Digital Preservation Systems Showcase

Time: Tuesday, 7th October 2014
Venue: Theatrette

Description: This session explores the functionality of digital preservation systems available to the user community. System developers showcase their systems in line with a pre-determined set of functions. These functions are derived from current standards, key literature and interest groups (e.g. PREMIS, OAIS, and the International Internet Preservation Consortium Preservation Working Group).

Running order:

09:00 – 09:15	Introduction
09:15 – 10:15	DuraSpace
10:15 – 11:15	Artefactual Systems
11:15 – 11:40	Break
11:40 – 12:40	KEEP Solutions
12:40 – 13:40	Lunch
13:40 – 14:40	Preservica
14:40 – 15:00	Break
15:00 – 16:00	Ex Libris
16:00 – 17:15	Questions and round-up

Presentation structure

The DP system providers explored the questions listed below, demonstrating how their products handle these core components of digital preservation. The fundamental issue was to highlight how their systems tackle the key areas, taking into account the contextual 'Considerations' listed below.

Showcase details

The preservation workflow for digital preservation systems can be simplified into three large groups.

- How do we get content in?
- How do we manage and preserve it once in?
- How can the content be accessed from the system?

More detailed functional areas are listed under each of these groups.

How do we get it in?
- Ingest flows / methods
  What are the flows that can be used to route digital content into the system. Are there difference between flows (for example, different assessment criteria, more detailed identification, stricter security?)
- Preconditioning / pre-ingest preparation

Does the system take care of any actions that may be considered 'preparation' of content for ingest. This may include such actions as adding correct file extensions, repairing 'broken' files, tidying of file names.

- Format identification
  How does your system identify formats? To what level is identification made? What tools and resources does it use? How is format identification used by the system?
- Metadata extraction
  How does your system extract metadata? This may include technical and descriptive metadata. How much metadata is extracted and for what purpose? What tools and resources does it use?
- Fixity checking/assignation
  Does the system check fixity supplied to it? What type of fixity recording/checking mechanisms are used?
- Virus checking
  Does the system check for malware? What tools does it use? What happens if a virus is discovered?

How do we manage and preserve it?
- Intellectual management
  Is intellectual management (e.g. cataloguing) done by the system, or this a dependency on another system?
- Risk analysis
  Does the system do risk analysis of content based on technical form of the content? If so, what is it checking? What information is given to system users and what can they do with that information?
- Preservation planning
  What is the process for preservation planning in the system? What tests/proofs/sign-offs are required? How does it relate, if at all, to current community practice in (for example PLATO1)?
- Preservation execution
  How does your system undertake preservation actions (migration and/or emulation)?
- Repository management (queries, monitoring, analysis, updates)
  What reports is the system capable of generating? Is there a repository dashboard to analyse contents? What statistics are delivered out-of-the-box? Can functions such as format identification, virus checking, fixity checking and metadata extraction be re-run as required? How are updates for third party tools dealt with?
- Exception handling
  What functions exist for users to deal with exceptions in any process (including ingest processes?)

How do we access it?
- Derivative generation (static, on-the-fly, options of types)
  How is access given to content? Are access copies made of masters? If so, what are the formats used (and what are their master formats)? How are these copies made?
- Access rights
  How are rights administered and managed by the system?
- Complex materials
  Are there special access methods available for particularly complex materials such as email, webharvests, full text, multiple-object materials?
- Handing over to other access methods
  Can the system hand over materials to other access mechanisms? How easy is this handover? Are there any constraints?

---

[1] http://www.ifs.tuwien.ac.at/dp/plato/intro/.

- Export of data
    Can data be exported from the system?

Considerations
- Flexibility/interoperability of the system
    What external (to the system) sources is the system dependent on? How is reflected in the concept of the Archival Information Package? How are updates in those dependencies managed (for example content management system changes to access rights, or identifiers).
- Exit strategy
    How locked-in are customers?
- Archival Information Package
    o Relationship to PREMIS and other metadata schemas.
        Does the system implement PREMIS? Is it a conformant implementation? If not, why not?
    o Data model
        What is the object model used by the system? What level of detail captured about the object (intellectual entity only, or all the way down to bitstream information [as per PREMIS])?
- Provenance
    What data is kept to track/note provenance of the content? What triggers the new generation of metadata in this trail?
- Large/small, bulk/single
    How does the system deal with the very large and the very small and boutique (both in terms of size and number?
- Testing
    What testing regimes/tools are in place for new releases?
- Storage
    Does the system promote a particular type of storage? Are there any constraints on the configuration of storage?

Moderator
Moderated by Ross King, Chairman of the Board at Open Planets Foundation and Senior Scientist at the Austrian Institute of Technology.

Digital Preservation Systems Showcase – Audience Notes

DuraSpace / Artefactual Systems / KEEP Solutions / Preservica / Ex Libris

How do we get it in?

Ingest flows / methods

....
....
....
....

Preconditioning / pre-ingest preparation

....
....
....
....

Format identification

....
....
....
....

Metadata extraction

....
....
....
....

Fixity checking/assignation

....
....
....
....

Virus checking

....
....
....
....

How do we manage and preserve it?

Intellectual management

....
....
....
....

Risk analysis

....
....
....
....

Preservation planning

....
....
....
....

Preservation execution

....
....
....
....

Repository management (queries, monitoring, analysis, updates)

....
....
....
....

Exception handling

....
....
....
....

How do we access it?

Derivative generation (static, on-the-fly, options of types)

....
....
....
....

Access rights

....
....
....
....

Complex materials

....
....
....
....

Handing over to other access methods

....
....
....
....

Export of data

....
....
....
....

Considerations

Flexibility/interoperability of the system

....
....
....
....

Exit strategy

....
....
....
....

Archival Information Package

....
....
....

- o Relationship to PREMIS and other metadata schemas.

....
....
....

- o Data model

....
....
....

Provenance

....
....
....
....

Large/small, bulk/single

....
....
....
....

Testing

....
....
....
....

Storage

....
....
....
....