Riksarkivet

Different Aspects of Digital Preservation

DCH-RP and EUDAT Workshop in Stockholm 3rd of June 2014

Börje Justrell

Table of Content

- Definitions
- Strategies
- The Digital Archive Lifecycle

Digital preservation - Definitions

A major difficulty in digital preservation is the lack of a precise and definitive taxonomy of terms. Different communities use the same terms in different ways which can make effective communication problematic. The following definitions are for practical use but may not necessarily achieve widespread consensus among the wide ranging of cultural heritage institutions.

Preservation is NOT concerned only with sustaining single digital objects. Digital objects should be preserved in **context** which makes them understandable.

In European calls for R&D project it is often said, that preservation is on hand when digital objects are **accessible** and **usable** to future users.

Digital preservation - Definitions

Digital preservation

DigitalPreservationEurope project defines digital preservation as: "*a set of activities required to make sure digital objects can be located, rendered, used and understood in the future*."

http://www.digitalpreservationeurope.eu/what-is-digital-preservation/

Digital curation

Digital curation is often used in parallel with digital preservation; it has wider coverage and involves "*maintaining, preserving and adding value to digital data throughout its life-cycle*".

http://www.dcc.ac.uk/digital-curation/what-digital-curation

Digital preservation - Definitions

Digital archiving

This term is used very differently within sectors. The library and archiving communities often use it interchangeably with digital preservation. Computing professionals tend sometimes to use digital archiving to mean the process of backup and ongoing maintenance (including storage) as opposed to strategies for long-term digital preservation.

Digital preservation - Definitions

Digital objects

Range from relatively simple, text-based files (e.g. word processing files), to highly sophisticated web-based resources which fully exploit the benefits of technology by combining sound with images, the ability to link to other resources, and the ability to interrogate.

Include **Born digital** objects, which are not intended to have an analogue equivalent, either as the originating source or as a result of conversion to analogue form (print out).

Digital preservation - Definitions

Digitisation

The process of creating digital files by scanning or otherwise converting analogue materials. The resulting digital copy, or digital surrogate, could then be classed as a digital object to sustain and consequently subject to the same broad challenges involved in preserving accessibility and usability to it, as "born digital" materials.

Digital preservation - Definitions

Authenticity

The digital material is what it purports to be.

In the case of electronic records, it refers to the trustworthiness of the electronic record as a record.

In the case of "born digital" and digitised materials, it refers to the fact that whatever is being cited is the same as it was when it was first created unless the accompanying metadata indicates any changes. Confidence in the authenticity of digital materials over time is particularly crucial owing to the ease with which alterations can be made.

Digital preservation - Definitions

Long-term preservation

Continued access to digital objects or at least to the information contained in them, indefinitely.

Medium-term preservation

Continued access to digital objects beyond changes in technology for a defined period of time but not indefinitely.

Short-term preservation

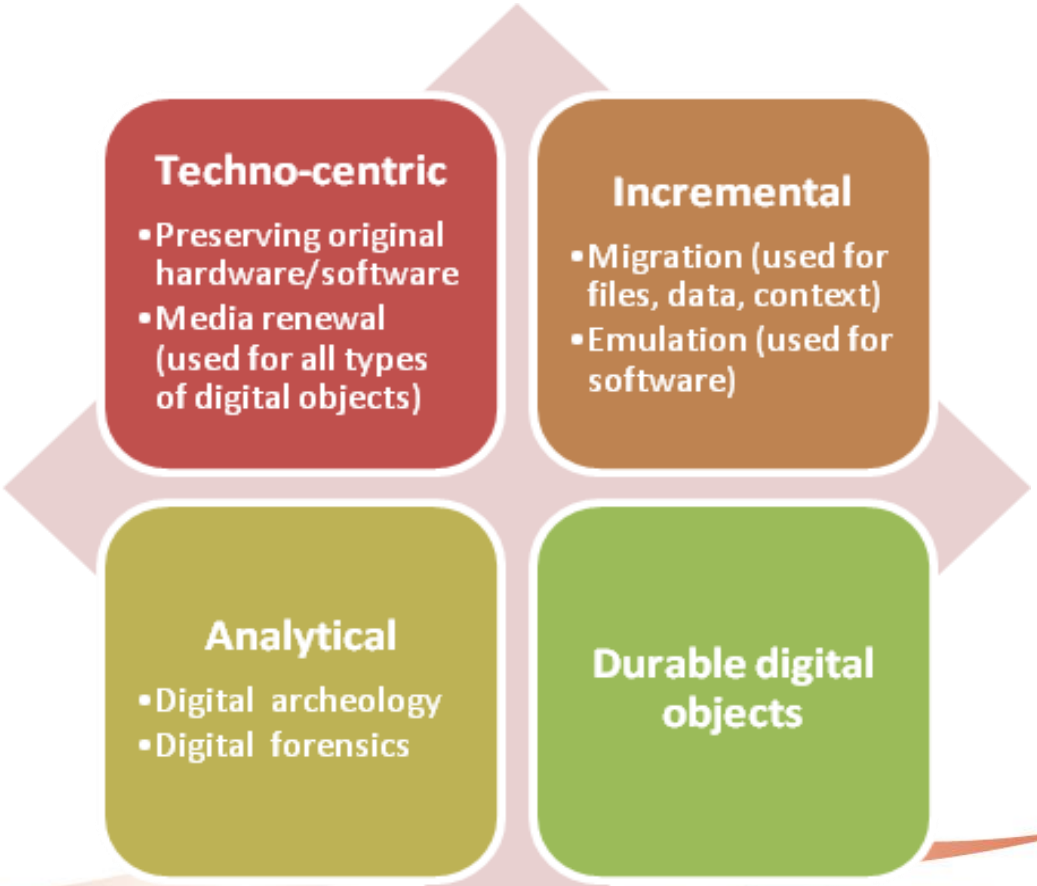
Access to digital objects either for a defined period of time while use is predicted but which does not extend beyond the foreseeable future and/or until they become inaccessible because of changes in technology.

Digital Preservation - Strategies

The key challenge in preserving the accessibility and usability of digital objects over time is to overcome technology obsolescence, but - other issues around managing collections of digital objects are also involved.

There are several strategies for sustaining future us of digital objects.

Digital Preservation - Strategies



Source: *Digital Preservation Services: State of the Art Analysis* (Raivo Ruusalepp and Milena Dobreva)

Digital Preservation - Strategies

The *techno-centric strategy* aims to preserve original hardware and software in a usable state; it involves regular storage media renewal to make sure that the physical digital objects are not corrupted.

Incremental change relies on either

- *migration of digital objects* into new formats; the migration strategy normally uses standardised file formats which are repeatedly converted to keep up with present technical generation

- or preserving the formats of the digital objects and *using emulation* to be able to use them; the emulation strategy preserves the original file formats and uses emulation at alternative levels to enter them into new technical environment in combination with preserved original software.

Digital Preservation - Strategies

Analytical strategies are currently based on techniques used in computer forensics; the underlying logic is to apply specialised methods for recovery of objects which are in demand in the future instead of ‘mass preservation’; the pioneering work in this domain was called *digital archaeology*.

Yet another strategy seeks for ways of changing the formats of the digital objects in a way which allows the objects themselves to call on preservation actions. Such objects are some times called *Durable digital objects*.

Digital Preservation - Strategies

The first three strategies require rigorous organisation of processes in organisations; the fourth one is still under development.

All four strategies outline the principles of preservation; in practice they are implemented within archival lifecycles that integrate various tools and/or services.

These lifecycles can be specific to organisations, depending on the type of objects they hold and their target users.

Digital Preservation - Strategies

Migration of digital objects into new formats is the strategy that has been the dominant one; it is used by most institutions working with digital preservation.

When converting the digital objects to new standardised file formats at technical generations changes, the conversions are expected to be done without information losses.

To be able to achieve that, it is of vital importance to *fully control the file formats.*

Digital Preservation - Strategies

Regardless of which strategy or combination of strategies that is chosen, cultural heritage institutions often make a distinction between the **master version** of digital data and at least one surrogate delivery version. The master version should contain as much intellectual, visual or audio content as possible, be saved in a standard (non-proprietary) file format, and preferably be duplicated across multiple locations.

Delivery versions of data may be re-sized, compressed, and saved in whichever format is suitable for delivery to the user. Delivery versions are typically of lower quality (more compressed) than their original master files.

Digital Preservation – the Digital Archive Lifecycle

The diversity of digital objects and types of institutions responsible for their preservation create a variety on the level of tools used in practice – but, the underlying process could today be described as universal.

The key standard is *ISO 14721:2012 Space data and information transfer systems – Open archival information system – Reference model,* known as the *OAIS model; it provides* a functional framework that presents the main components and the basic data flows within a digital preservation system.

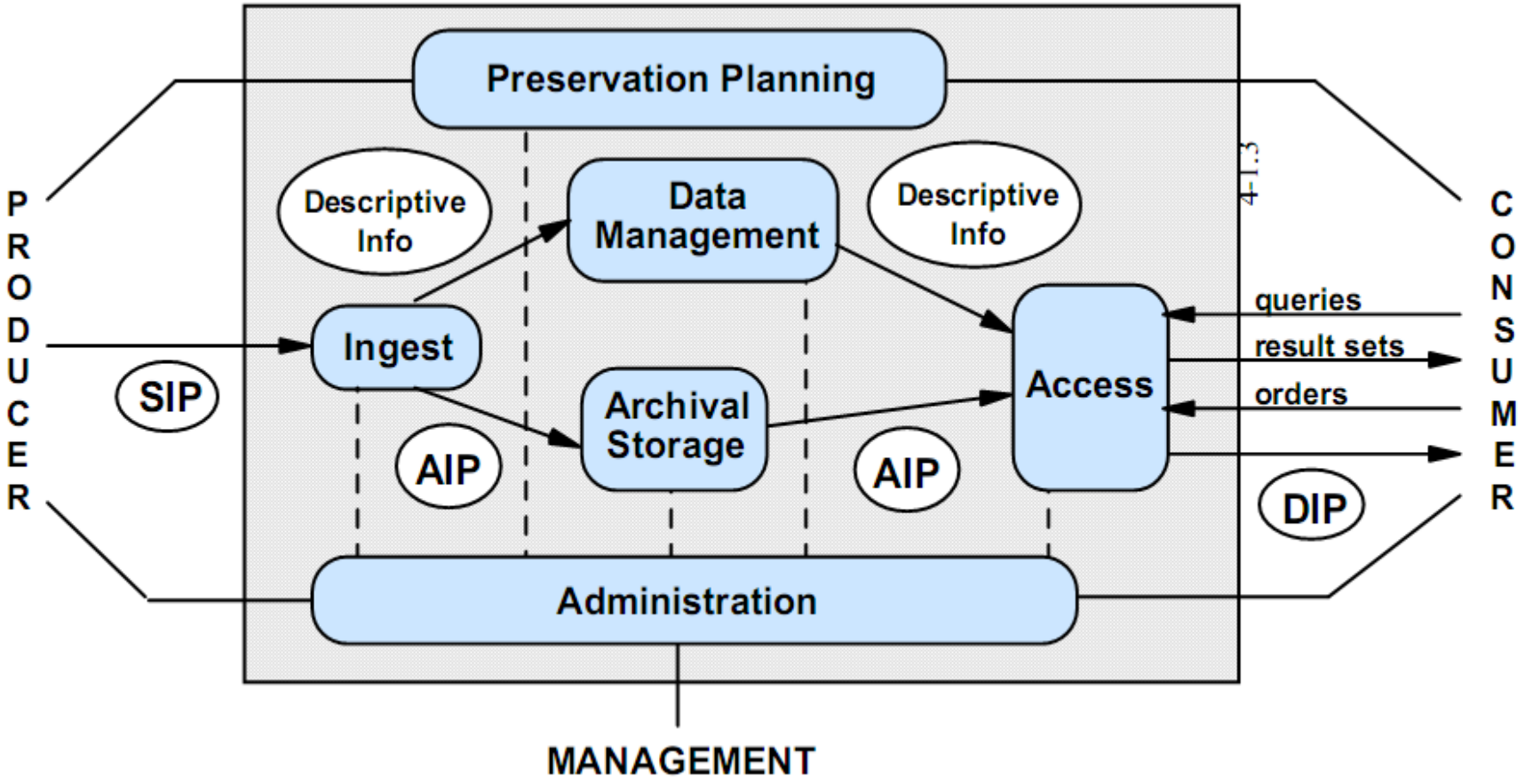
The OAIS model defines six functional entities that synthesis the most essential activities within a digital archive: ingest, preservation planning, archival storage, data management, administration, and access.

Digital Preservation – the Digital Archive Lifecycle

The OAIS model looks at data stored in the digital archive as a fluid object that can (co-)exist as three types of information packages:

- submission (SIP) is used to transfer data from the producer to the archive,
- archival (AIP) is used for the archival storage and preservation,
- dissemination (DIP) is used within the access function when consumers request archived materials.

The OAIS model



OAIS Model and the DCH-RP Project

The traditional OAIS model is no longer sufficient.

When we bring distributed digital preservation (e-Infrastructure and the Cloud) into the model, there needs to be something else in the model, too.

The DCH-RP project is working towards identifying what this “something” should be!

Thank you for your attention!

Börje Justrell