CONUL: Survey of Digital Infrastructure & Services

Guidelines for completing this survey
This survey is structured around the Open Archive Information System (OAIS) Reference Model. For a brief introduction to OAIS, see www.dpconline.org/docs/lavoie_OAIS.pdf. For a glossary of OAIS terms used in this survey, see http://www.dcc.ac.uk/digital-curation/glossary

Please complete this survey for each digital repository system managed by your organisation.

1. Background

1.1 Please indicate the digital repository system/stack used by your library:

- DSpace
- Eprints
- Digital Commons
- Fedora-Commons
- Fedora-Commons + Hydra
- Fedora-Commons + Islandora
- Safety Deposit Box
- ContentDM (OCLC)
- Rosetta (ExLibris)
- Archivematica
- Other, please specify

1.2 How long has your library used this system?

2. Functional Model

2.1 Ingest Function

2.1.1 Please estimate the percentage of content ingested in each of the following broad categories:

- Non-library staff (e.g. researchers) deposit individual items through the repository web interface _____ %
- Library staff deposit individual items through the repository web interface _____ %
- Non-library staff batch load multiple items through the repository web interface _____%
- Library staff batch load multiple items through the repository web interface ____ %
- Library staff batch load multiple items using an external client or script _____ %
- Remote deposit of content from a 3rd-party system or client (e.g. direct deposit from a publish system using SWORD protocol or equivalent API) ____ %

2.1.2 Which of the following actions is carried out as part of the ingest process (as opposed to

pre-ingest by another system). Tick all that apply:

- virus checking
- file characterization
- generation of a normalized preservation version of files (e.g. create a PDF/A version of a deposited PDF file)
- generation of derivative/access files (e.g. thumbnail generation, OCR process)
- file fixity checking (e.g. comparing checksums of ingested files to known pre-ingest values)
- creation of descriptive metadata
- creation of structural metadata or relationships between files
- creation of administrative/rights metadata
- extraction of technical preservation metadata, e.g. using JHOVE or equivalent
- Other, please specify: _____

2.2 Archival Storage Function

2.2.1 Which of the following best describes the library's responsibility for procuring and administering the storage infrastructure used by . Tick all that apply.

- The library procures and administers local hardware
- The library procures & manages hosted storage (e.g. Amazon Web Services)
- Hardware is procured and administered by the library's parent institution (e.g. space allocated & managed by a University)
- Storage is provided & administered by a 3-party company (e.g. as part of a hosted repository solution)

<u>Primary/"Live" Data</u>

2.2.2 For the canonical copy of content in the *Archival Information Package*, please indicate what percentage is stored on the following storage tiers:

- Online (immediately available to the repository): _____ %
- Nearline (stored on external storage media but automatically retrievable upon request, e.g. a LTO Tape Library): _____ %
- Offline (data is stored on external media which must be manually retrieved and accessed, e.g. CDs/DVDs, disconnected external HD, etc.): _____ %
- Other/Not Applicable, please specify: _____

2.2.3 For content stored online, which of following best describes the storage architecture (for a definition of terms, please see http://docs.openstack.org/trunk/openstack-compute/install/yum/content/terminology-storage.html):

- Block Storage (SAN)

- File Storage (NAS)
- Object Storage
- Other, please specify_____

2.2.4 Please indicate the current size of all Archival Information Package content managed by your repository system:

- < 1TB
- 1 - 10 TB
- 10 - 30 TB
- 30 - 60 TB
- 60 - 100 TB
- > 100 TB

Possible question here on how data is arranged on the filesystem, e.g.
a) locally managed directory structure (e.g. bagit, dflat conventions, local structure),
b) managed by the repository software (e.g. date-based folder structure, Akubura, Modeshape),
c) managed by storage device (e.g. content addressed storage, object storage etc)

Backup

2.2.5 For backups of content *Archival Information Package*, please indicate what type of storage tier is used. Tick all that apply

- Online (e.g. local backup-to-disk, or cloud backup)
- Nearline (backups stored on external storage media but automatically retrievable upon request, e.g. a LTO Tape Library)
- Offline (baackups stored on external media which must be manually retrieved and accessed, e.g. CDs/DVDs, disconnected external HD, etc.)
- Other/Not Applicable, please specify: _____

2.2.6 How regularly do you test backup media and processes:

- Annually
- Monthly
- Weekly
- No regular schedule
- Other, please specify: _____

Disaster Recovery & Business Continuity

2.2.7 Which of the following best describes your disaster recovery and business continuity setup. Tick all that apply.

- backups of Archival Information Package (AIP) content is kept offsite. The full repository technology stack can (eventually) be reconstituted to the point of the last backup
- consistent snapshots of AIP content & all systems and data used by the repository are

stored offsite, so that the entire technology stack can be brought back online to the point of the last snapshot
- AIP content and repository systems are continuously replicated and can immediately be made available in the case of a disaster
- Other, please specify: _____

Data Integrity/File Fixity

2.2.8 Do you run regular file fixity audits of repository AIP content (e.g. checksum comparison)? Y/N

2.2.9 Please give details of how file fixity audits are carried out (e.g. built-in feature of repository, built-in feature of storage device, custom script, etc.) _____

Other

2.2.10 Does your repository keep older versions of files when a digital object is updated?

2.2.11 Does your repository allow content to be permanently deleted?

2.3 Access Function

2.3.1 Please indicate the ways in which authorised users can access repository content. Tick all that apply

- Content can be viewed or downloaded through repository system's built-in web-interface
- Content can be directly viewed or downloaded from a separate discovery layer used by your library (e.g. Encore, Summon, Vufind)
- Content can be directly viewed or downloaded from specialised content portals (e.g. a digital humanities project or omeka exhibition which directly interface with your repository system)
- Content is directly accessible through documented APIs (e.g. International Image Interoperability Framework)
- Content can be harvested and kept in sync by external systems (e.g. using OAI-ORE)
- Metadata can be harvested and kept in sync by external systesms (e.g. using OAI-PMH)
- Content is published on the open web (e.g. metadata published as linked data and content accessible via Cool URIs)
- Other, please specify _____

Questions on Security/Authentication/Authorisation?

2.4 Data Management Function

2.4.1 Which of the following best describes how descriptive metadata for objects in the

repository is maintained following the initial ingest process. Tick all that apply

- Updates are carried out in the repository system (e.g. through the repository systems native web interface)
- Updates are carried out in an external system & synchronised to the Archival Information Package in the repository (e.g. Archival Descriptions managed in an Archival Management System; MARC records in an LMS)
- Descriptive metadata is never altered following ingest
- Other, please specify _____

2.4.2 Which of the following best describes the management of Authority Control within your repository system following the initial ingest/creation of descriptive metadata

- the system provides authority control functionality (e.g. an author name or subject heading can be corrected & applied descriptive metadata for all relevant AIPs)
- authority control is not carried out
- authority control is handled by in external system & updates to descriptive metadata is synchronised to the Archival Information Package in the repository (e.g. Archival Management System; LMS)
- Authority control is inherently handled outside of the repository (e.g. descriptive metadata in the AIPs only stores references to Linked Data URIs)
- Other, please specify _____--

2.4.3 What types of reports do you make available regarding the administration, performance, and use of the repository system? Tick all that apply

- Reports on quantity of material (e.g. number of objects, number of objects by type)
- Reports on provenance of material (e.g. items by producer)
- Reports on usage (e.g. number of downloads)
- Security reports (i.e. login logs etc.)
- Other, please specify _____-

2.5 Preservation Planning Function

2.5.1 Does your library have a designated unit or individual responsible for planning preservation actions for your repository? Y/N

2.5.2 Does your library have a documented set of preservation policies and strategies for content in your repository? Y/N

2.5.3 Has your library assessed the long-term trustworthiness of your repository using Trusted Repository Audit Checklist (TRAC) or equivalent?

2.6 Administration Function

No questions under this function? Covered by staffing question? Difficult to quantify interaction with external stakeholders (Producers/Management/Consumers)

Could have additional questions here regarding deposit agreements, rights management policies etc.

2.7 Resourcing & Skills

Please indicate the amount of library staff time dedicated to each of the OAIS Functions. Note, the Archival Storage Function and Access Function have been omitted from the table below; in terms staffing, these repository functions are most likely serviced by the Administrative Function

Role	FTE	Skills/Qualifications/Years of experience
Ingest function, i.e. staff involved in the process of loading and verifying content		
Data management function, i.e. staff involved in the maintenance of metadata, reports, and statistics		
Preservation Planning function, i.e. staff members involved in defining, enforcing, monitoring and reviewing preservation policies and processes		
Administration function: Management role: including managing repository staff and liaising with Producers, Consumers, and wider the wider library Management Team		
Administration function: Software development		
Administration function: Systems Administration/ Operations		

3. Information Model

3.1 Submission Information Package

3.1.1 Do you use any standards to define the structure of content packages deposited to your repository::

- Bagit
- METS
- OAI-ORE
- Other, please specify _____

3.2 Archival Information Package

Descriptive Metadata

3.2.1 What schemas are used to describe AIPs. Tick all that apply

- DC
- MODS
- MARC21
- PB Core
- VRA Core
- EAD
- Other, please specify _____

3.2.2 Which of the following best describes how the permanent, canonical version of descriptive metadata is stored within the system for each digital object. Tick all that apply. Note, do not include caching mechanisms: for example, if metadata is stored in XML but cached in a relational database or full-text index to facilitate the operation of the system, only tick XML.

- An XML file containing only descriptive metadata (e.g. MARCXML file)
- Part of an XML file which wraps various types of metadata (e.g. a METS XML file)
- A plain text file containing a non-xml serialisation of descriptive metadata (n3, json, etc)
- A binary file (e.g. MARC21 as ISO2709)
- A relational database
- A triple-store
- Don't know
- Other, please specify _____

Structural Metadata

3.2.3 Please indicate which of the following is used to store structural information about the relationship between the files which make up the AIP:

- METS

- Other non-rdf based XML schema (e.g. locally defined schema, or Hydra conentMetadata schema), please specify _____
- RDF (e.g. OAI-ORE Resource Map, or Fedora's RELS-EXT datastream)
- Structural relationships are maintained by the repository system's relational database schema

Administrative, Rights, and Technical Metadata

3.2.4 Please indicate which of the following is used to store administrative information about the files which make up the AIP (including representation information such as technical characteristics, and rights and access metadata). Tick all that apply

- PREMIS XML Schema
- MIX XML Schema
- PREMIS or MIX XML Schema wrapped in METS
- Other XML schema (e.g. locally defined schema, or Hydra rightsMetadata schema), please specify _____
- Structural relationships are maintained by the repository system's relational database schema

3.2.5 Do you extract and store technical metadata separately from the information embedded in content files themselves?

3.2.6 Do you keep an audit log of actions/events carried out on each digital object in the repository (i.e. a log of each Create, Read, Update, Delete operation)?

3.2.7 Does your organisation use a globally unique persistent identifier scheme for objects in the repository. Tick all that apply:

- DOI
- ARK
- Handle
- PURL
- EZID
- Locally defined URI scheme, with a strong organisational commitment to long-term maintenance, i.e. "Cool URIs"
- Other, please specify _____

3.3 Dissemination Information Package

other questions here? largely covered under Access under Functional model

4. Operating Environment

4.1 Producers/Depositors

4.1.1 Please give a rough estimate of the percentage of material deposited under the following

broad category of producer:

- Library Digitisation Programme ____ %
- Non-library units belonging to your parent organisation (e.g. Records Office) ____ %
- Individuals belonging to your parent organisation (e.g. researchers) ____ %
- Individuals not belonging to your parent organisation (e.g. acquisition of papers) ____ %
- Organisations not belonging to your parent organisation (e.g. e-legal deposit) ____ %
- Other, please specify _____

Question on Content Types, i.e. what content is collected?

4.2 OAIS Management

both establish current management environment (policies, organisational buy-in etc) & ask questions in relation to potential for CONUL collaboration

4.2.1 Funding streams
4.2.2 Strategic Plan & Organisational Support/Buy-in
4.2.3 Relevant policies devloped
4.2.4 External partnerships

4.2.5 Would your organisation consider migrating to another repository system in order to align with other CONUL members?

4.3 Consumers

4.3.1 Has your organisation identified and defined a “designated community” or “designated communities” as defined by the OAIS Reference Model, i.e. “a subset of Consumers expected to independently understand the archived information in the form in which it is preserved and made available by the OAIS.”?