



Digital **Preservation** Coalition



William Kilbride
@williamkilbride

Getting Started in Digital Preservation: what do I need to know?

It won't go away
It won't do itself

You already have many of
the skills you need!

Digital preservation typically makes bleak reading ...



When asked about how long their digital resources would be available for, JISC-funded projects said ...

‘In perpetuity’

‘Indefinitely’

‘50 years’

‘10 years then elsewhere’

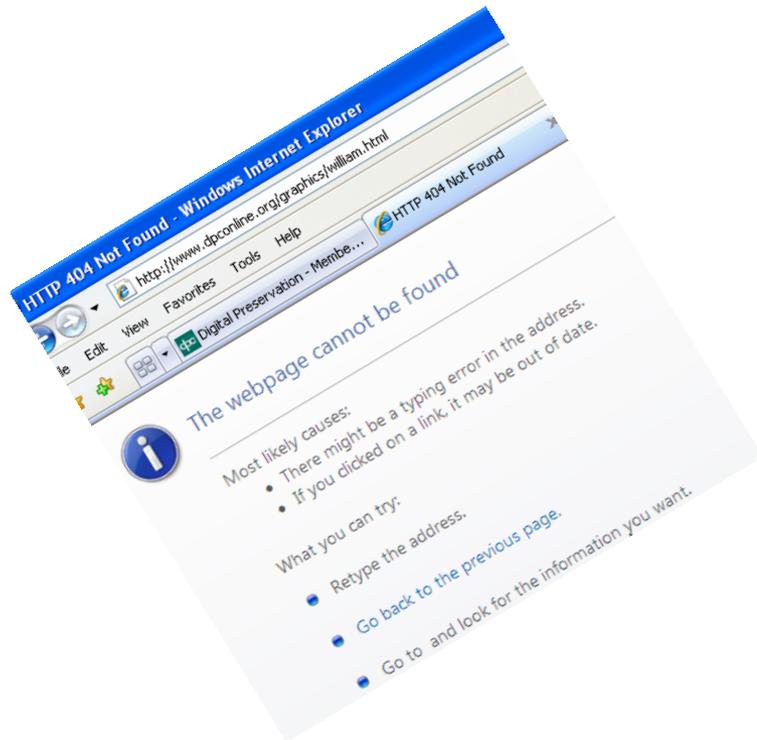
‘until 2014’

‘forever or for three years’

DPC/Portico/ULCC 2010



Digital preservation typically makes bleak reading 2



“...of all the web links cited in answers to parliamentary questions 1997-2006, 40 percent are now broken”
(Spencer et al 2009)

Digital preservation typically makes bleak reading 3



<Enter your details here>

.....

.....

.....



What's the problem?

- Digital data (images, documents etc) have value

- They create opportunities

...but...

- Access depends on software hardware and people

- Technology and people change

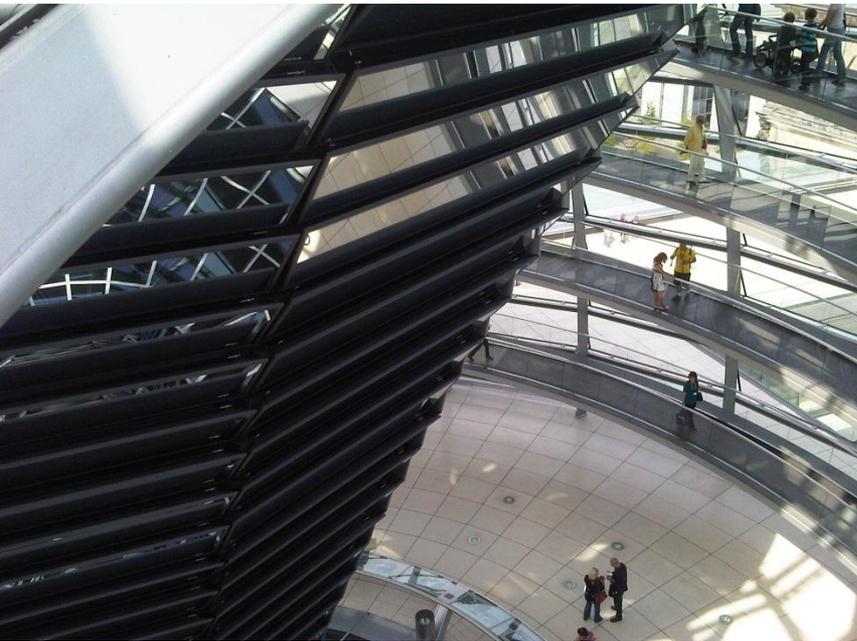
...therefore...

- Technology can create barriers to reuse

- So, managing data in the long term protects and creates opportunities

- Not something we do for its own sake

We do preservation because we want to be:



1. Transparent

e.g. Data Protection, Freedom of Information ... childcare, human tissue

2. Safer

e.g. preparedness, detection, disaster, recovery, audit

3. Smarter

e.g. scientific value, access to heritage, value of social knowledge

4. Wealthier

e.g. efficient business, management of IP, employment, planning, creative

5. Healthier

e.g. managed life history, research and safe innovation

6. Greener

e.g. evidence-based policy development, efficient data retention

And because of



1. Legal Compliance

e.g. Sarbanes-Oxley, Data Protection

2. Regulatory Compliance

e.g. power generation, aviation, banking

3. Legal protection

e.g. patents, mis-selling, detection, audit

4. Unanticipated exploitation

e.g. petro-chemical, music,
pharmaceuticals

5. Business Continuity and improvement

e.g. product recall, disaster recovery

6. Business Value

e.g. getting the right information to the
right people at the right time in a format
they can use



Digital **Preservation** Coalition

Digital preservation is not just about 'data':

Digital preservation is not just about 'access':

Digital preservation is not just about 'risk':

it's about

people and

opportunity

A bright yellow, multi-pointed starburst graphic with a black outline, positioned to the right of the word 'and'.

***Start
Here!***



Digital **Preservation** Coalition



Four basic responses to long-term preservation:

- *Some fancy words*



1. Migration

Changing the format of a file to ensure the information content can be read

2. Emulation

Intervening in the operating system to ensure that old software can function and information content can be read

3. Hardware preservation

Maintaining access to data and processes by maintaining the physical computing environment including hardware and peripherals.

4. etc

Research and development field, new solutions and new approaches continue to emerge, eg virtualisation for preservation



Six basic challenges and how to address them

- *Tools*
- *Services*
- *More Fancy words*



Challenge 1:

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 this configuration and use it to enable access.

Metadata, documentation, representation information

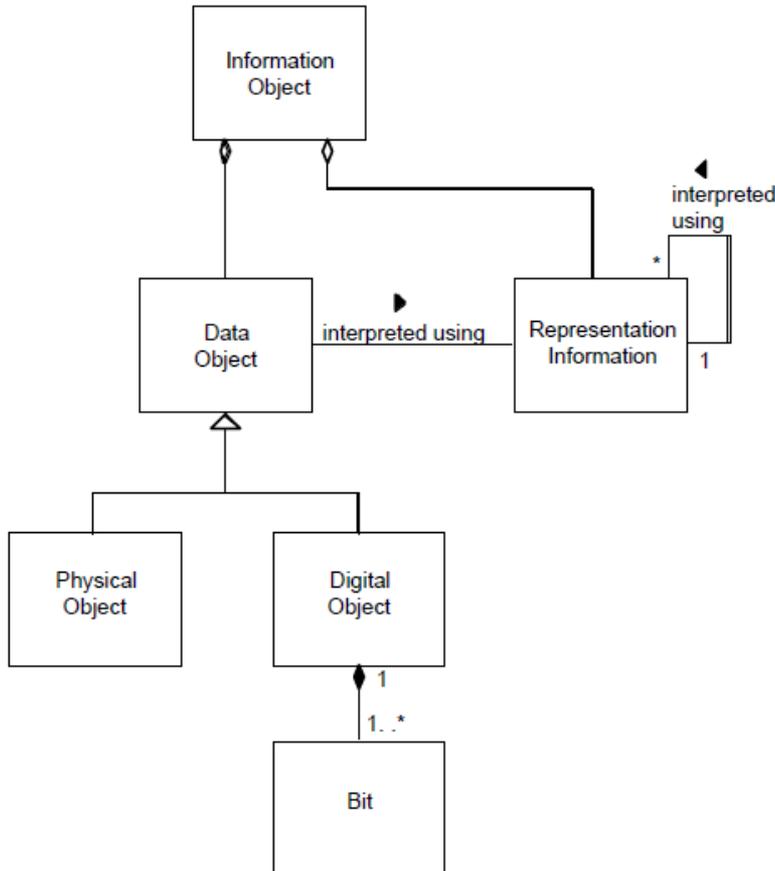


Challenge 1:

Metadata, documentation, representation information, archival description ...

Various answers:

- *OAIS Information Model*
- *PREMIS Data Dictionary*
- *METS for wrapping data*
- *Registry services (e.g TOTEM, PRONOM etc)*

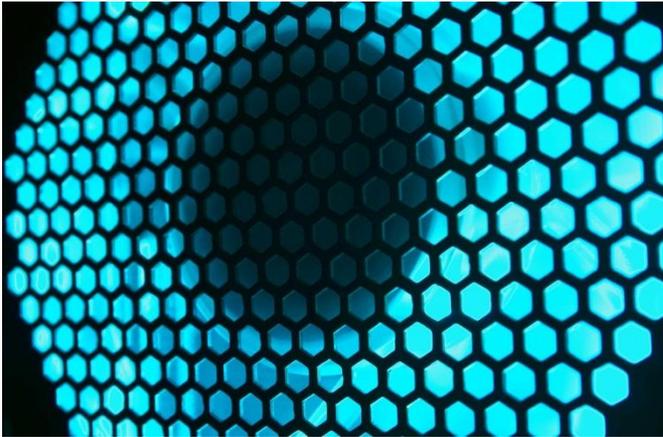




Digital Preservation Coalition

Challenge 2

Technology continues to change creating the conditions for obsolescence.

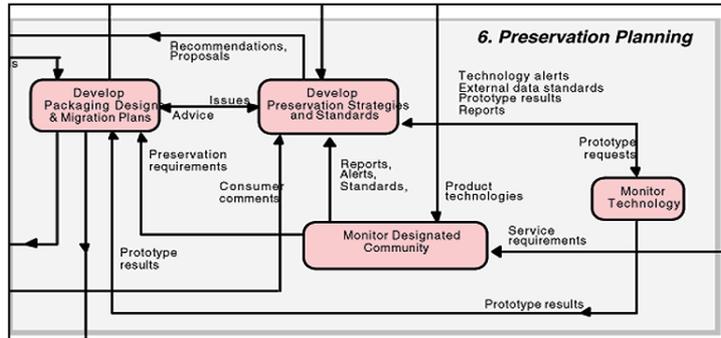


... so we need to plan accordingly, expecting that our current plans may need to change.



Challenge 2

Planning and learning



Be a learning institution Different solutions:

- *OAIS Planning Functions*
- *PLATO: Tool – Library and Methodology*
- *PLANETS Testbed*
- *Audit and certification: DANS, TRAC / 16363, DIN 31644*



[Introduction](#) | [Documentation](#) | [Case Studies](#) | [Events](#) | [History](#)

Introduction

What is Plato?

The fast changes of technologies in today's information landscape have considerably shortened the lifespan of digital objects. Digital preservation has become a pressing challenge. Different strategies such as migration and emulation have been proposed; however, the decision for a specific tool e.g. for format migration or an emulator is very complex. The process of evaluating potential solutions against specific requirements and building a plan for preserving a given set of objects is called preservation planning. So far, it is a mainly manual, sometimes ad-hoc process with little or no tool support. The planning tool **Plato** is a decision support tool that implements a solid preservation planning process and integrates services for content characterisation, preservation action and automatic object comparison in a service-oriented architecture to provide maximum support for preservation planning endeavours.

This software is licensed under [Apache version 2.0](#) or later. We are going put the source code of Plato on sourceforge in the near future. In the meantime please do not hesitate to contact us at plato@ifs.tuwien.ac.at to receive a copy of the source code.

[Click here to enter Plato.](#)
(ports 8080 and 8443 must be open)





Challenge 3



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



Challenge 3

Storage and refreshment

Different Solutions:

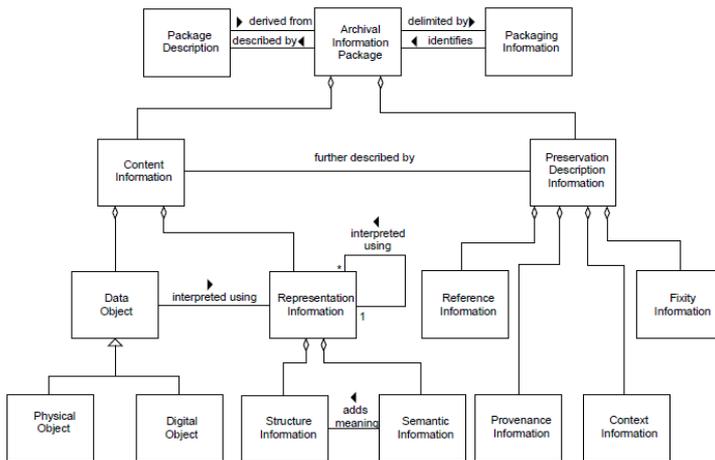
- *Multiple media*
- *Controlled storage*
- *Self reporting media*
- *Lots Of Copies Keeps Stuff Safe*
- *Cloud storage*



Beware: proliferation can become a problem

Challenge 4

Digital preservation systems are subject to the same obsolescence as the objects they safeguard.



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

Submit

Archive

Disseminate

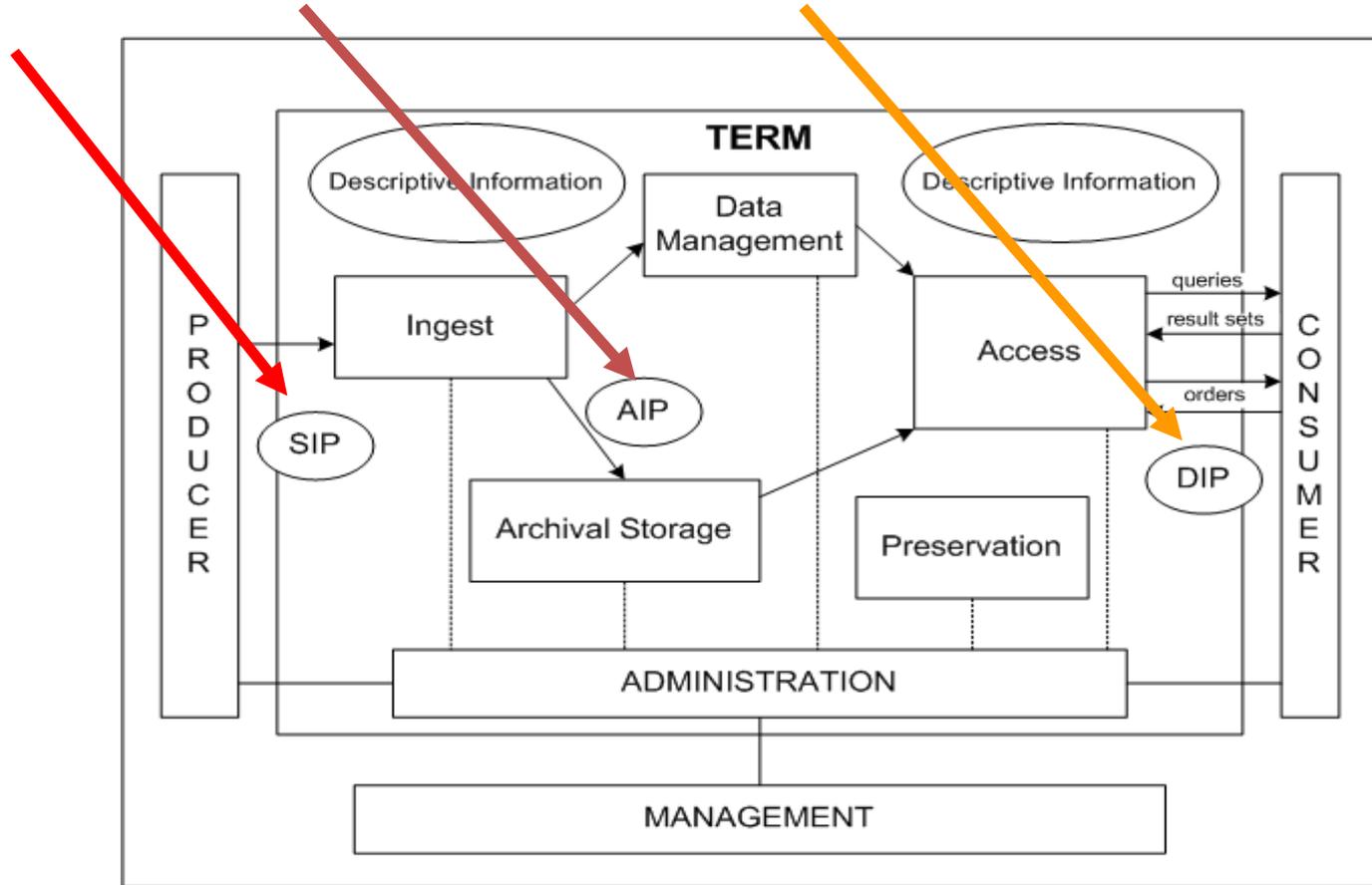


Fig. 1. Major functions of the OAIS Reference Model from Consultative Committee for Space Data Systems (CCSDS), CCSDS 650.0-W-1, Producer-Archive Interface Methodology Abstract Standard, (OAIS), White Book, Issue 1, Draft Recommendation for Space Data System Standards.

Picture from DLib



Digital**Preservation**Coalition



Courtesy NASA/JPL-Caltech

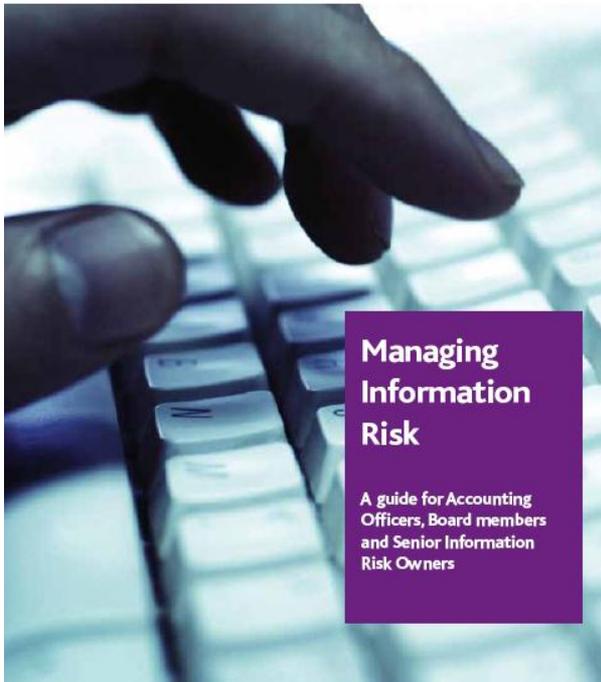
Consultative Committee on
Space Data Systems
Inadvertent comparison with
NASA
Scales up really well



Challenge 5

Digital resources are intolerant of gaps in preservation.

HM Government



We need to act early and we need to act on an on-going basis. Lends itself to risk management approaches

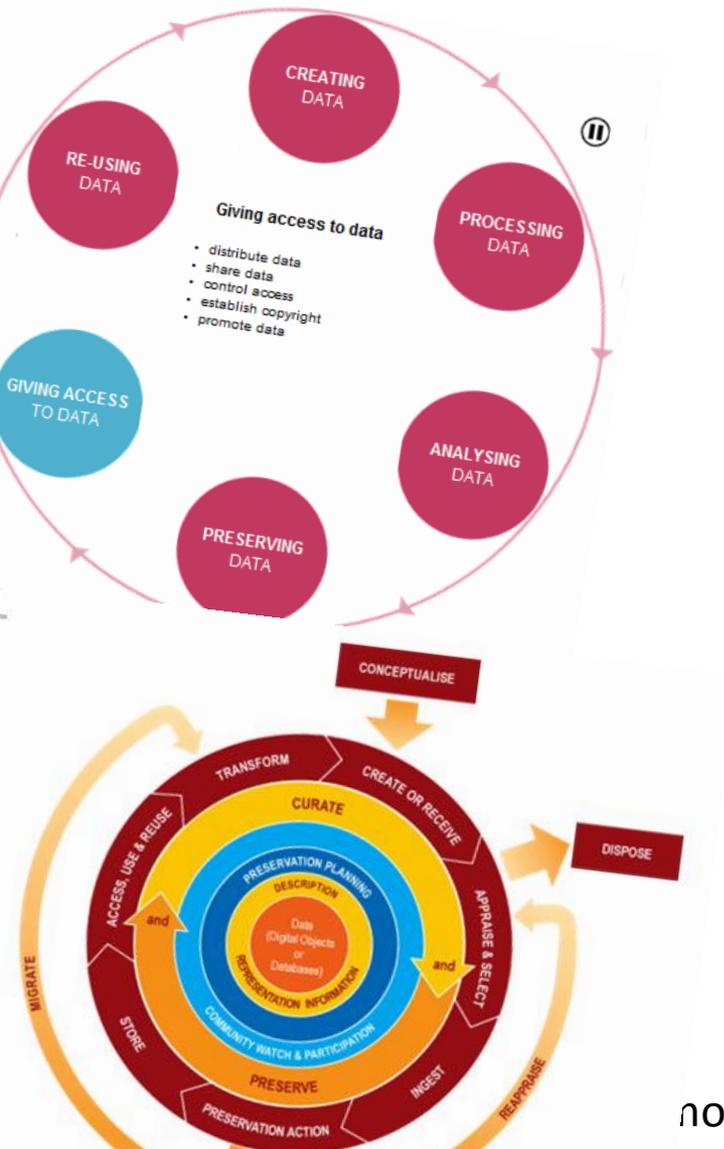


Challenge 5

On-going preservation

Different responses:

- *Intervene early in lifecycle*
- *Transferable AIPs*
- *Risk management approach e.g. DRAMBORA*
- *Monitor community*





Challenge 6

Resources can be corrupted or tampered without trace

Need to fixity and authenticity checks

Challenge 6

Fixity and authenticity



A variety of solutions:

- *Checksum*
- *Forensic tools*
- *Authenticity Evidence Records*
- *Data security protocols*



Digital **Preservation** Coalition



Getting Started in Digital Preservation: what do I need to know?

It won't go away
It won't do itself
Don't wait for perfection

William Kilbride
william @dpconline.org
@williamkilbride