A Framework for Applying OAIS to Distributed Digital Preservation

Spring CNI 2011 Project Briefing

Martin Halbert, UNT Dean of Libraries
Matt Schultz, Educoopia institute
CNI Spring 2011, San Diego, California
Overview

- Background
  - The Need for a Framework
- Justifying the Need for a Framework
  - A MetaArchive Case Study
- Enlarging the conversation
  - Helpful Follow-up Questions
- Next Steps
  - Proposing a Mutual Set of Work
Background

The Need for a Framework
What is this all about?

- OAIS Reference Model now widely accepted and driving repository development
  - DAITSS
  - Archivematica
- Derived auditing metrics (TRAC) becoming a benchmark for “trustworthy” repositories
- Digital preservation increasingly being accomplished through distributed approaches
Distributed Digital Preservation (DDP) Initiatives

- LOCKSS-based
  - LOCKSS, CLOCKSS, MetaArchive, DataPASS, PeDALS, COPPUL, LOCKSS-KOPAL, Synergies, ADPNet
- iRODS-based
  - Chronopolis, NARA-TPAP, SHAMAN
- DuraCloud
- NetArchiveSuite
A Framework for Applying OAIS

- Supplemental document to OAIS
  - Enlarge the community’s understanding of distributed digital preservation concepts & approaches
  - Guidance on trustworthy digital preservation activities for DDP developers & practitioners
  - Effectiveness for auditing DDP preservation solutions with metrics like TRAC
Justifying the Need for a Framework

A MetaArchive Case Study
MetaArchive Cooperative

- Established in 2004 (support from NDIIPP and NHPRC), preserving content for 18 members
- Uses LOCKSS software to provide peer-to-peer distributed digital preservation infrastructure
- Sustainable organizational framework: Membership organization with a 501c3 host (Educopia)
- 286 TB network capacity (and growing)
MetaArchive & TRAC Self Audit

- Completed self-audit with external auditor in 2010 (see http://metaarchive.org/resources)
- MetaArchive successfully conforms in all 3 categories and 84 criteria
- “trustworthy digital repository”...ensures that processes, policies and workflows meet the standard for long-term preservation
Use Cases & Gap Analysis

- Use Cases (experiences applying OAIS)
  - Open Pull vs. Push on Ingest
  - Dark Archiving, Full Preservation & Designated Communities

- Gap Analysis
  - OAIS Section 6 & MetaArchive
Open Pull vs. Push on Ingest

- Dominant model
  - Producer → “pushes” → SIP → Archive
- MetaArchive-LOCKSS model
  - Archive ← “pulls” ← SIP ← Producer
- MetaArchive: The formation of an AIP is primarily the responsibility of the Producer
  - Data Wrangling & Plugin Development
- Expanding on roles & responsibilities
  - Overlap between Producers, Consumers & Mgmt
Dark Archiving, Full Preservation & Designated Communities

- OAIS *primarily* assumes that most archives will build active dissemination pathways (AIPs → DIPs)
- MetaArchive is a “dark archive” that limits access
  - Focus is on Content vs. End User Experience
- Producers (also Consumers) assume responsibility for “Full Preservation”
  - Transforming the notion of a Designated Community?
- Voluntary association that carefully weighs its preservation activities in concert
  - Involving and engaging libraries, archives, and cultural memory organizations in the process
OAIS Section 6 & MetaArchive

- Responding to OAIS Section 6: Technical Levels of Interaction Between OAIS Archives
  - Independent Archives
  - Cooperating Archives
  - Federated Archives
  - Archives with Shared Functional Areas

- Network Driven Differences?
  - Participatory; Peer-to-Peer
Central & Distributed Preservation
Independent Archives

- Independent Archives
  - Serving a single Designated Community
  - Design decisions and standards NOT geared toward interoperability
  - May be distributed across many sites
    - Archival Management still distinct from Producers and Consumers
  - MetaArchive is distributed with Archival Management located at Producer/Consumer ends
Central & Distributed Preservation Cooperating Archives

Figure 6-1: Cooperating Archives with Mutual Exchange Agreement

Figure 6-2: Cooperating Archives with Standard Ingest and Access Methods
Cooperating Archives

- MULTIPLE archives serving SIMILAR Designated Communities
- Design decisions and standards ARE geared toward interoperability (DIP becomes a SIP)
- Archives order and ingest data from one another
  - But each separately responsible for its own archival management
- MetaArchive has shared standards and exchanges data but still mutually responsible
Central & Distributed Preservation Federated Archives

Figure 6-3: An OAIS Federation Employing a Common Catalog
Central & Distributed Preservation
Federated Archives

- Federated Archives
  - MULTIPLE archives serving SIMILAR Designated Community(ies)
    - Local and Global
  - Shared finding aids but local access
  - Problem of multiple copies but only one original
    - Extra effort needed upon access to deliver the original
  - MetaArchive makes sure that each “archive” has copies as authoritative as the original
Central & Distributed Preservation Archives with Shared Functions

Figure 6-4: Archives with Shared Storage
Central & Distributed Preservation Archives with Shared Functions

- Archives with Shared Functional Areas
  - MULTIPLE archives serving DIFFERENT Designated Communities
  - Shared resources (Ex. Common Storage) and standards (Ex. Common Ingest and Access)
  - Archive may also be Cooperative and Federated
    - Archival Management still distinct from Producers and Consumers
  - MetaArchive is a SINGLE archive with shared resources and standards
Outcomes & Directions

- OAIS & TRAC helped us identify where we could improve our policies and documentation
- TRAC audit also showed us the need for articulating DDP in OAIS terms:
  - Bridging terminology
  - Expanding on functional areas
  - Describing roles and responsibilities
- Not another Reference Model
Summary Points

- Expanding roles & responsibilities
  - Overlap between Producers, Consumers & Mgmt
- Flexibly applying definitions & concepts
  - Full Preservation & Designated Communities
- Elaborating on Archive Interoperability
  - Independent Archive that absorb qualities of multiple interoperable archives Vs. Independent Archives being interoperable
Enlarging the Conversation

Helpful Follow-Up Questions
Characteristics of Trustworthy DDP

- Organizational distribution
- Geographic distribution
- Replication of content
- Technical redundancy & overlap
- Other?
With whom are agreements made?
What happens if a replication site drops out?
Tracking who has curatorial responsibility (is it transferred? To the network or to individual repositories)?
What data management is handled by the Producer and what is handled by the Repository?
When needed, which copy becomes the most appropriate Dissemination Information Package (DIP)?
DDP and Technical Questions

- Repository relationships
  - P2P, hub and spoke, data grid, other
- Security
  - How to assess the security of each repository? Of the network?
- Preservation metadata
  - Where has a copy lived? Where are others? Are all equal?
- Copies, copies, copies
  - How many copies are enough? What if they don’t match? What if content changes?
Next Steps

Proposing a Mutual Set of Work
Setting the Stage

- Constellation Group
  - Educopia, Library of Congress & PLNs
- Core Committee
  - Nancy McGovern (ICPSR)
  - Martin Halbert (UNT)
  - Richard Pearce-Moses (Clayton State)
  - Tyler Walters (Virginia Tech)
  - Matt Schultz (Educopia)
Statement of Purpose

The Educopia Institute and the National Digital Stewardship Alliance seek to convene a working group with the broader digital preservation community to develop an approach for implementing the OAIS Reference Model and the derived Trustworthy Repository Auditing & Certification metrics (TRAC) in a distributed digital preservation (DDP) environment.
Focused Objectives

- **Describing a Framework for Applying OAIS to DDP**
  - Describe the roles and responsibilities at institutions which are working to replicate and preserve digital information in an organizationally distributed environment;
  - Provide a common vocabulary and set of recommendations for trustworthy interoperable systems development; and
  - Inform administrative and technical practices between groups that seek to collaboratively preserve digital information using distributed mechanisms.
Activities Proposed

- 1-2 Year Project – Three Phases
  - Research & Recruitment: Gap Analysis White Paper (OAIS Section 6 & Various Use Cases)
  - Production: Collaborative drafting of Framework (modeled on PAIMAS)
  - Dissemination: Submission to CCSDS and promotion in the DP community
Gap Analysis White Paper

- NDSA Wiki
  - Gap Analysis – OAIS Section 6
    - Model alignments & divergences
  - Use Cases (experiences applying OAIS)
    - Document specific distributed digital preservation elements that need to be accounted for
- White Paper
  - Summarize findings and form recommendations
- Virtual Meetings
Drafting a Framework

- **Process**
  - Build on White Paper recommendations
  - 1 or 2 virtual and physical meet-ups
  - Model on CCSDS Producer Archive Interface Methodology Abstract Standard (PAIMAS)
    - Bridging terminology
    - Expanding functional areas
    - Describing roles & responsibilities
About PAIMAS

- **Producer Archive Interface Methodology Abstract Standard** (http://public.ccsds.org/publications/archive/651x0m1.pdf)
  - Purpose is to identify, define, and provide structure to the relationships and interactions between an information Producer and an Archive.
  - Elaborates on the SIP to AIP phases (preliminary, formal def, transfer, validation)
  - Provides guidelines & recommendations for creating a sub-standard of the Methodology
  - Points back to OAIS
Disseminating the Framework

- Comment period
  - Solicit on major digital preservation listservs
- Submission of Framework to CCSDS
  - Supplemental document to OAIS
- Launch event with all groups involved
Practical Next Steps

- Meet with interested groups at CNI
  - Feedback & criticisms
  - Expand on core committee
  - Exchange contacts w/Matt (matt.schultz@metaarchive.org)
- Visit the NDSA Wiki – Standards & Infrastructure
  - Provide use cases & gap analysis
- Watch the NDSA listserv for virtual meet-ups prior for White Paper drafting
Feedback & Critique

Thoughts? Questions?