Digital Curation 101

STORE

Store is the sixth sequential action in the curation lifecycle, following *Preservation Action*.

Topics:

- Store
- Storage requirements
  - Organisational structure and continuity
  - Technical infrastructure and practices
- Ensuring quality of storage
  - Open Archival Information System (OAIS) Reference Model
  - Trusted Digital Repositories (TDRs)
- The next action in the curation lifecycle

Store

Store is the sixth sequential action of the data curation lifecycle. Its activity is:

- Storing the data in a secure manner adhering to relevant standards.

After data have been appraised and ingested, and appropriate preservation actions have been applied, they need to be stored securely.

Storage facilities should ensure that:

- Data are stored securely over time: their integrity is not compromised
- The storage is reliable: data is not lost
- They meet the requirements of relevant standards
- Data in storage can be accessed: they can be got out of storage for use and reuse.

Storage requirements

To be suitable for long-term curation purposes, a storage facility must be more than just an installation of repository software such as DSpace or Fedora. The key requirement it should meet is *sustainability*. This means that a long-term curation facility needs to be sustainable both in its organisational structure and continuity, and in its technical infrastructure and practices.

Organisational structure and continuity

A long-term curation facility should be sustainable in the way it is set up and organised. The facility should:

- Demonstrate a commitment to continuing maintenance of digital objects for an identified community (or communities)
- Ensure adequate and appropriate finance and staffing to fulfil the commitment
- Negotiate the requisite contractual and legal rights, and fulfil these legal responsibilities
- Develop an effective and efficient policy framework
- Develop a strategic program for preservation planning and action.

Technical infrastructure and practices

To be sustainable a long-term curation facility should have a technical infrastructure that provides continuing maintenance and security of the digital objects stored in it. That is, the integrity, authenticity and usability of the digital objects it stores can be maintained over time.

A sustainable facility bases its operations on appropriate practices. First, it should acquire and ingest data based upon stated criteria that correspond to its commitments and capabilities. Next, to store these data sustainably the facility should:

- Store the data in formats that:
    - Do not apply any form of manipulation which causes data loss or loss of authenticity
    - Are widely implemented and supported
    - Preferably are open or non-proprietary
    - Have a potentially long life
    - Are most likely to have migration pathways to the next format available
    - Store enough metadata and representation information to support identification, access and preservation processes

- Use a reliable storage format on at least two types of carrier

- Make multiple copies, which are checked and verified regularly
- Replace carriers and software as the market demands, with plans to migrate the content to the next type of reliable carrier[1], [2].

Ensuring quality of storage

Two relevant standards that define criteria for long-term storage are the OAIS (Open Archival Information System) Reference Model, and the Trusted Digital Repository (TDR) model.

OAIS Reference Model

The key standard is the OAIS Reference Model (ISO 14721:2003).

Three of the five functions defined by OAIS are especially relevant to the *Store* action:

- The *Archival Storage* function which ‘handles the storage, maintenance and retrieval of the Archive Information Packages (AIPs) held by the archive’
- The *Data Management* function which ‘coordinates the Descriptive Information pertaining to the archive’s AIPs, in addition to system information used in support of the archive’s function’
- The *Administration* function which ‘manages the day-to-day operation of the archive’.

Archival Storage defines requirements for the effective storage of AIPs, routine integrity checking, and disaster recovery. Indicative components include:

- For effective storage of AIPs: moving AIPs from Ingest into permanent storage, refreshing the storage media, providing the information needed to allow objects to be disseminated from the repository
- For routine integrity checking: checking the stored bitstream regularly to assess whether it is identical to the original bytestream
- For disaster recovery: ongoing media refreshment, ongoing monitoring of media for deterioration, geographically distributed backup systems.

[1] Based on Kevin Bradley, Junran Lei, Chris Blackall, *Towards an Open Source Repository and Preservation System, 2007* http://portal.unesco.org/ci/en/ev.php-URL_ID=24700&URL_DO=DO_PRINTPAGE&URL_SECTION=201.html

[2] http://www.crl.edu/content.asp?l1=13&l2=58&l3=162&l4=92

Digital Curation 101

Data Management defines all aspects of the OAIS repository's management. Indicative components include:

- Access controls
- Security information used to authenticate users of the repository
- Statistical information to improve operation
- Policies for and monitoring of the allocation of unique identifiers.

Archive Administration defines the services needed for day-to-day maintenance of the repository. Indicative components include:

- Negotiating submissions agreements with content producers and providers
- Reviewing procedures
- Maintaining systems configurations for hardware and software
- Developing and maintaining repository policies and standards
- Negotiating user access agreements with service providers or others.

Trusted Digital Repositories (TDRs)

A TDR is defined as,

> 'one whose mission is to provide reliable, long-term access to managed digital resources to its designated community, now and in the future[3].'

A TDR must meet certain requirements. One is that it should be compliant with the OAIS Reference Model. Other requirements are to:

- Accept responsibility for the long-term maintenance of digital resources on behalf of its depositors and for the benefit of current and future users
- Have an organizational system that supports not only long-term viability of the repository, but also the digital information for which it has responsibility
- Demonstrate fiscal responsibility and sustainability
- Design its system(s) in line with commonly accepted conventions and standards to ensure the ongoing management, access, and security of materials deposited within it
- Establish methodologies for system evaluation that meet community expectations of trustworthiness

[3] RLG, *Trusted Digital Repositories: Attributes and Responsibilities An RLG-OCLC Report* (2002) http://www.oclc.org/programs/ourwork/past/trustedrep/repositories.pdf

- Be trusted to carry out its long-term responsibilities to depositors and users openly and explicitly
- Have policies, practices, and performance that can be audited and measured

How do you know if the digital repository can be trusted? What assurance is there that the digital repository can maintain its contents over time? Audit and certification of digital repositories are developing. Current audit and certification of digital repositories is based on self-assessment. Examples are:

- TRAC: the Trustworthy Repositories Audit and Certification (TRAC) Criteria and Checklist[4] (US Center for Research Libraries), which lists characteristics (for example, ‘Repository has procedures and policies in place’) that trusted repositories should demonstrate but does not indicate how these can be assessed

- nestor’s Catalogue of Criteria for Trusted Digital Repositories (version 1, 2006)[5], also checklist-based

- DRAMBORA: the *Digital Repository Audit Method Based on Risk Assessment (DRAMBORA)*[6] toolkit*,* developed by the DCC and DigitalPreservationEurope, which provides a more formal auditing methodology based on risk assessment, resulting in a structured registry of risks

- PLATTER (Planning Tool for Trusted Electronic Repositories)[7] has been developed to complement DRAMBORA by assisting with planning of repository goals, objectives and performance targets to establish trusted repository status.

The next action in the curation lifecycle

The next sequential action in the curation lifecycle is *Access, Use & Reuse* which describes how data are made accessible to legitimate users and used and reused by them.

[4] http://www.crl.edu/content.asp?l1=13&l2=58&l3=162&l4=91

[5] http://www.crl.edu/content.asp?l1=13&l2=58&l3=162&l4=91

[6] http://www.repositoryaudit.eu/

[7] http://www.digitalpreservationeurope.eu/publications/reports/Repository_Planning_Checklist_and_Guidance.pdf