Standard Models and Formats for Digital Preservation

Börje Justrell
Director
Swedish National Archives

Table of Content

- Definitions
- Strategies
- The Digital Archive Lifecycle
- The Growing Tool Box
- Control of File Formats

Digital preservation - Definition

- 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 project it is often said that preservation is on hand when digital objects are accessible and usable to future users.

Digital preservation - Definition

- 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' 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*".

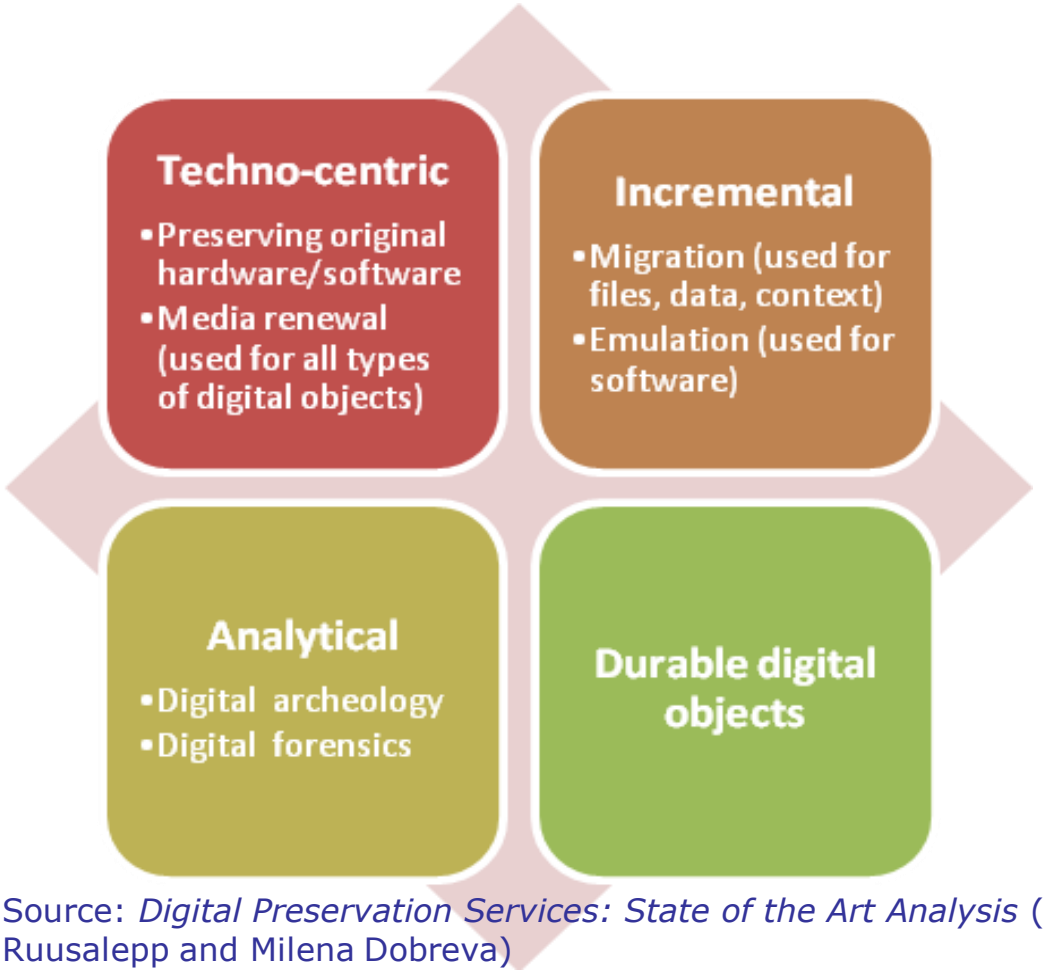
  http://www.dcc.ac.uk/digital-curation/what-digital-curation

Digital Preservation - Strategies

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

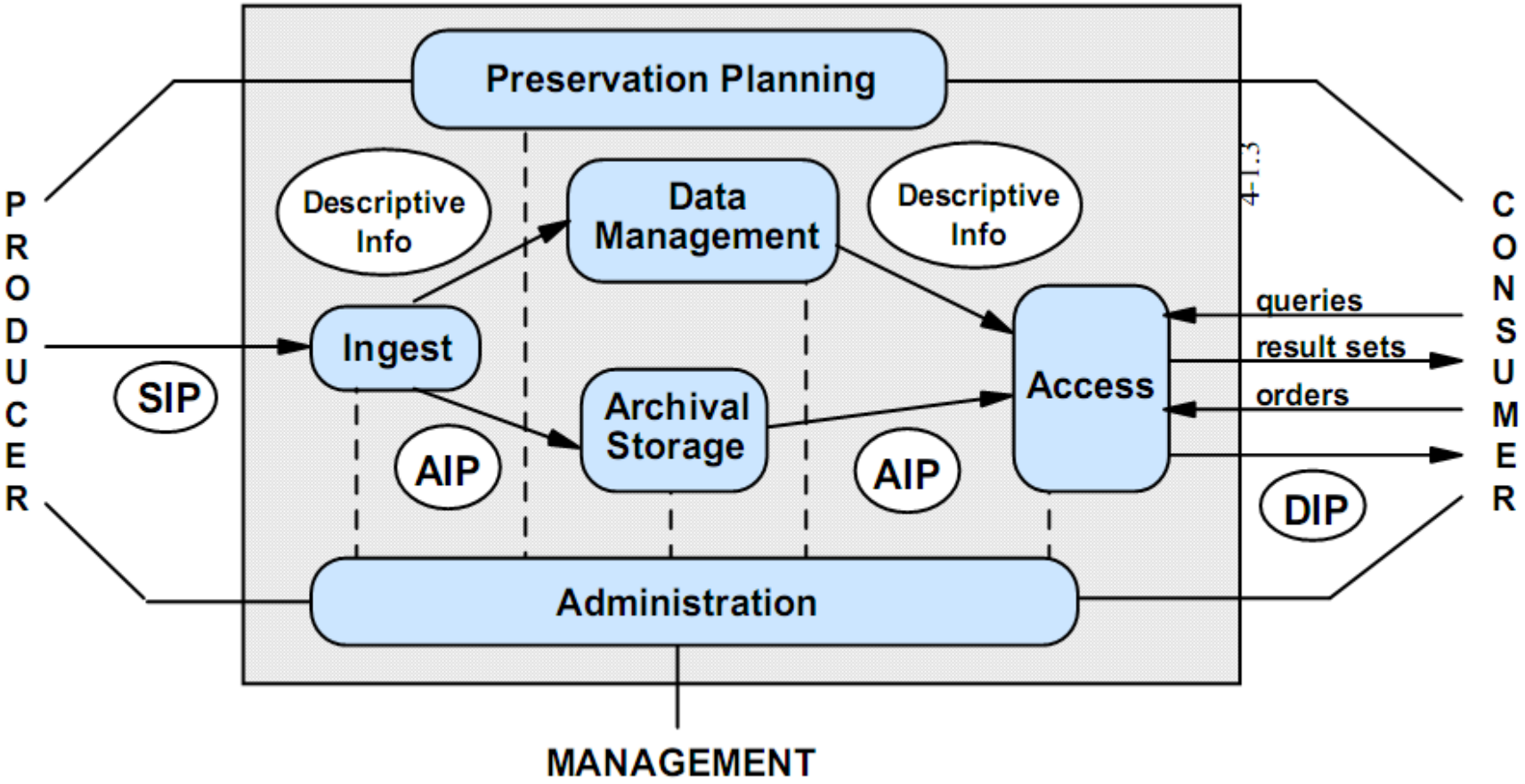
Digital Preservation – the Digital Archive Lifecycle

- 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 14721:2003 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

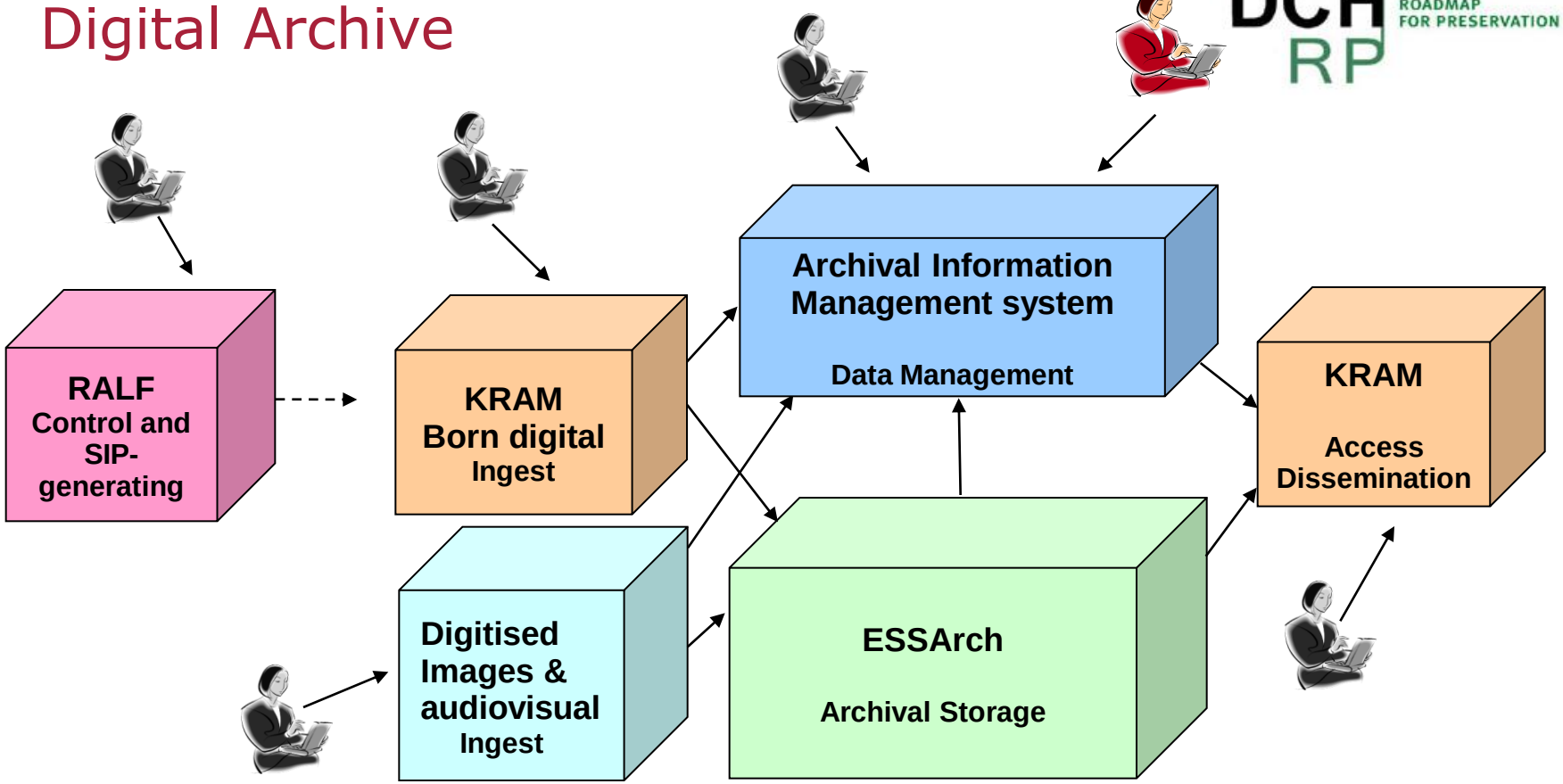


Digital Preservation – the Digital Archive Lifecycle

- As a reference model, the OAIS standard does not imply a specific design or formal method of implementation.

- Instead, it is left to the readers to develop their own implementation by analysing existing business processes and matching them to OAIS functions.

The Swedish National Archives´ Digital Archive



Digital Preservation – the Growing Tool Box

A vital pre-requisite when the cultural heritage community is to promote interoperability, encourage widespread access and control costs in its digital preservation programmes is a wide use of relevant and open standards.

- Descriptive metadata standards for a number of domains – for example, museums, archives, libraries, etc. – were surveyed in the *Athena and Linked Heritage projects* and elsewhere.
- The *Athena* project also looked at the mainstay standards for various categories of digital content commonly encountered in cultural heritage preservation.

Digital Preservation – the Growing Tool Box

- It important to understand and communicate the license agreements and terms of usage associated with digital resources, whether these are "born digital" or digitised representations of cultural heritage artefacts; the *Linked Heritage* project investigated this topic
- *Linked Heritage* also listed and explained a set of basic standards, whose use is widespread across digital preservation and other internet-enabled areas
- Standards for Search, Retrieval, and Harvesting; the *Athena project* identified three protocols of particular importance in this area.

Digital Preservation – Control of File Formats

- Metadata is normally stored in XML and specified in different schemas controlled by the community of professional curators through international organisations.
- Data content are normally stored in specific file formats for documents, images, sound, video etc. These files are usually produced by software from different vendors.
- Even if files are in standard formats, the implementation of standards cannot be guaranteed. The software implementing standards for the production of the electronic files is not in control neither by the institutions that produces them nor by the memory institutions.

Digital Preservation – Control of File Formats

- File formats conformance tests of transfers are done, but are today not totally reliable. Different software for testing could end up in different results
- This poses problems in long-term preservation. Data objects meant for preservation, passing through an uncontrolled generative process, can jeopardise the whole preservation exercise.
- A new EU project *PREFORMA* will look into that, with the overall intention to research critical factors in the quality of standard implementation in order to establish a long-term sustainable ecosystem around developed tools with a variety of stakeholder groups. These tools should be innovative and provide a reference implementation of the most common file format standards for the assessment of the collections to be archived and for the correction of the collections.

DIGITAL CULTURAL HERITAGE ROADMAP FOR PRESERVATION