



# Overview of the Presentation

---

- Goals for the system
- Records likely to work well with the system
- Overview of the workflow
- Overview of the hardware and software
- Practicalities: Joining the Project

# Goals

---

- LC/NDIIPP-funded research project
- Use middleware to automate processing of large volumes of records
  - Follows OAIS reference model
  - Can be audited using RLG/OCLC
- Test LOCKSS for as storage system
  - Automated integrity checking and error correction
- Articulate a curatorial rationale that serves as the basis for a community of shared practice
- Work as inexpensively as possible

# Keys for Success

---

- Records that grow out of a regular process
- Records have associated metadata

§

- Likely: Certificates, licenses, Email
- Less likely: Individual documents in file systems
- Possible: Publications, web documents

# Workflow 1. Creation

---

- Creation
  - Records created or received
  - Create metadata for internal access
  - Create Submission Information Package (SIP)
    - Records
    - Metadata
  - Reality: We'll take what we can get
  
- Negotiation key
  - Format and frequency of records and metadata

# Workflow 2. Description

---

- Metadata objectives and means
- To support administration
  - Accession
  - Chain of custody
- To support preservation
  - Track file formats that need migration
  - Demonstrate authenticity

# Workflow 2. Description

---

- Traditional description (APPM/DACS, AACR/RDA)
  - Provenance-level entry in catalog
  - Series-level entry in catalog
  - Acquisition-level entry
- Write business rules for item-level metadata
  - Each series has different business rules
  - Business rules can often be reused (email, similar SIPs)
  - Map received metadata
  - Generate preservation metadata
  - Supply additional metadata

# Metadata Normalization – Examples

---

- Title
  - MARY ALICE CICERALE and ROBERT PORTER : marriage certificate, 2008
- Extent
  - 1 Adobe Acrobat PDF (34379 bytes)
- Date
  - 3 Jan 2008 : Marriage date
  - 12 Jan 2008 : Recording date
- Access
  - CICERALE, MARY ALICE : Bride
  - PORTER, ROBERT : Groom

# Workflow 3. AIP Processing

---

- Middleware runs business rules
  - Normalizes metadata using business rules
  - Adds item-level metadata to administrative catalog
  - Creates Archival Information Packages (AIP)

# Workflow 4. Quality Control

---

- Curator checks to see that business rules ran correctly
- Curator checks error reports
  - Hash values don't match
  - Missing files
  - Extra files

# Workflow 5. Ingest

---

- Middleware

- Packages AIPs for long-term storage
- Exposes the records for access into LOCKSS
- Generates statistics and administrative reports

- LOCKSS

- Harvests records for storage
- Signals middleware that ingest is complete
- Routinely checks for integrity and corrects problems

# Workflow 6. Dissemination

---

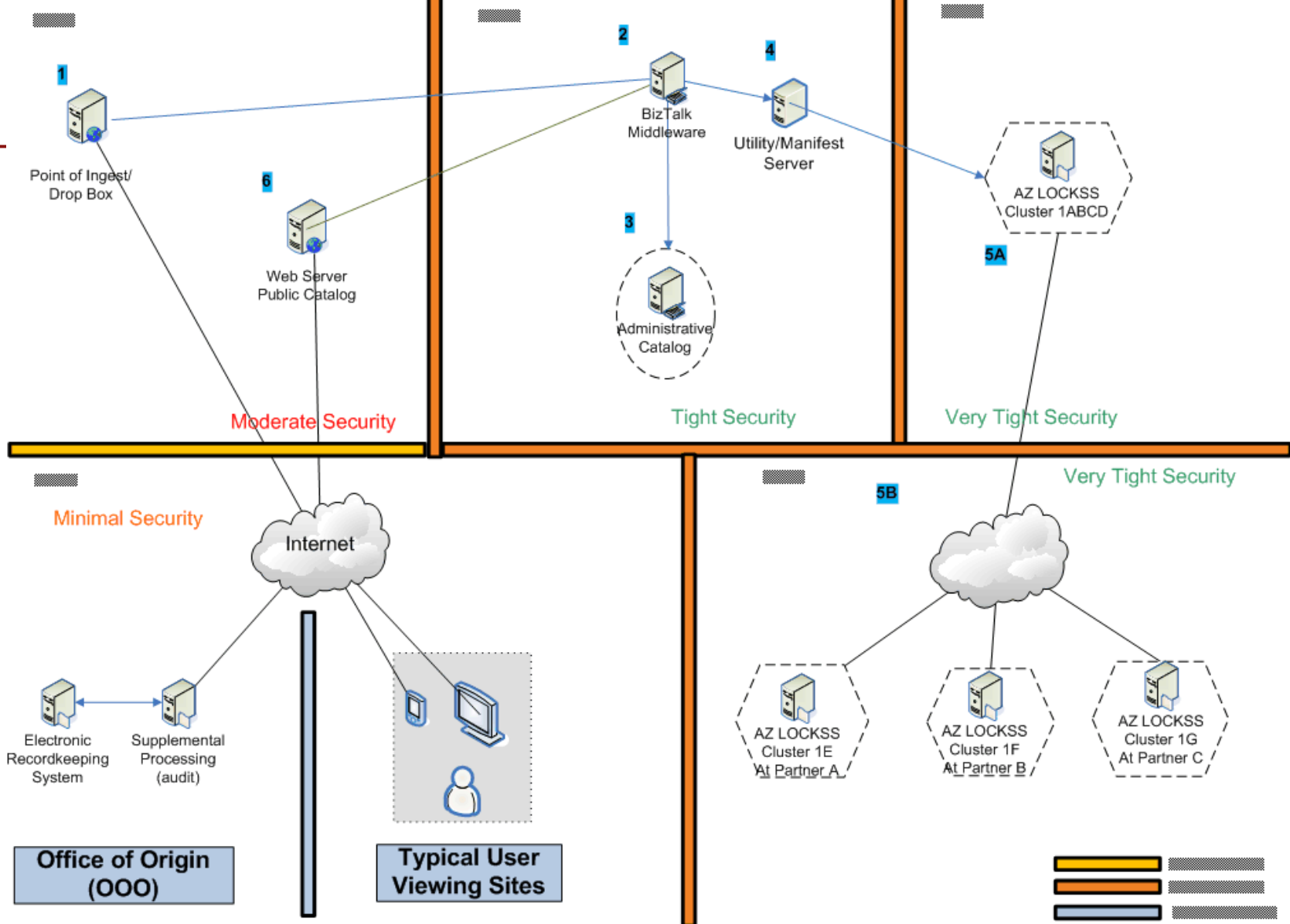
- For records are open to the public
  - Project can include confidential records
  - Partners should focus on accessible records
  - Challenge of personally identifying information
- Middleware
  - Creates Dissemination Information Package (DIP)
    - Smaller
    - Browser friendly
  - Middleware updates public catalog

# Hardware and Software

---

- Microsoft BizTalk middleware
- MS SQL Server for admin and public catalogs
- Windows Server
- Microsoft IIS webserver
  
- LOCKSS for archival storage system
  - Seven boxes, at least two in other states
  - 4TB storage capacity, expandable to 16TB
  - One Ubuntu workstation (virtual option) for plugins

# Arizona Persistent Digital Archives and Library System (PeDALS) – Network Architecture



# Project Status

---

- Core Metadata v. 1 completed
  - First iterative review in progress
- Admin Catalog database designed and implemented
- Admin Catalog web interface nearing completion
- Simple AIP schema completed
- Proof of concept
  - Marriage certificates ingested
  - Marriage certificates code reused for Commission orders

# Next Steps

---

- Begin acquisition, ingest
  - Iterative review with each ingest
  - More complex record types
- Public catalog

# Being a Partner

---

- Project covers hardware and software costs
- Partner contributes staff time
  - Curators: archivists, librarian – 20% each
  - IT: PeDALS admin – knowledge of programming
  - IT: network, system administrator – as needed
  - \$25,000 in staff support from grant
- Memorandum of agreement

# Content

---

- Permanent collection
  - Not a records management system
- Open to the public
  - Implications of web access
- Nine to twelve series
  - New acquisitions (preferred) or existing collections
- Generated out of a regular process, with metadata
  - Looking for ways to work with other materials
- May be digitized; cost of digitization not covered

# Content

---

- Legislatures
  - Bills and amendments
  - Enacted and codified laws
  - Reports from legislative agencies and bodies
  - Journals
  - Floor proceedings
  - Procedural rules

# Content

---

- Courts
  - Decisions and opinions
  - Reports from judicial agencies and bodies
  - Civil and criminal rules or procedures
  - Recordings of trial and appellate court proceedings

# Content

---

- Executive/regulatory agencies or entities
  - Reports and hearings
  - Regulatory policies and decisions
  - Rulemaking documentation
  - Agency policy and operations manuals

# Content

---

## Special Subjects

- Directories of government-wide structure, personnel, and responsibilities
- Historical state tax forms and instructions, along with descriptions (i.e., rates and brackets) of each state tax
- Documentation (such as memoranda of understanding) of regional government consortiums, mutual aid agreements, and other interstate arrangements, including Councils of Government
- State plans for the Temporary Assistance for Needy Families (TANF) program
- State adjutants-general documentation relating to National Guard and State Guard units and activities, particularly with regard to military deployments and civilian disaster assistance

# Staffing

---

- Local teams
  - Archivist, Librarian: negotiate to acquire content, describe acquisitions, map metadata, quality control
  - PeDALS/BizTalk Admin: implement rules, help develop code
  - IT: systems and network support
  - Partner lead: coordination, reporting
- Curatorial Committee
  - Metadata mapping
  - Iterative review of Core Metadata standard
- Tech Committee
  - Developing catalog
  - Implementing business rules

# Meetings

---

- Bi-weekly project conference calls
- Committee conference calls as needed
  
- Annual project meeting (2-3 reps)
- Annual LC/NDIIPP partners meeting (1-2 reps)
  
- End-of-project symposium (2011)
  
- BizTalk training

# Timeline

---

- MOA asap
- Hardware and software
  - Purchase asap, install within six weeks
- Begin identifying content asap
- Begin prep for first ingest
- First ingest in fall
- Partner meeting: Fall 2010
- End of project: Dec 2010