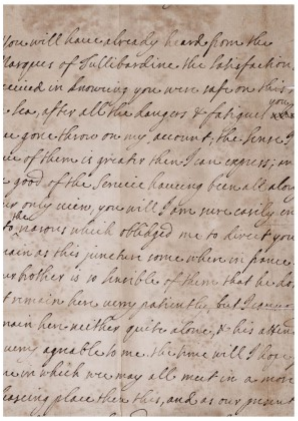
Introduction to Digital Preservation

DigitalPreservationCoalition

Traditional Media



- Robust
- Tangible
- Independently understandable
- Experienced in assigning value

- Traditional objects are generally quite robust
- They are tangible, we can hold them in our hands
- Are generally independently understandable (if you speak the language they are written in…..)
- We are quite experienced in understanding their worth and assigning value to such objects

Digital Information

- Ephemeral
- Obsolescence
  - Media
  - Formats
  - Documentation
- Rights
  - Copyright
  - Moral
  - Data Protection
- New skills and solutions
- New benefits

- Digital objects are ephemeral by their very nature
- They very susceptible to obsolescence as they are entirely dependent on the media they are stored on, the accessibility of their file format and often require documentation to use and understand them
- Managing issues such as rights can also be much more difficult, from protecting copyright to ensuring personal data is protected
- They require us to gain new skills to care for them, or for us to work with new groups of colleagues with different skills groups (particularly IT specialists)
- But they do also bring a whole host of new benefits, in particular the ability to make content accessible to users.

What's the Problem?

- Digital data (images, documents etc.) have value
- They can create opportunities

...but...

- Access depends on software hardware and people
- Technology and people change

...therefore...

- Technology can create barriers to reuse
- Managing data in the long term protects and creates opportunities

In a Bit More Detail….

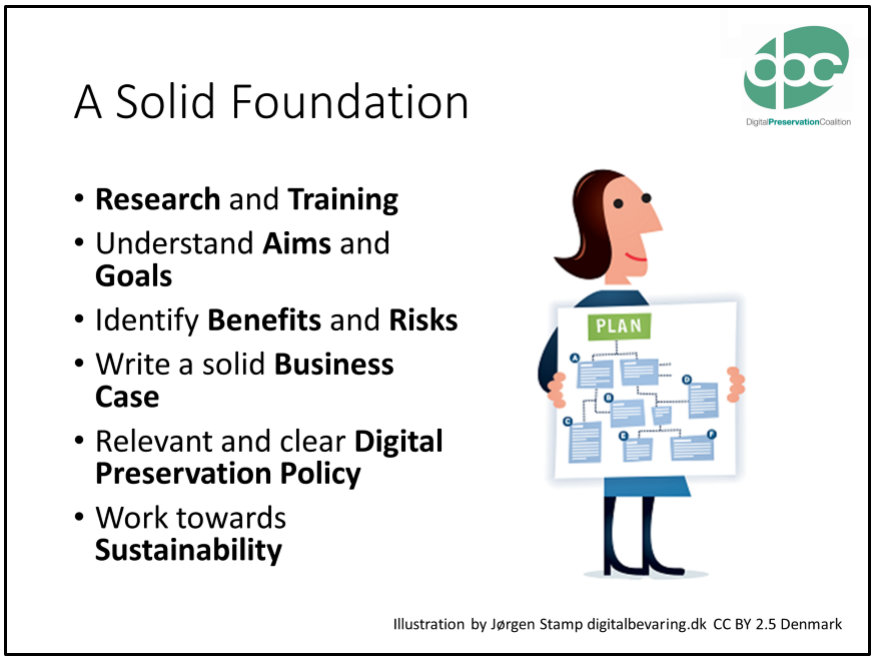
- Seven basic challenges
- Ways to address them:
  - Tools
  - Services
  - Standards
- Fancy words and acronym bingo

Illustration by Jørgen Stamp digitalbevaring.dk CC BY 2.5 Denmark

Challenge One

Starting a digital preservation programme seems expensive and it's difficult to know where to start…

…but there's plenty of help on how to build a business case and tools to aid with planning



Research and **Training** – don't get bogged down, there's a lot of info out there so don't try to read anything. The Digital Preservation Handbook is a good place to start, giving a good overview of Digital Preservation issues and providing sign posts to useful resources and case studies: http://handbook.dpconline.org

Understand **Aims** and **Goals –** Clear articulating these will help shape and solidify future work, using a maturity model can help figure these out.

Identify **Benefits** and **Risks –** Important to clarify the risks your organisation will face if digital preservation activities are not undertaken, and on the flip side, the benefits that will accrue from good DP practices

Write a solid **Business Case –** Important to clearly articulate why DP is important in a format familiar to those in management

Relevant and clear **Digital Preservation Policy –** Provides a fundamental reference point for all future DP work

Work towards **Sustainability** – embedding DP in the every day workflows of your organisation should be the ultimate aim

Challenge Two

Digital preservation systems can be complex and are subject to the same obsolescence as the objects they safeguard…

… so we need systems which are resilient, based on standards and which can be tested on an on-going basis

OAIS Functional Model

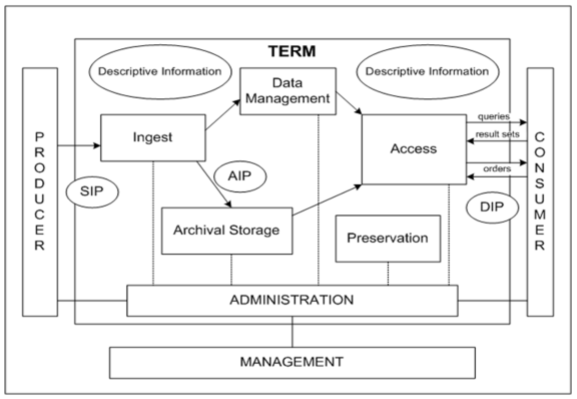


Fig. 1. Major functions of the OAIS Reference Model from Consultative Committee for Space Data Systems (CCSDS), CCSDS 650.0-W-1, Producer-Archive Interface Methodology Abstract Standard, (OAIS), White Book, Issue 1, Draft Recommendation for Space Data System Standards.

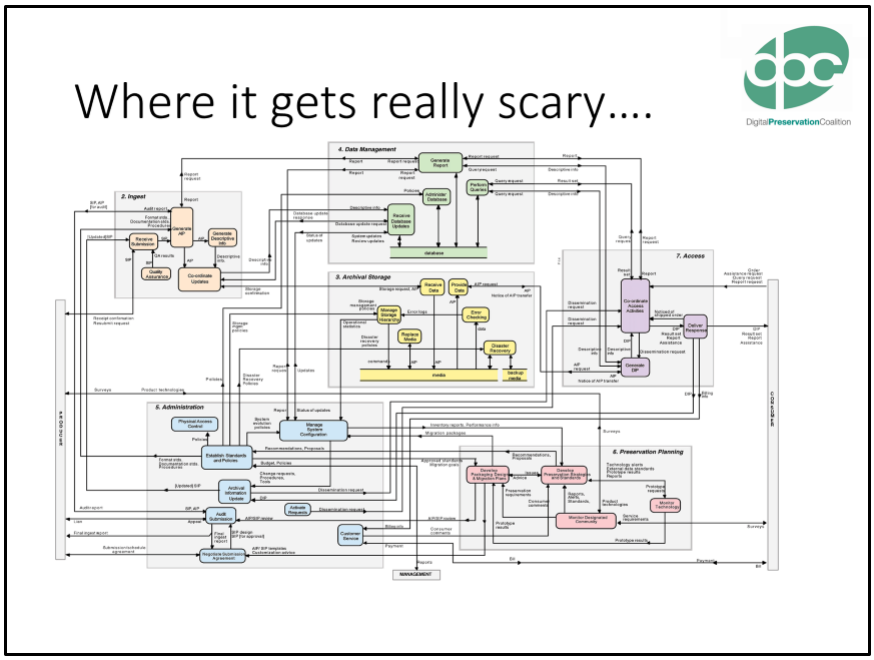
While it is far from perfect the Open Archival Information System model is one of the keystones of Digital Preservation. In particular it provides much of the terminology used within the field. This diagram represents it's functional model at the highest level. As well as the key functions of an OAIS that it shows (such as Ingest, Preservation and Access) it also includes various information packages. These information packages contain the digital material to be preserved along with its accompanying metadata and within OAIS these exist in 3 different forms across the lifecycle:

1. The Submission Information Package
2. The Archive Information Package
3. The Dissemination Information Package

To accompany this functional model the OAIS also describes an information model that lays out what types of metadata (specifically called Representation Information in OAIS) should be included in the information packages to facilitate preservation.

OAIS is an internal standard originally developed by the space data discipline and our Executive Director at the DPC likes to say by comparing your library or archive to this,

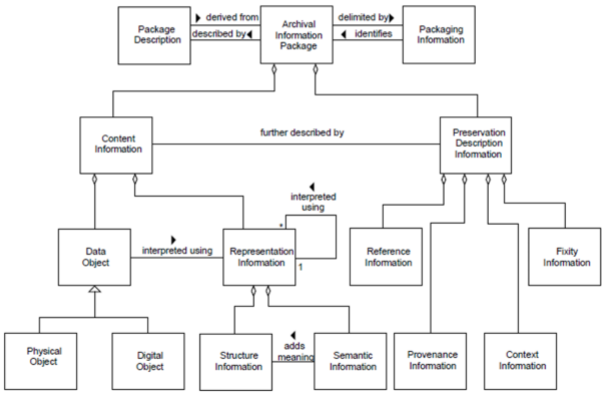
you’re comparing yourselves to NASA. This makes a little more sense when you start delving a little further into the standard and things like the full functional model……



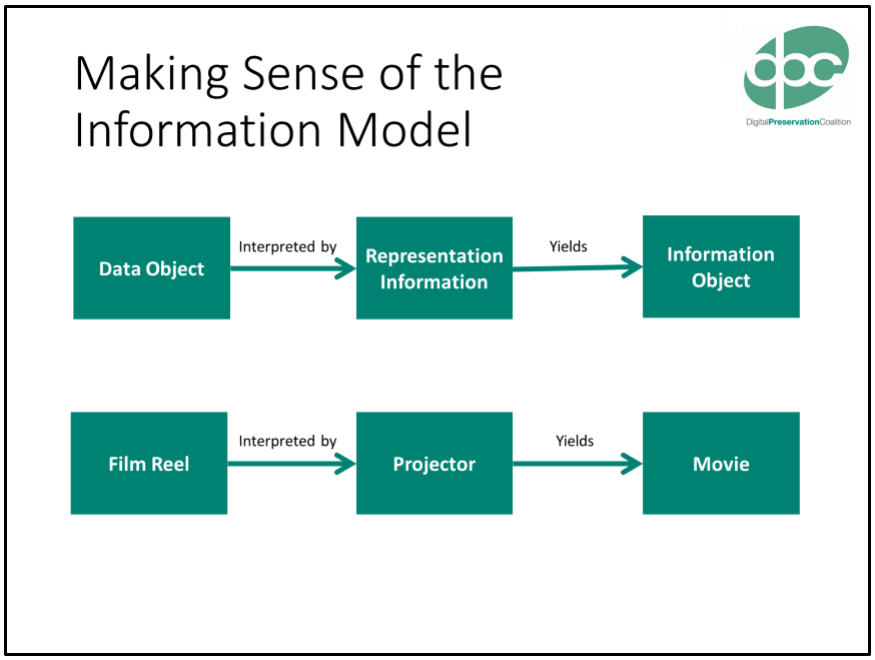
…..although there are more accessible ways in. The DPC has published a very accessible Tech Watch Report on OAIS for which there is also an accompanying webinar you can watch on our website. The standard is also currently up for review and we'd encourage anyone with an interest to take part, more details are available on the DPC site.

Brian Lavoie's Tech Watch Report: http://dx.doi.org/10.7207/twr14-02

I'm More Terrified Of.....

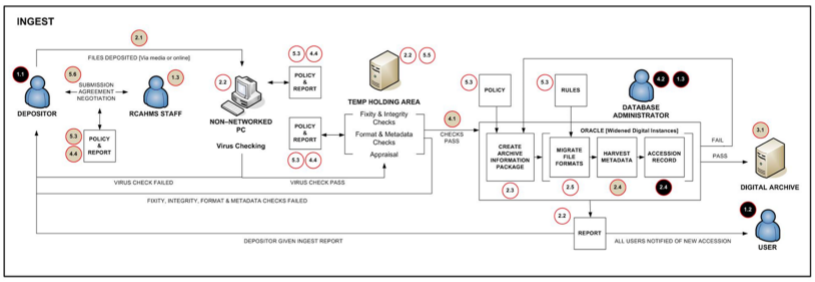


The OAIS Information Model details a structure that includes the digital object itself, the representation information required to render the object on screen and the preservation description information necessary for managing the object.



Representation information is perhaps the most difficult part of the information model to grasp. It can be useful to consider in relation an object like a reel of film. To show this onscreen a film projector is required. The projector fulfilling the role of the representation information. In the digital world representation information can be anything from the file format to detailed information about the file and the software and operating system environment in which to access it (in some cases perhaps even the software itself).

But still useful….

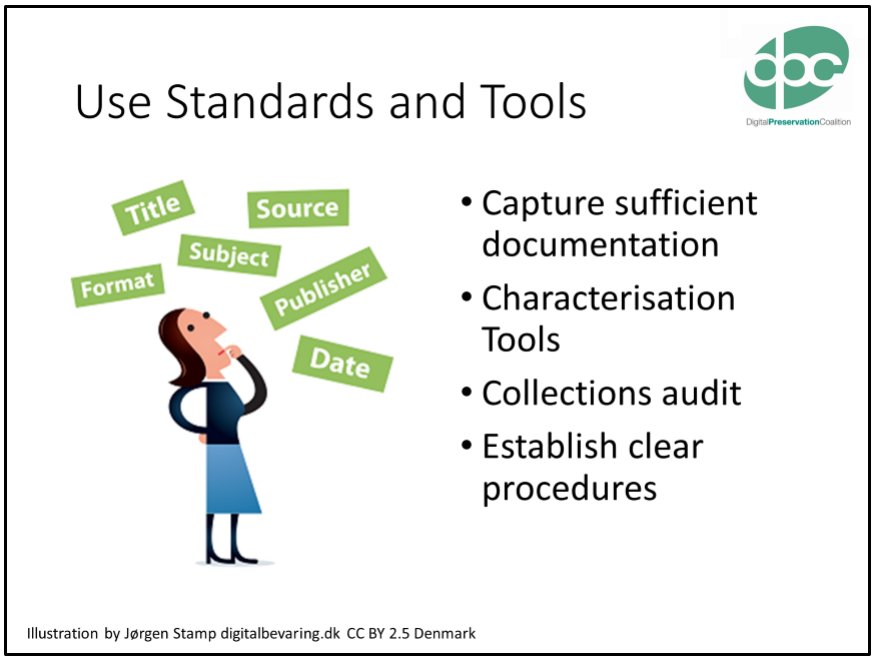


This a diagram we produced at my previous place of work, the Royal Commission on the Ancient and Historical Monuments of Scotland. The aim was to start building what a digital preservation workflow might look like then carry-out a gap analysis comparing the current systems with the OAIS functional model. The black circles are people, systems or process that were in place and meeting standards, the tan circles are those which were in place but not sufficiently developed/meeting requirements and the white circles where things that did not currently exist. It provided a very useful, clear visual aid for presenting to managers as they could clearly see the large gaps in our current systems and processes.

Challenge Three

Access and long term use depends on the constant configuration of hardware, software, data and the capacity of the operator….

… so we need to capture information on this configuration and use it to enable access.



Capturing sufficient documentation is essential and you may want to consider using established Metadata Standards

PREMIS Data Dictionary

http://www.loc.gov/standards/premis/

METS for wrapping data

http://www.loc.gov/standards/mets/

Collecting information can also be done using software such as characterisation tools or via more traditional methods such as a collections audit.

Once you have gathered this information it is important to establish clear procedures for updating it on a regular basis.

Challenge Four

Technology continues to change creating the conditions for obsolescence…

… so we need to plan and take action accordingly, expecting that our current plans may need to change.

Active Preservation

- OAIS Preservation Planning functions
- Preservation Planning tools
- Different Approaches
  - Migration
  - Emulation
  - Hardware Preservation
  - Virtualisation
  - Etc…….

Illustration by Jørgen Stamp digitalbevaring.dk CC BY 2.5 Denmark

- Migration – from one file format to another, or one version of a format to a newer version (i.e. .doc to .docx)
- Emulation – creating a programme that emulates the original hardware or software environment the object was created in. Allows an authentic experience although they are rarely perfectly

Challenge Five

Storage media fail, have a short life and storage devices are subject to obsolescence.

… so we need a storage strategy which includes error checking and refreshment

Storage and Refreshment

Different Solutions:

- Multiple media
- Controlled storage
- Self-reporting media
- Lots Of Copies Keeps Stuff Safe
- Cloud storage

Beware: proliferation can become a problem

Illustration by Jørgen Stamp digitalbevaring.dk CC BY 2.5 Denmark

Be aware of the storage plan for your organisation and make sure digital material well identified and is stored where it is secure and will be backed-up.

Lots Of Copies Keeps Stuff Safe

http://www.lockss.org/

Cloud storage

http://www.archives.org.uk/images/documents/Cloud_computing_report_final-1.pdf

Challenge Six

Digital resources are intolerant of gaps in preservation…

…We need to act early and we need to act on an on-going basis.

Ongoing Preservation

Plan your responses:

- Intervene early in lifecycle
- Build in Sustainability
- Risk management approach

Illustration by Jørgen Stamp digitalbevaring.dk CC BY 2.5 Denmark

Plan your responses:
Intervene early in lifecycle
Build in Sustainability
    Transferable Archive Information Packages
    E-ARK - http://www.eark-project.com/
Risk management approach

Challenge Seven

Resources can be corrupted or tampered without trace…

…Need to carry out fixity and authenticity checks on your data

Fixity and Authenticity

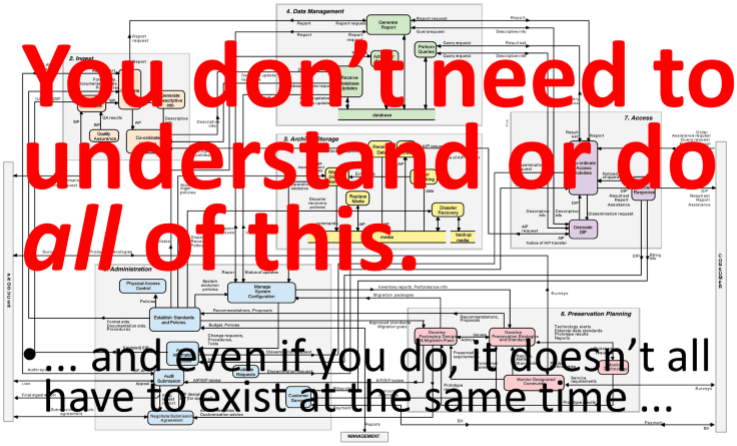
A variety of solutions:

- Checksums
- Forensic tools
- Authenticity Evidence Records
- Data security protocols
  - (ISO 27000 series)

Illustration by Jørgen Stamp digitalbevaring.dk CC BY 2.5 Denmark

Fixity and Checksums - http://handbook.dpconline.org/technical-solutions-and-tools/fixity-and-checksums
Digital Forensics and Preservation Report - http://dx.doi.org/10.7207/twr12-03
Authenticity from the APARSEN Project - http://www.alliancepermanentaccess.org/wp-content/uploads/sites/7/downloads/2014/06/APARSEN-REP-D24_1-01-2_5_incURN.pdf

But ‘Don’t Panic’…..



You don’t need to understand or do *all* of this.

- … and even if you do, it doesn’t all have to exist at the same time …

The Reality…

- It won't go away……

And……

- It won't fix itself……

But……

- You already have many of the skills you need!

Illustration by Jørgen Stamp digitalbevaring.dk CC BY 2.5 Denmark