OAIS reference model and digital archives in practice

Raivo Ruusalepp
Estonian Business Archives
ERM vs Digital Archive

ERM
- Established functional requirements
- Good guidance for implementation
- Metadata standards
- Extensive (best) practice

Digital archive
- OAIS reference model + a few others
- Lots of literature on digital preservation
- Metadata: OAIS-based, ISAD(G), etc.
- Little practice: tests and research projects
Acknowledgements

- JISC – Institutional repository infrastructure development programme
- “Assessment of UK Data Archive and The National Archives compliance with OAIS/METS” (2004-2005)
- UK National Archives
  - Matt Palmer
- UK Data Archive
  - Hilary Beedham
  - Julie Missen
Introduction to UKDA

- Established in 1967
- Acquires, preserves and disseminates social science and humanities datasets
- Has built its digital archive and storage system over years, following mostly its own and its users’ needs
- Member of an international network of data archives
- Legal place of deposit in co-operation with the TNA as of 2005
Introduction to TNA

- Formed as an amalgamation of the Public Record Office and the Historical Manuscripts Commission
- Digital Preservation Department founded in 2001
- Built the Digital Archive software system for preserving born digital government records in 2003
- Also has a Web Archive for government web sites and the PRONOM registry
Project Overview

- Map the current preservation systems of the two organisations against those in the OAIS Reference Model
- Test the compliance with the OAIS model
- Assess how the operational structure can be informed by OAIS (and vice versa)
- Explore the potential for interaction between existing metadata standards and METS
No fixed methodology for testing for OAIS compliance

Easy to meet the OAIS ‘mandatory requirements’ — mostly policy level

To test the compliance with OAIS functional entities, workflow analysis at both archives, which was then compared to the OAIS functions

Produced workflow diagrams and mappings to OAIS functional entities
Outcomes: functional model

- There is a good match between the reference model and the practice.
- OAIS is only a reference model, not an implementation guide - it does not offer solutions for ‘how things should be done’, but does provide a list of ‘what needs to be done’.
- Sometimes the general level of the model is not helpful for understanding the logic behind functions (e.g., in the ingest, administration, management functions).
Some OAIS functional entities have very little knowledge “built into them” – they perform a very limited set of actions and rarely take decisions.

In a real archive, systems and workflow are designed so that decisions are taken closer to where they are needed, sometimes it is a single person who makes the decision, performs the task and tests its outcome.

The OAIS does not preclude this, but does not recommend it either.
Outcomes: functional model

- It is a concern that the model is not obviously scalable. A possible solution is the development of an ‘OAIS Lite’ version to help smaller archives gain value from it as a reference model.

- The task of mapping the UKDA and TNA functions to the OAIS reference model would have been easier if there had been a guideline or manual to aid and inform the process.
More difficult to match as the metadata schemas in use do not decompose naturally into the separate information packages of SIP, AIP and DIP – metadata encompasses the entire workflow of managing the digital objects.

OAIS model has very little to say about the structure of DIPs and does not attempt to model the ongoing management metadata that are required for a functioning archive.
Outcomes: Ingest

- Workflow compared with both the OAIS and the “Producer-Archive Interface Methodology Abstract Standard” (draft)
- Identified overlap between ingest and pre-ingest activities which needs to be considered in the context of the OAIS
- Some discrepancies found:
  - transfer arrangements
  - validation levels
  - metadata validation and creation
  - DIP creation
The whole ingest process needs more elaboration from the ‘records creator’ point of view.

The pre-ingest activities should be included in the OAIS model itself, rather than in a separate standard(s) – pre-ingest activities are not without cost and they ensure quality, understandability and accessibility of information packages.

Quality assurance should receive more attention at the ingest stage.
The OAIS model points to the strong link between the user community and the way the material in the archive should be described and preserved.

In reality it is often difficult to limit the user communities to groups as narrow as the OAIS standard examples.

The OAIS has an inbuilt limitation of overly assuming an identifiable and relatively homogeneous consumer community.
Analysis of the access function demonstrated that the UKDA and TNA have various limitations on access to their collections and each has its own mechanisms for controlling it (e.g., user agreements, FoI requests, unrestricted access), which are not part of the OAIS. The DIPs are not created ‘on the fly’ from AIPs, as suggested by the OAIS, but converted from SIPs at the ingest stage and stored separately from AIPs.
Recommendations: Dissemination

- Consideration should be given to the relationship between AIPs and the created DIPs within the model.
- Creating the DIP at the time of ingest means that there is access to a richer information source (the depositor) for the staff carrying out the process.
- The mechanisms of attaching different metadata to DIPs requested by different designated communities remains unclear from the OAIS standard.
Utility of the OAIS language as a means of communication between the UKDA and TNA was noticeable. The working terminologies of each organisation is different but as dialogue progressed each found that it was useful and less confusing to use OAIS terminology as a common means of communication.

Recommendation: encourage use of OAIS terminology.
Outcomes

Organisations wishing to undertake a similar exercise to this need to be aware that it is resource intensive and requires the full support of senior management to enable it to be completed successfully.

Recommendation: do not undertake the exercise without plenty resources and support from senior management.

There may be benefits in providing a guide, manual or self-testing toolkit for mapping functions to the model.
Report is available from...


or contact: publicity@esds.ac.uk for hard copies
Questions?

Raivo Ruusalepp
raivo@eba.ee
tel: +372 559 456 13
www.eba.ee