Curation and Preservation Services: Adapting Frameworks and Tools to Enable Sustainable Programs

Nancy Y McGovern, TCDL, May 2013
Topics

- Community Landscape
- Building Durable Programs
- Ensuring Durable Skills
- Demonstrating Good Practice
Community Landscape
Digital Preservation

“the active management of digital content over time to ensure ongoing access” (NDIIPP*)

- Encourage quality creation by producers
- Document actions taken over the life of digital objects
- Ensure access over time
  - handshakes across generations of technology
  - proven technologies for preservation to contemporary for access

* National Digital Information Infrastructure and Preservation Program
  Library of Congress
Data Curation

“active and on-going management of data through its life cycle of interest and usefulness to scholarship, science, and education” enables discovery, ensures quality, adds value, and provide for re-use over time [UIUC]

• Predates the digital community
• Value-added steps by curators to enhance utility
• Intersection of data science (curators) and research (producers and consumers)
Digital Curation

“maintaining and adding value to a trusted body of digital information for future and current use”

- active management and appraisal over entire life cycle
- builds upon underlying concepts of digital preservation
- emphasizes opportunities for adding value through annotation and continuing resource management
- Preservation is a curation activity - both are concerned with managing digital resources with no significant (or only controlled) changes over time

Source: JISC
Digital Preservation + Data Curation

Digital Curation

Digital Curation Centre definition, circa 2004
Community Context

Curation and Preservation are ongoing not new issues to manage

- **1960s**: national archives, data archives
- **1970s**: increasing interest and concern
- **1980s**: digitization developments
- **1990s**: library, museum, Web collections
- **2000s**: digital art, geospatial, e-science…
- **2010s**: research data, analog archives…

variations by nation, domain, size, complexity…
Preserving Digital Information (PDI), 1996

Commission on Preservation and Access & RLG
Standards and Practice

- **TDR**: Trusted Digital Repositories, 2002
- **OAIS**: Open Archival Information System Reference Model (ISO 14721), 2003 update approved in 2012
- **PAIMAS**: Producer Archive Interface Method Abstract Standard (ISO 20652), 2006 plus update
- **NISO Building Good Digital Collections**, v3.0 2007
- **PREMIS**: Preservation Metadata Implementation Strategies, 2005 plus updates
- **BRTF**: Blue Ribbon Task Force on Sustainable Preservation and Access, 2010
- **TRAC**: Trustworthy Repositories Audit and Certification, 2007 and ISO 16363: 2012
Trusted Digital Repositories

Characteristics of a TDR:

- OAIS Compliance
- Administrative Responsibility
- Organizational Viability
- Financial Sustainability
- Technological and Procedural Suitability
- System Security
- Procedural Accountability
Core OAIS Requirements

- Negotiate for and accept appropriate information from Producers
- Obtain sufficient control of information for Long-term Preservation
- Determine Designated Community
- Ensure information is Independently Understandable to the Designated Community
- Follow documented policies/procedures for preservation against reasonable contingencies
- Make information available to Designated Community ... with evidence for Authenticity
Producer Archive Interface Method
Abstract Standard (PAIMAS)
OAIS standard since 2005

Phase objective
- Define the information to be archived
- Develop agreement (data schedule, complementary elements)
- Actual transfer of the objects
- Validate the transferred objects

Phases:
- Preliminary Phase
  - Preliminary Agreement
- Formal Definition Phase
  - Dictionary Formal model Submission Agreement
- Transfer Phase
- Validation Phase
  - Transferred object files
  - Anomalies
  - Validation agreement

AIP Creation
TRAC

Audit and Certification of Trustworthy Digital Repositories (ISO 16363:2012)

Builds on:
Sustainable Access

Effective and sustainable DP programs address:

- Value – understand and stress content value
- Roles – identify stakeholders and involve them
- Incentives – identify “carrots” for preserving

Identify and address costs across life cycle

Building Durable Programs
Addressing Long-term Access

Preservation makes long-term access possible...

**Preservation** vs. **Access**

- proven ➔ technologies ➔ cutting edge
- accumulate ➔ metadata ➔ relevant now
- access over time ➔ purpose ➔ access now
- future users ➔ focus ➔ current users
Holistic Management

An effective approach addresses:

- Organizational requirements and objectives (what?)
- Technological opportunities and change (how?)
- Resources – funding, staff, equipment, etc. (how much?)

Digital Preservation Management Workshop: dpworkshop.org
Organizational

High-level organizational policies
reflect the intentions of the organization

Lower-level organizational policies
document the decisions of the organization

Individual policy statements
regulate the actions of the organization

Encoded policy statements
translate organization’s policies into actions

Technological
Role of Policies

Benefits of policy development:

- Specifies institutional commitment
- Developing policy builds DP team
- Demonstrates compliance – meet requirements
- Manages expectations – message to stakeholders
- Identifies issues and challenges
- Raises awareness
- Defines roles and responsibilities
Policy Development

who (producers, consumers, curators, managers, auditors) can do what (actions specific to a life cycle stage)
when (at what stage of the life cycle)

In what circumstances (rules derived from policy decisions) – past, present, and future

Types of life cycle activities:
Real time – collection/object (e.g., processing, delivery)
Over time – repository (e.g., preservation planning, audit)
Planning

• Preservation Planning (ongoing)
• Self-assessment (internal process)
• Audit (external review by peers)

Also

• Business Continuity (Protect)
• Disaster Planning (Protect)
From NIST Contingency Planning Guide for Information Technology Systems
Building Sustainable Programs

- Enabled by standards-based frameworks
- Demonstrate good practice
- Document decisions and resulting actions
- Devise cost-effective strategies
- Maintain relevant skills
- Monitor changing technological landscape
- Respond to evolving requirements
- Contribute to community efforts
Ensuring Durable Skills
Digital Curation Capabilities

- Devise strategies
- Develop policies
- Collaborate
- Raise awareness
- Define good practice
- Develop programs
- Address legal issues
- Investigate problems
- Develop workflows
- Design object packages
- Identify dependencies
- Enable interoperability
- Develop competencies
- Build/maintain registries
- Balance risks and costs
- Monitor technology
- Invest in solutions
- Manage repositories
- Promulgate standards
- Manage metadata
Developing and Maintaining Skills

DPM Program Stages
1. Acknowledge
2. Act
3. Consolidate
4. Institutionalize
5. Externalize

Skills Development
1. Interest
2. Self-study or Course(s)
3. Credential
4. Specialization
5. Instruction / Mentoring
Levels and Skills

Different roles need different skills

<table>
<thead>
<tr>
<th>Role</th>
<th>Organizational</th>
<th>Technological</th>
</tr>
</thead>
<tbody>
<tr>
<td>Executive</td>
<td>Fund</td>
<td>Invest</td>
</tr>
<tr>
<td>Managerial</td>
<td>Plan</td>
<td>Select/Administer</td>
</tr>
<tr>
<td>Operational</td>
<td>Use</td>
<td>Coordinate/Build</td>
</tr>
</tbody>
</table>
Digital Curator Vocational Education Europe

- Area 1: Knowledge and principles
- Area 2: Skills and competences
- Area 3: Audience/profile types
- Area 4: Part of digital curation lifecycle
- Area 5: Teaching methods/training delivery
- Area 6: Professional context
Managing Skills

- Perspectives: organizations, teams, individuals
- Range: generalist to specialist
- Stage: early, mid and later career
- Evolution: technologies, requirements, skills

Tools needed
- Organizations: skills bank
- Teams: roles manager
- Individuals: career portfolio
Defining Positions and Roles

- Same job title + different job description = confusion for employers and employees
- Solution: define competencies, formalize
- Reporting lines - what level position?
- Factors: experience, development, costs
- Balance of organizational and technical
- Required vs. Desired skills – which degrees?
- Communication skills
Metadata Skills Example

We have:

- Metadata specialists
- Preservation specialists

What might a preservation metadata specialist look like?
NOTE: the Dream Team refers to the ‘67 Red Sox - of course
Maintaining Durable Skills

- Anticipate change (flexible)
- Assess technical capabilities (aware)
- Track relevant technologies (current)
- Balance monitoring and doing (adaptive)
- Make informed decisions (prudent)
- Invest in technologies (savvy)
- Collaborate on solutions (innovative)
Demonstrating Good Practice
Ten Principles (TRAC, DRAMBORA, nestor)

1. Commits to continuing maintenance of digital objects for identified community/communities.
2. Demonstrates organizational fitness (including financial, staffing structure, and processes) to fulfil its commitment.
3. Acquires and maintains requisite contractual and legal rights and fulfils responsibilities.
4. Has an effective and efficient policy framework.
5. Acquires and ingests digital objects based upon stated criteria that correspond to its commitments and capabilities.
6. Maintains/ensures the integrity, authenticity and usability of digital objects it holds over time.
7. Creates and maintains requisite metadata about actions taken on digital objects during preservation as well as the relevant production, access support, and usage process contexts before preservation.
8. Fulfils requisite dissemination requirements.
9. Has a strategic program for preservation planning and action.
10. Has technical infrastructure adequate to continuing maintenance and security of its digital objects.\(^6\)
<table>
<thead>
<tr>
<th>Strategic Objective Plan</th>
<th>Responsibilities</th>
<th>Corresponding Core Principle(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business Plan</td>
<td>Financial planning, monitoring, and reporting</td>
<td>2</td>
</tr>
<tr>
<td>Staffing Plan</td>
<td>Acquisition and maintenance of relevant skillset for managing repository</td>
<td>2</td>
</tr>
<tr>
<td>Data Plan</td>
<td>Specification of data and metadata objects, formats, and structures for ingest, storage, and dissemination, together with the relevant transformations and mappings.</td>
<td>5, 6, 7, 8</td>
</tr>
<tr>
<td>Acquisition Plan</td>
<td>Management of the relationship with depositors and other data providers. Appraisal policy.</td>
<td>3, 5</td>
</tr>
<tr>
<td>Access Plan</td>
<td>Management of relationship with end users. Access Policy.</td>
<td>1, 8</td>
</tr>
<tr>
<td>Preservation Plan</td>
<td>Ensure that access and usability of material in repository is not adversely affected by technological change and obsolescence</td>
<td>9</td>
</tr>
<tr>
<td>Technical System Plan</td>
<td>Specifies goals for hardware, software and networking</td>
<td>10</td>
</tr>
<tr>
<td>Succession Plan</td>
<td>Manage obligation to ensure preservation of material beyond the lifetime of the repository</td>
<td>1</td>
</tr>
<tr>
<td>Disaster Plan</td>
<td>Respond to rapid changes to the repository environment</td>
<td>1, 6</td>
</tr>
</tbody>
</table>
Role of Audit

Benefits of audit (and self-assessment):
- Raises awareness during self-assessment
- Includes gap analysis
- Produces development plan
- Provides evidence for stakeholders
- Enables transparency for DP program
Examples of TRAC Review Results

- Formalize policies
- Define roles and responsibilities
- Consider succession planning
- Designate funding
- Rationalize metadata
- Address preservation rights
- Prioritize technical developments
Trusted Repositories Audit and Certification

This page provides a place for an organization to document its evidence for meeting the requirements of Trusted Repositories Audit and Certification (TRAC) checklist. A TRAC review is a method to demonstrate good practice and common responsibilities for each requirement.

Responsibilities: Each entity is assigned a role for each requirement using the RACI responsibility assignment matrix. RACI is especially useful in requiring distributed responsibilities. See the Responsibilities for TRAC page for more information on RACI committees that have roles in TRAC conformance.

Requirements: Each TRAC requirement has its own page. Sub- and Sub-sub requirements are referred to the page. Current compliance with TRAC requirements is assessed on a rating system from 0 to 4 (see exam:

0 = non-compliant
1 = slightly compliant
2 = half compliant
3 = mostly compliant
4 = fully compliant
# Requirement Status

## 3.1 Governance and Organizational Viability

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Compliance Rating</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1.1 Mission Statement</td>
<td>Slightly Compliant</td>
<td>Reviewed</td>
</tr>
<tr>
<td>3.1.2 Preservation Strategic Plan</td>
<td>Half Compliant</td>
<td>Not done</td>
</tr>
<tr>
<td>3.1.3 Collection Policy</td>
<td>Non-Compliant</td>
<td>Drafted</td>
</tr>
</tbody>
</table>

## 3.2 Organizational Structure and Staffing

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Compliance Rating</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2.1 Identified Required Competencies and Appointed Staff</td>
<td>Non-Compliant</td>
<td>Not done</td>
</tr>
</tbody>
</table>
3.1 Governance and Organizational Viability

3.1.2 Preservation Strategic Plan

The repository shall have a Preservation Strategic Plan that defines the approach the repository will take in the long-term support of its mission. For more information

3.1.2.1 The repository shall have an appropriate succession plan, contingency plans, and/or escrow arrangements in place in case the repository ceases to operate or the governing or funding institution substantially changes its scope.

3.1.2.2 The repository shall monitor its organizational environment to determine when to execute its succession plan, contingency plans, and/or escrow arrangements.

Supporting Text

This is necessary in order to help the repository make administrative decisions, shape policies, and allocate resources in order to successfully preserve its holdings.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Preservation Strategic Plan, meeting minutes, documentation of administrative decisions which have been made.

Discussion

The strategic plan should be based on the organization’s established mission, and on its defined values, vision and goals. Strategic plans typically cover a particular finite time period, normally in the 3-5 year range.

Evidence Examples

This would be examples of possible evidence.

Evidence Provided

Here we list the evidence provided.

<table>
<thead>
<tr>
<th>Compliance Rating</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Half Compliant</strong></td>
<td><strong>Not done</strong></td>
</tr>
</tbody>
</table>
3.1.2.1 The repository shall have an appropriate succession plan, contingency plans, and/or escrow arrangements in place in case the repository ceases to operate or the governing or funding institution substantially changes its scope.

Supporting Text

This is necessary in order to preserve the information content entrusted to the repository by handing it on to another custodian in the case that the repository ceases to operate.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Written and credible succession and contingency plan(s); explicit and specific statement documenting the intent to ensure continuity of the repository, and the steps taken and to be taken to ensure continuity; escrow of critical code, software, and metadata sufficient to enable reconstitution of the repository and its content in the event of repository failure; escrow and/or reserve funds set aside for contingencies; explicit agreements with successor organizations documenting the measures to be taken to ensure the complete and formal transfer of responsibility for the repository’s digital content and related assets, and granting the requisite rights necessary to ensure continuity of the content and repository services.

Discussion

A repository’s failure threatens the long-term sustainability of a repository’s information content. It is not sufficient for the repository to have an informal plan or policy regarding where its data goes should a failure occur. A formal plan with identified procedures needs to be in place.
Responsibilities

Stakeholder groups are assigned responsibilities using RACI. The RACI Matrix describes participation by various organizational roles in completing tasks for a project. RACI is especially useful in clarifying roles in projects and processes requiring distributed responsibilities.

Definitions of the RACI Categories:

- **Responsible**: person or group who performs an activity or does the work
- **Accountable**: person or group who is ultimately accountable and has Yes/No/Veto
- **Support**: person or group that assists in completing task
- **Consulted**: person or group that needs to feedback and contribute to the activity
- **Informed**: person or group that needs to know of the decision or action

Note for clarification: Resources are allocated to Responsible who gets input from Consulted, assistance from Support, and guidance from Accountable. For more about RACI, see the following resources:

- RACI model | RACI chart | RACI method
- Wikipedia, Responsibility Assignment Matrix
- Accountability: Great Info On The RASCI / RACI Matrix
- How to Do RACI Charting and Analysis
The list below consists of the stakeholder groups that have a role to play in TRAC compliance. TRAC roles and responsibilities are enumerated.

- Senior Management
- Coordination Group
- Operations Group
- Information Technology
- Administration: Finance or HR
- Acquisitions
- Preservation
- Dissemination
- Rights Management
- External Advisory Group
Senior Management

This page consists of a RACI chart to assist in assigning and tracking the responsibilities in regards to TRAC compliance. See the Responsibilities for TRAC for an outline of roles included in the following RACI chart. For additional guidance, please also see the Suggestions for Performing Assigned Roles page.

Responsible

3.1.2 Preservation Strategic Plan
3.1.3 Collection Policy

Accountable

3.1.1 Mission Statement
3.2.1 Identified Required Competencies and Appointed Staff

Support

Consulted

Informed

If you have trouble accessing this page and need to request an alternate format, please contact Nance McGovern.
SafeArchive in Action

safearchive.org
Implementation

The SafeArchive System

Collaborating Institutions

Policies

TRAC Input Form

Schema Form

Content

Content

Preservation Network

Replication

Discovered Network Data

Policies

Audit Schema Manager

Create Audit Schema

Edit Audit Schema

Schema Comparison Tool

Preservation Enforcer

Network Monitor

Cache Status Extractor

Cron Service

Console

LockSS Management Tool (future development)

Report Generator

View Audit Reports

View Operational Reports

Manually Refresh Audit

Generate Historical Reports

Repair or Reprovision

Enforcement

Comparison

Monitoring

Audit Report

Operational Report

Reports
Project Description

Dissemination Information Packages (DIPS) for Information Reuse (DIPIR)

DIPIR is an IMLS-funded project led by Dr. Ixchel Faniel and Dr. Elizabeth Yakel. Together with partners at The Inter-university Consortium for Political and Social Research, the University of Michigan Museum of Zoology, and Open Context, they are studying data reuse in three academic disciplines to identify how contextual information about the data that supports reuse can best be created and preserved. The project focuses on research data produced and used by quantitative social scientists, archaeologists, and zoologists. The intended audiences of this project are researchers who use secondary data and the digital curators, digital repository managers, data center staff, and others who collect, manage, and store digital information. Knowledge gained from the study will help guide current and future international practices for curating and preserving digital research data.
Reverse engineered TRAC to develop social science domain survey instrument

- In what ways might consumers be aware of TRAC?
- To what extent might it matter?
- How might their awareness affect (re)use?

Use of terms across domains:

- Data, metadata, use, re-use...
Repository Trust Concepts

Integrity
Benevolence
**Transparency**
Identification-based trust
**Social Factors**
**Structural Assurances**
Performance Expectancy

Trust

Intent to continue using
ANADP

Aligning National Approaches to Digital Preservation

- Envisioning an International Community of Practice
- National examples (Estonia, USA, Sweden)
- Alignment aspects:
  - Legal
  - Organizational
  - Standards
  - Technical
  - Resources
  - Education
- Alignment Opportunities (with Cliff Lynch)

ANADP released August 2012
http://www.educopia.org/publications