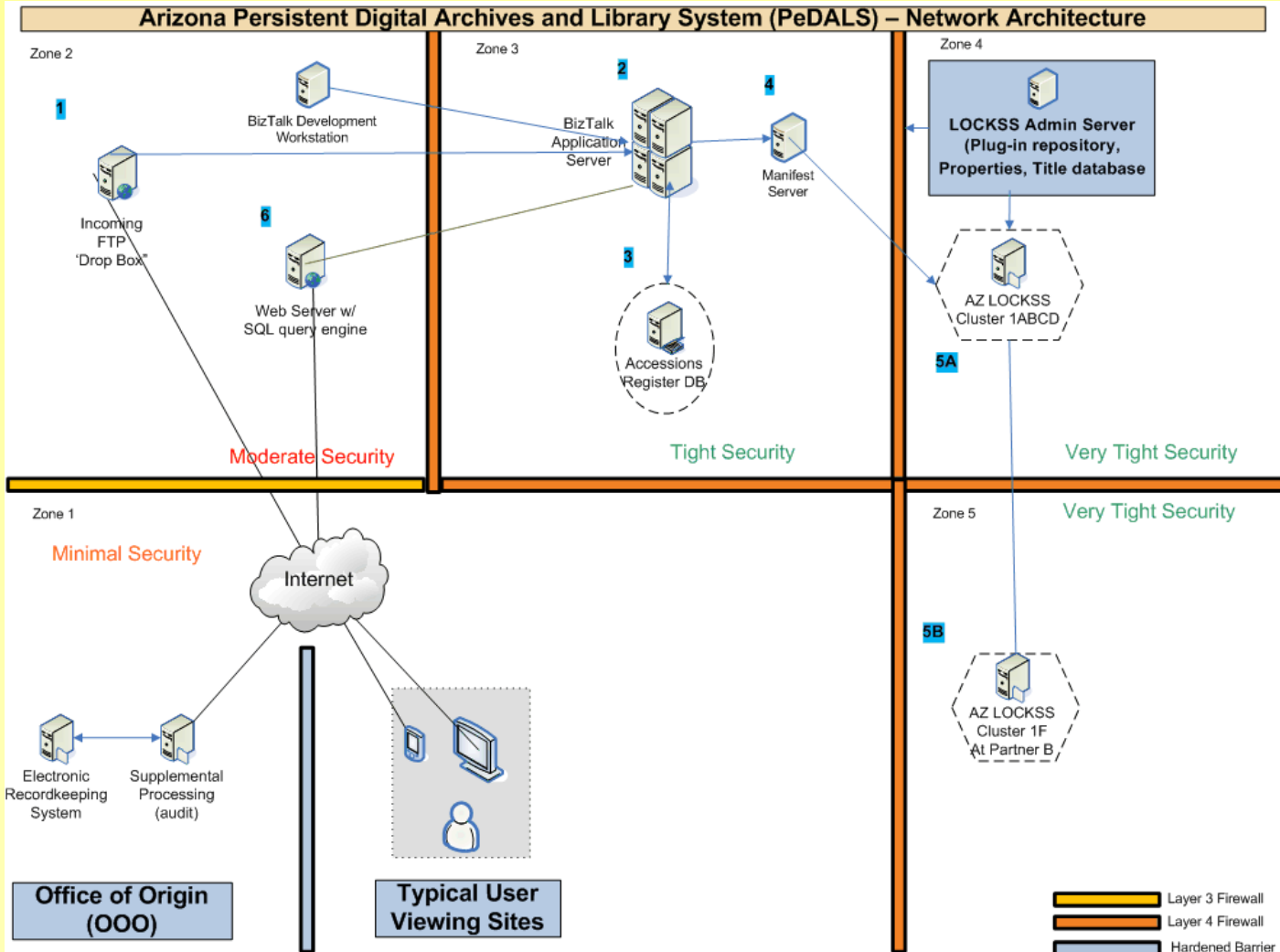
- Transformation of traditional, paper-based practices into the digital arena
  - Appraisal
  - Acquisition
  - Arrangement and description
  - Housing and storage
  - Reference and access
  - Preservation

# Data Flow

Arizona Persistent Digital Archives and Library System (PeDALS)  
OAIS Information Flow



# Architecture



# Automated Processing

---

- Curators work with rules, not records
  - Describe business processes (rules)
  - Monitor the process for quality assurance
- Rules expressed in software
  - A “pipeline” that transforms records as they move through the system
- Based on Microsoft BizTalk middleware

# 1. Preparatory Work with OOO

---

- For each series of records selected for transfer
  - Negotiate metadata you will receive
  - Negotiate format of the records
  - Negotiate format of the submission information package
  - Negotiate frequency and manner of transfer
  
- OOO develops procedures to create SIPs
  
- Archives describes business rules in middleware

## 2. Ingest: Receive SIPs From OOO

---

- Transfer to a drop box in DMZ
  - FTP
  - Tape
  - CD
- Each OOO has a directory
  - Each series has a subdirectory
- Isolated for virus scanning
- Option: simple Linux box

# Submission Information Packages

---

## □ OOO Metadata

- "Well number" , "Owner" , "Title" , "File name"
- "56-000001","CITY OF TUCSON","2003 annual report","56 files\56-000001\_0000.pdf"
- "56-000001","CITY OF TUCSON","2004 annual report","56 files\56-000001\_0000\_E52B0.pdf"
- "56-000001","CITY OF TUCSON","2005 annual report","56 files\56-000001\_0000\_E8578.pdf"
- "56-000001","CITY OF TUCSON","2006 annual report","56 files\56-000001\_0000\_EC3F8.pdf"

## □ Records

- XML
- PDF
- Other formats

# 3. Execute Business Rules

---

- Validation
  - Were all records received without corruption?
  - Were any false records received?
- Generate administrative metadata
  - Assign accession number, date
  - Assign unique system number
  - Record source, transfer authority, restrictions
- Generate series-level metadata
  - OOO: Creator, Provenance, Source of Records
  - Series title, Date ranges
  - Scope note
  - Access points: activities, topics

# 3. Execute Business Rules – con't

---

- Generate item-level discovery metadata
  - Creator, Provenance
  - Title
  - Extent: How much of what?
  - Description: abstract, first1024
  - Access points
    - Parties to the record: individuals and roles
    - Dates: Execution, filing
    - Subjects: location, keywords, form/genre

## 3. Execute Business Rules – con't

---

- Generate Preservation Metadata
  - Fixity: MD5 or SHA hash value
  - File format: JHOVE or Pronom
  - Rendering software
  - Digital signature
  - Operating system

# Mapping and Creating Metadata

---

<b>Metadata Element</b>	<b>Received</b>	<b>PeDALS Core</b>
Office of Origin		Arizona. Dept of Water Resources : AACR
Var Names		Water Resources : keyform
Series Title		Annual Reports
Series Description/Scope note		[Narrative text]
Arrangement		[Narrative text]
Series Date Range		2003 – 2006
Series Subjects		Water wells Water Supply
Activity/Function		Regulation Health and safety
Transfer authority		Retention schedule 89-403
Restrictions		Open

# Mapping and Creating Metadata

---

Metadata Element	Received	PeDALS Core
Title	2003 Annual Report	2003 Annual Report
Date		2003
Party to Record	City of Tucson	City of Tucson: Author
Location		Tucson
Subjects		Water wells Water supply
Description		[autogenerated]
First1024		[autogenerated]
OOO Id	56-00001_0000	56-00001_0000
File format		PDF
Fixity		[autogenerated]

# Final Processing of AIP

---

- Update Accessions Register database
- Create AIP
  - <AIP>
    - <Hash> </Hash>
    - <Metadata> </Metadata>
    - <Record> </Record>
  - </AIP>
- Deposit in Digital Stacks (LOCKSS)
  - Generate manifest list to expose to LOCKSS
  - LOCKSS harvests from manifest server

# LOCKSS: Digital Stacks

---

- ❑ Redundant Array of Inexpensive Servers
- ❑ Automatic integrity checking
- ❑ Automate error-correction
- ❑ Geographically dispersed copies
- ❑ Bitstream preservation

LOCKSS.ORG

Victoria Reich

Stanford University

# Dissemination Information Packages

---

- ❑ Middleware creates DIPs, puts on public webserver
  - Doesn't include administrative, preservation metadata that users will not likely want
  - In format easily supported by common browsers
- ❑ Middleware updates public webserver search database

# For more information

---

- <http://rpm.lib.az.us/PeDALS/>
  
- Principal Investigator
  - Richard Pearce-Moses  
[rpm@lib.az.us](mailto:rpm@lib.az.us)
  
- Project Coordinator
  - Sara Muth  
[smuth@lib.az.us](mailto:smuth@lib.az.us)