DigitalPreservationCoalition

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

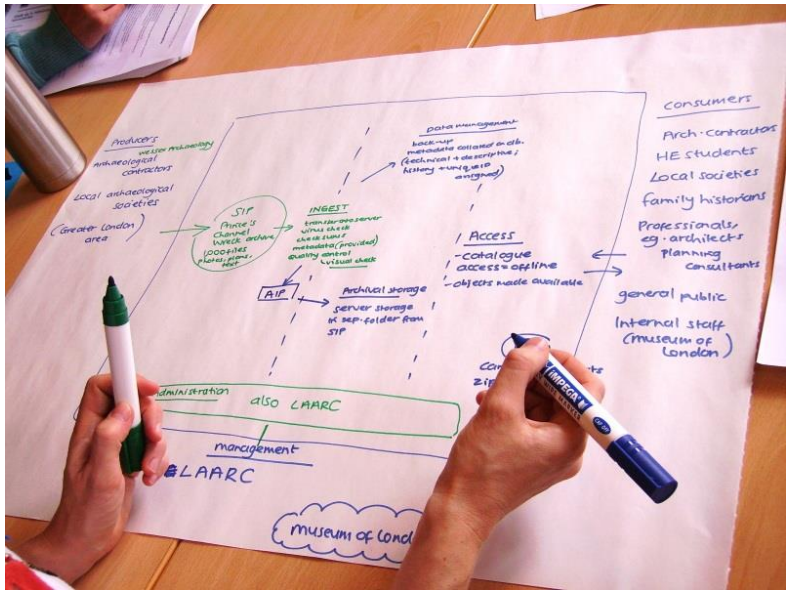
@williamkilbride
#dpc_getstarted

What has brought you here today?

Write down what you want to hear about …

… later we'll find out if we've answered the questions

DigitalPreservationCoalition



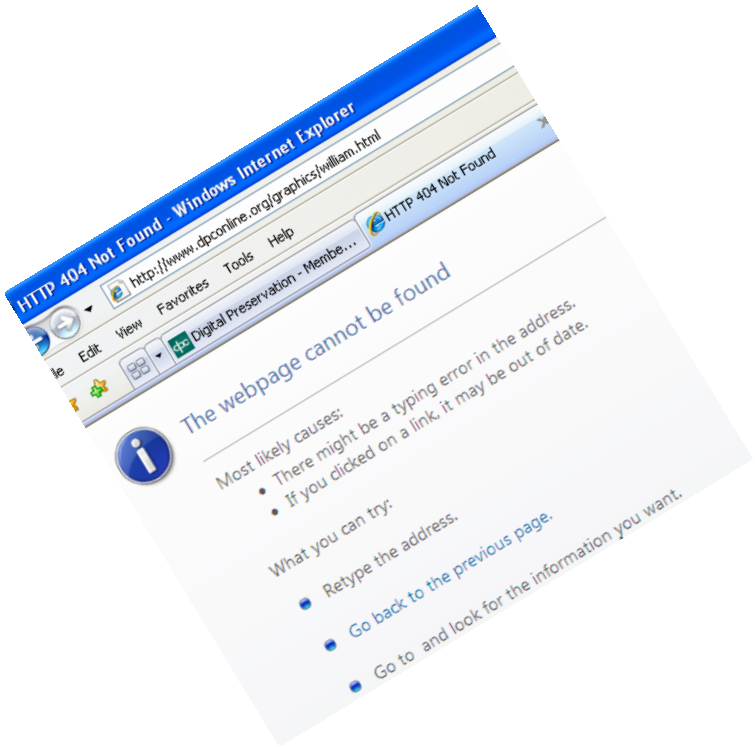
Getting Started in Digital Preservation: what do I need to know?

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

You already have many of the skills you need!

DigitalPreservationCoalition

Digital preservation typically makes bleak reading

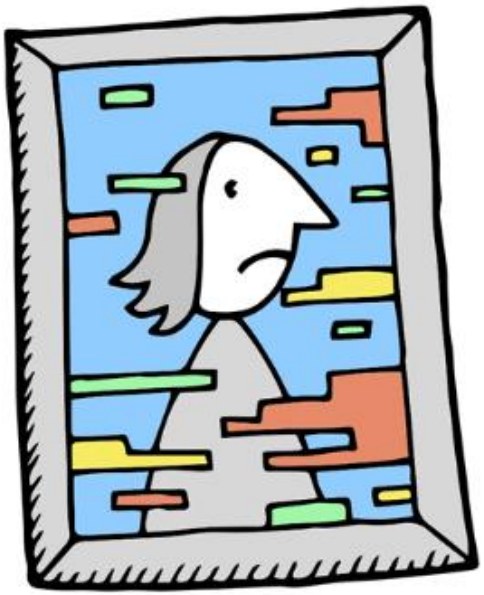


"…of all the web links cited in answers to parliamentary questions 1997-2006, 40 percent are now broken"

(Spencer et al 2009)

DigitalPreservationCoalition

Digital preservation typically makes bleak reading 2



<Enter your details here>

.....

.....

.....

DigitalPreservationCoalition

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

- It's not for the good of the bits

DigitalPreservationCoalition

We do preservation because we want :

1. Transparency
e.g. Data Protection, Freedom of Information …
2. Safety
e.g. detection, disaster, recovery, audit …
3. Knowledge
e.g. scientific value, access to heritage …
4. Wealth
e.g. efficient business, management of IP …
5. Health
e.g. research, safe innovation …
6. Environmental improvement
e.g. evidence-based policy development …

1. Legal compliance
e.g. Sarbanes-Oxley, Data Protection …
2. Regulatory compliance
e.g. power generation, aviation ….
3. Legal protection
e.g. patents, mis-selling, detection …
4. Unanticipated exploitation
e.g. petro-chemical, pharmaceuticals ….
5. Business continuity
e.g. product recall, disaster recovery …
6. Business value
e.g. the right information to the right people at the right time in a format they can use

DigitalPreservationCoalition

Our c

DigitalPreservationCoalition

Digital preservation is not just about ‘data’:
Digital preservation is not just about ‘access’:
Digital preservation is not just about ‘risk’:

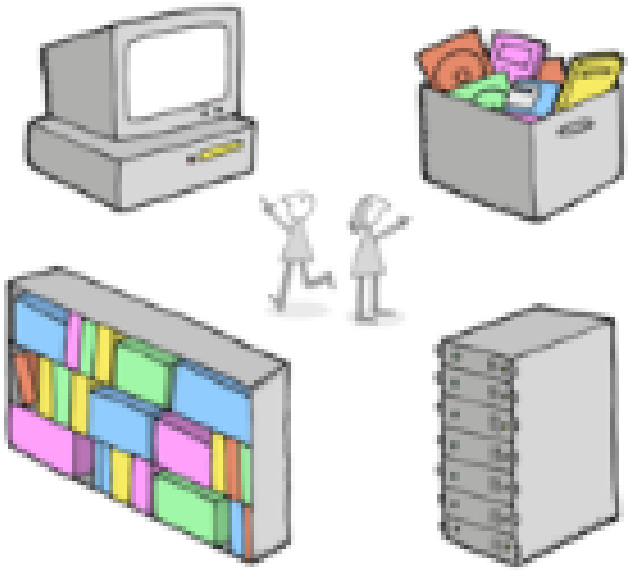
it’s about

people and opportunity

DigitalPreservationCoalition

Four basic approaches
and some fancy words

DigitalPreservationCoalition



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.

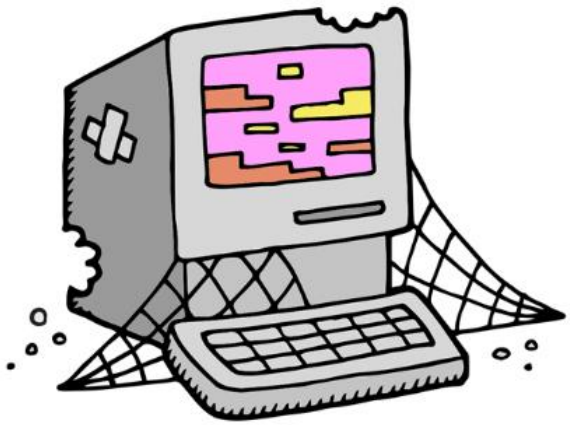
4. etc

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

DigitalPreservationCoalition

Seven challenges some solutions and more fancy words

DigitalPreservationCoalition

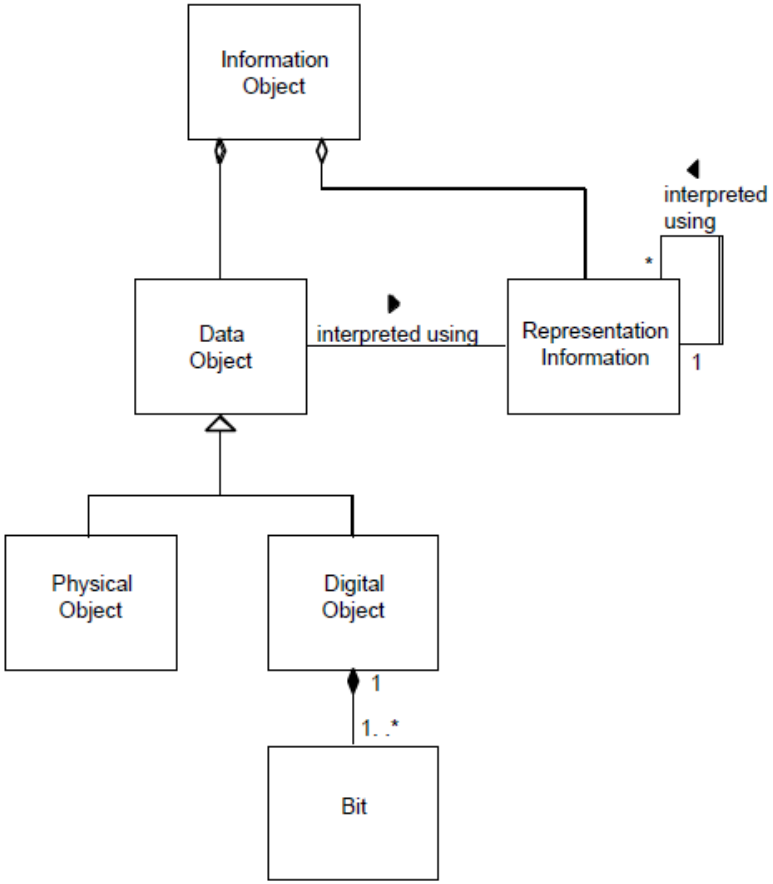


Challenge 1:

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

Different levels of answer:

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

DigitalPreservationCoalition

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.

DigitalPreservationCoalition

Challenge 2

Planning and learning

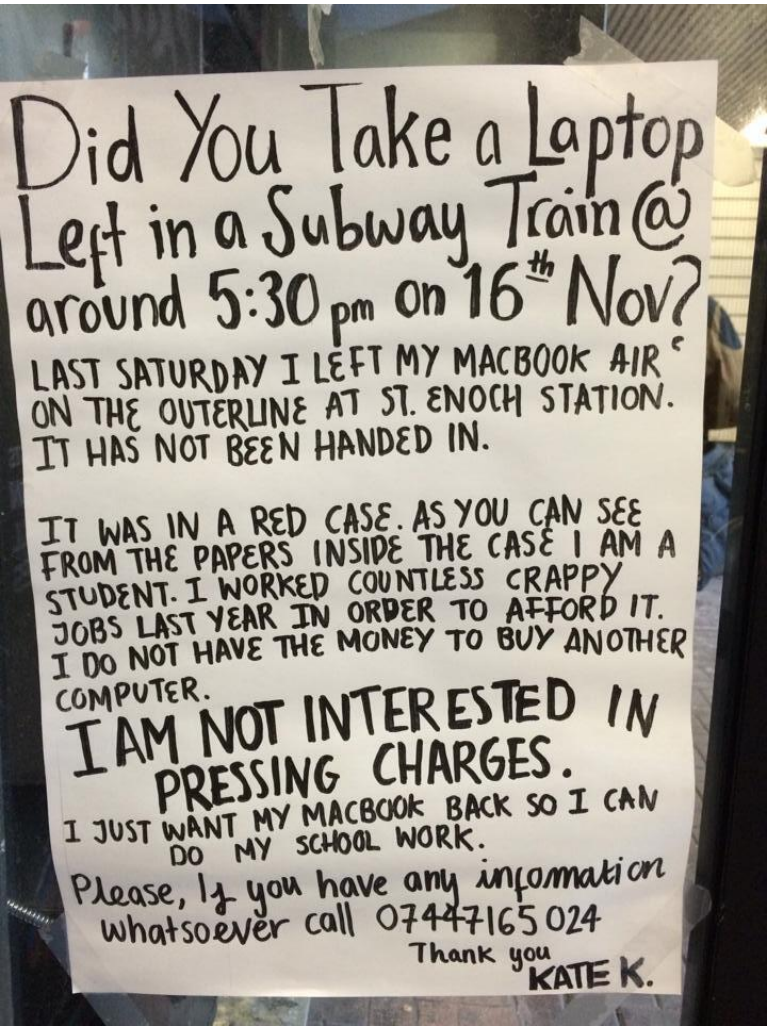
Be a learning institution

Different solutions:

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

DigitalPreservationCoalition

Challenge 3



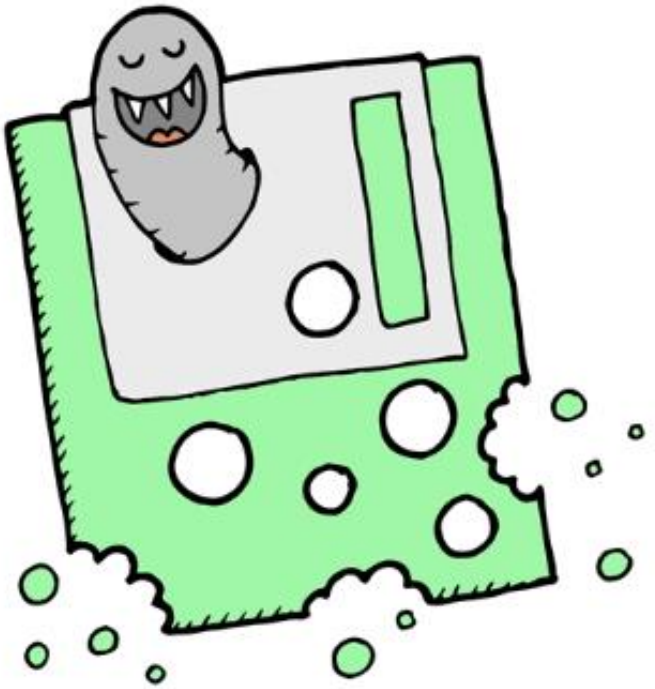
Storage media fail, have a short life and storage devices are subject to obsolescence.

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

DigitalPreservationCoalition

Challenge 3

Storage and refreshment



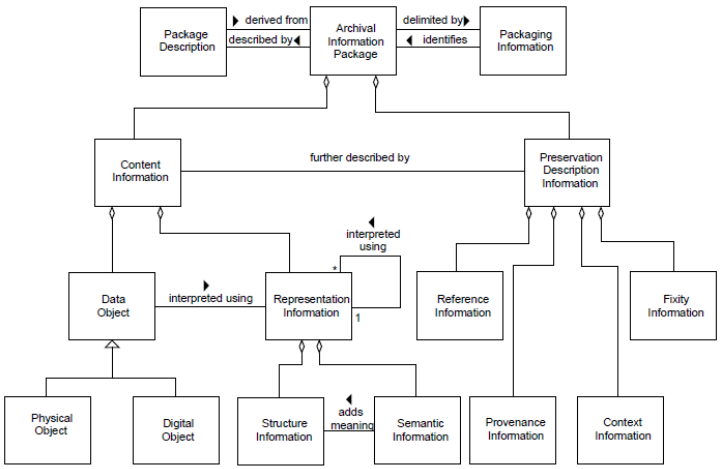
Different Solutions:

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

Beware: proliferation can become a problem

DigitalPreservationCoalition

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

Archive **Disseminate**

Submit

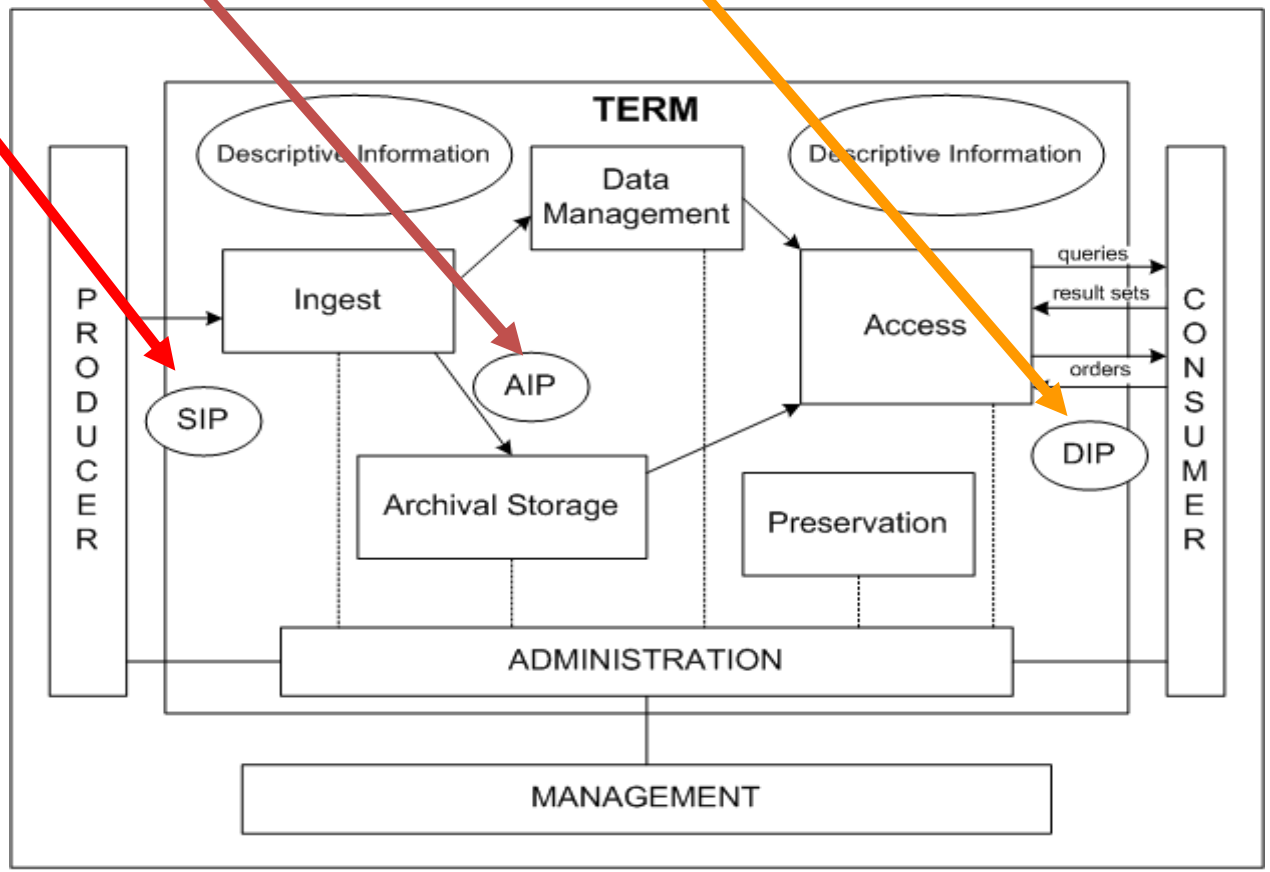
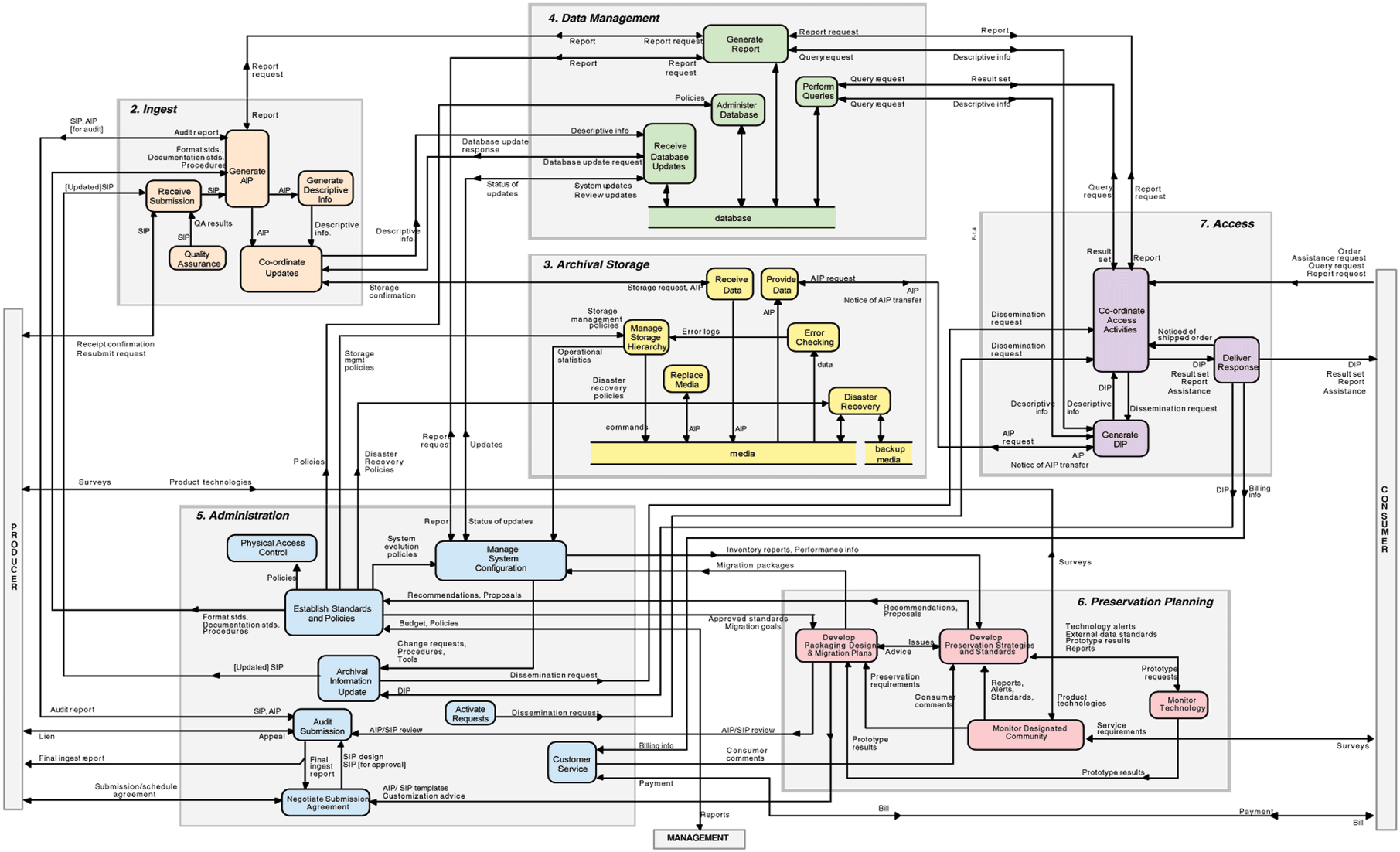
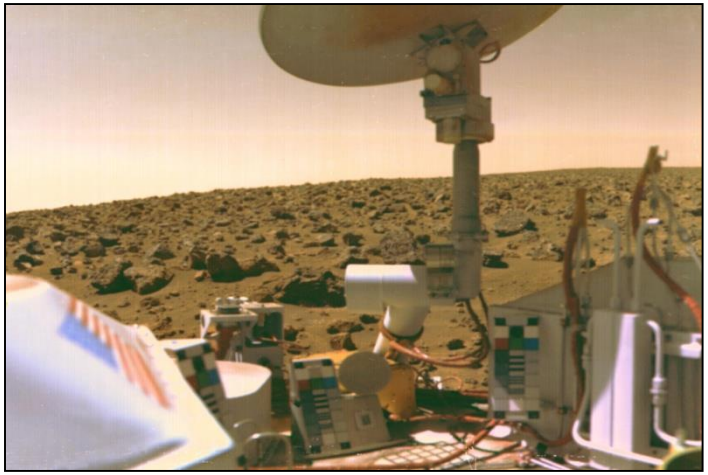


Fig. 1. Major functions of the OAIS Reference Model from Consultative Committee for Space Data Systems (CCSDS), CCSDS 650.0-W-1, Producer-Archive Interface Methodology Abstract Standard. (OAIS). White Book. Issue 1. Draft Recommendation for Space Data System Standards.

Picture from DLib

DigitalPreservationCoalition





Courtesy NASA/JPL-Caltech

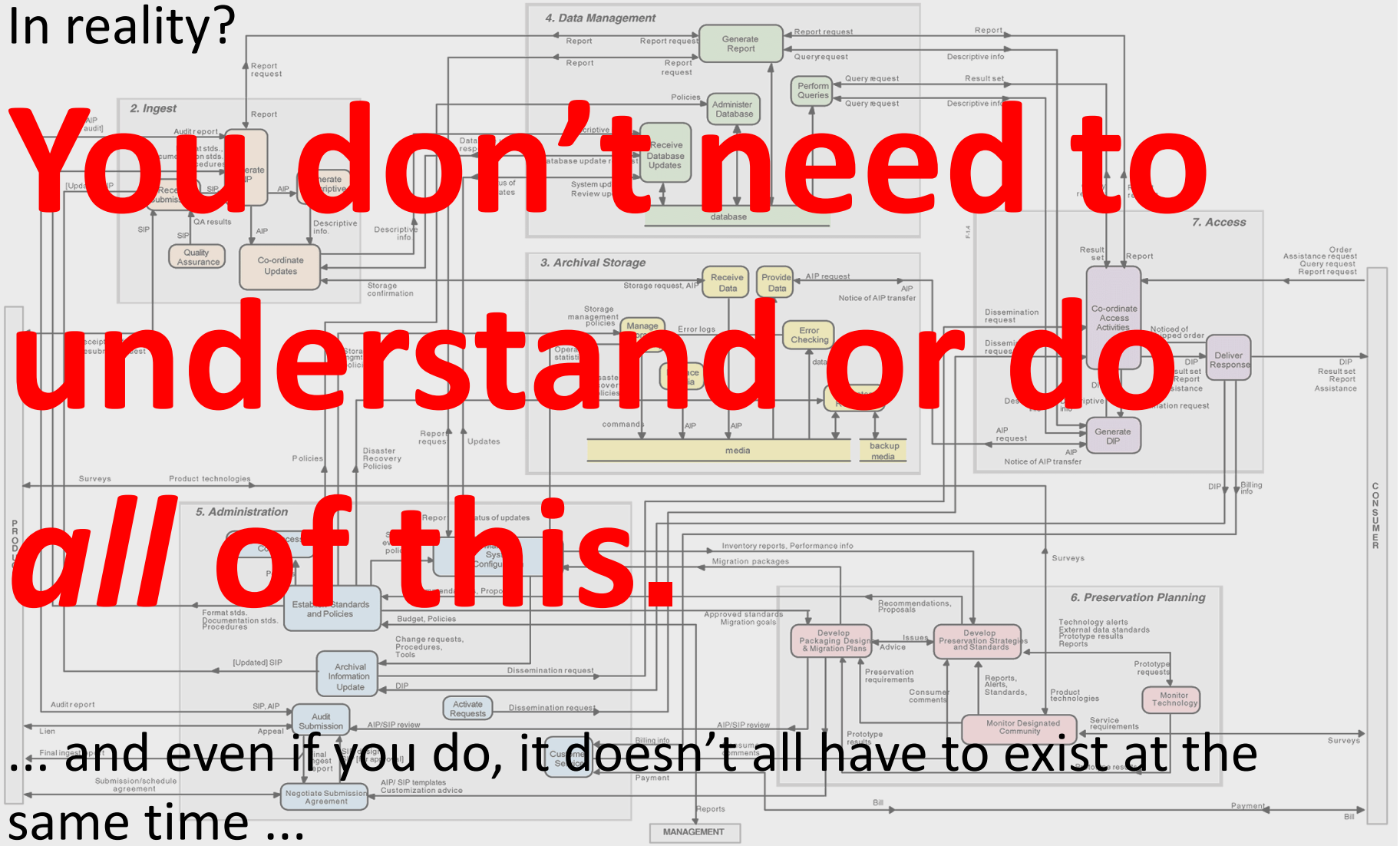
Consultative Committee on Space Data Systems

Inadvertent comparison with NASA

Scales up really well

DigitalPreservationCoalition

In reality?



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 ...

DigitalPreservationCoalition

Challenge 5

Digital resources are intolerant of gaps in preservation.

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

DigitalPreservationCoalition

Challenge 5

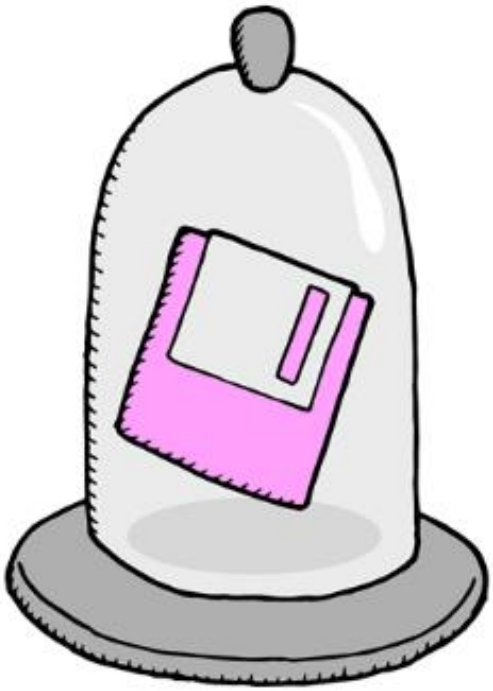
On-going preservation

Different responses:

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

DigitalPreservationCoalition

Challenge 6



Resources can be corrupted or tampered without trace

Need fixity
and authenticity checks

DigitalPreservationCoalition

Challenge 6

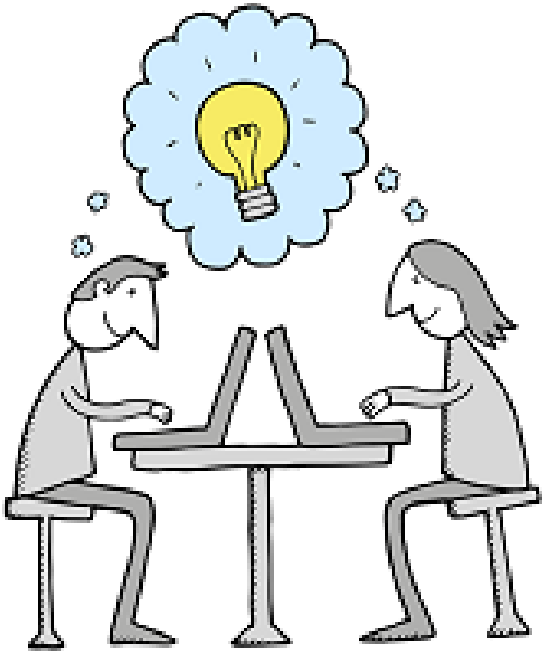
Fixity and authenticity

A variety of solutions:

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

DigitalPreservationCoalition

Challenge 7



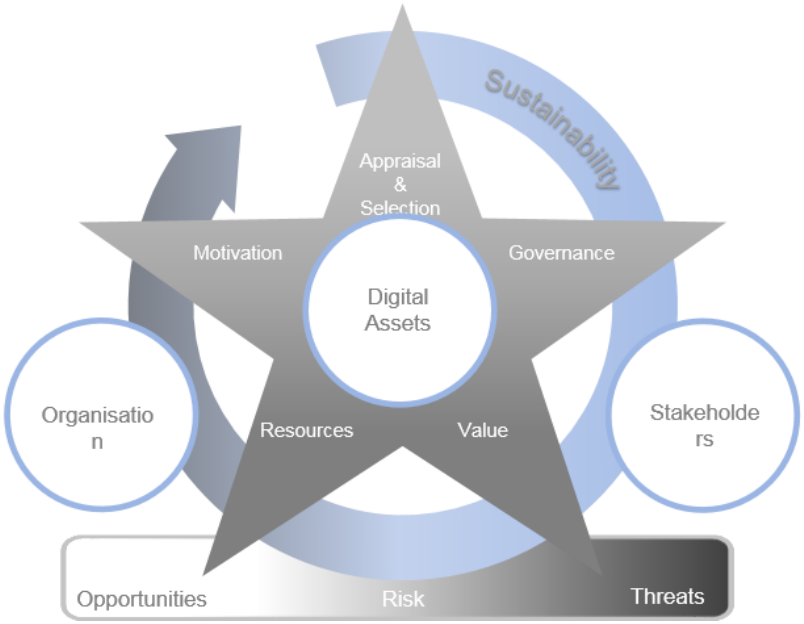
Digital resources need long term financial commitment to be sustainable

It's an unfunded mandate
Benefits unclear
Costs uncertain

DigitalPreservationCoalition

Challenge 7

Sustainability



Business Case Toolkit
Curation Costs Exchange
Digital Curation Sustainability Model
Mandates
Revenues

DigitalPreservationCoalition

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