# Digital Preservation and long term access:
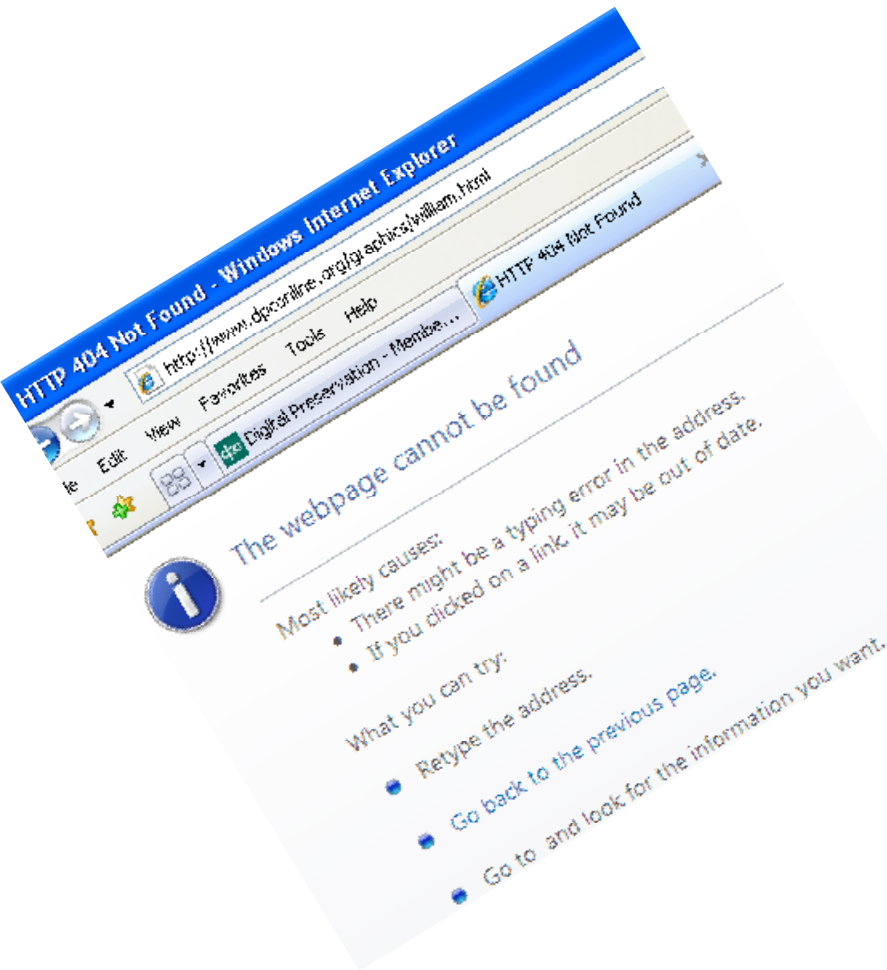
## challenges, opportunities approaches, tools ….

…and a couple of words about the DPC

**William Kilbride**
**william@dpconline.org**

www.dpconline.org

- What *are* the challenges?
- What *are not*?

- What is the solution?
- What *is not?*

# Digital preservation typically makes bleak reading …



**Digital Archives**

**ads** ARCHAEOLOGY DATA SERVICE

**S**trategies for **D**igital **D**ata

Findings and Recommendations from Digital Data in Archaeology: A Survey of User Needs

Cadw

DoENI

English Heritage

The Herita Coun

**ads** ARCHAEOLOGY DATA SERVICE

**Digital Ar**

The present state of digital archiving in

Netw ork 13%

CD-ROM 14%

Tape 22%

Hard disc 28%

Floppy disc 23%

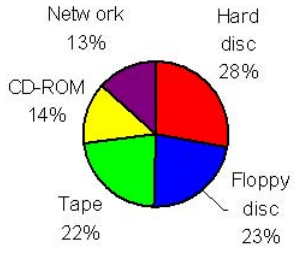**The popularity of certain storage options**

**ads** ARCHAEOLOGY DATA SERVICE

**Digital Archives**

In short, the archaeological record could be decaying faster in its digital form than it ever did in the ground

**DigitalPreservationCoalition**

- Let's go back to first principles …

- Digital data has value. It is an asset.
- It has potential and creates new opportunities.
- Use gives rise to direct and indirect outcomes.
- …but…

- Deployment depends on software, hardware and people.
- Software, hardware and people change.
- …therefore…

- Access is not guaranteed without (some) action
- Value, opportunity, impact not guaranteed
- Potential outcomes – ie health or research – lost

Digital preservation is not about 'data':
Digital preservation is not about 'access':
Digital preservation is not about 'risk':

# it's about people and opportunity

Digital preservation is not about 'data':

Digital preservation is not about 'access':

Digital preservation is not about 'risk':

Digital preservation is about:

# Healthier -Wealthier Safer - Smarter - Greener

people and communities

*Challenge 1*

*Access depends on the configuration of hardware and software and the capacity of the operator.*

Documentation can capture these configurations

Emulation or Migration can create the conditions where access is possible.

*Challenge 2*

*Technology continues to change creating the conditions for obsolescence.*

Technology watch services can give advanced notice of obsolescence.

Plan for the long term early not late.

Self-preserving technology

*Challenge 3*

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

Storage media can be refreshed and can self-check.

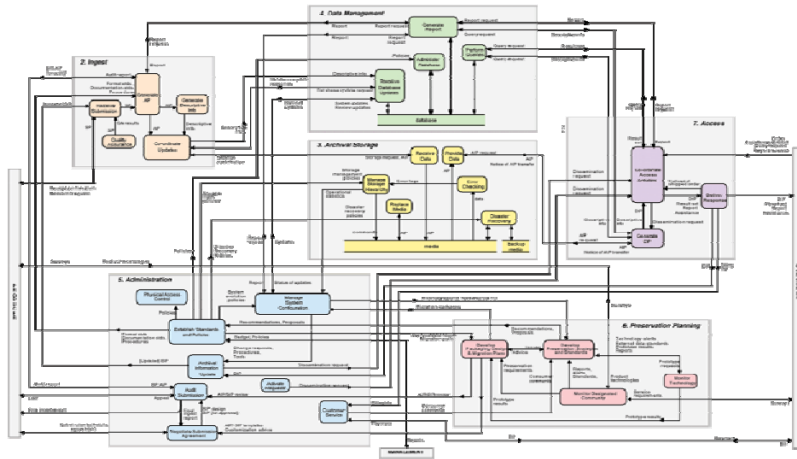Storage densities continue to improve offering greater capacity at less cost.

(storage is cheap – discovery is expensive)

*Challenge 4*

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

Systems can be modular and transparent.

Fitness for purpose can be monitored through time.

Recursion of process

www.dpconline.org

*Challenge 5*

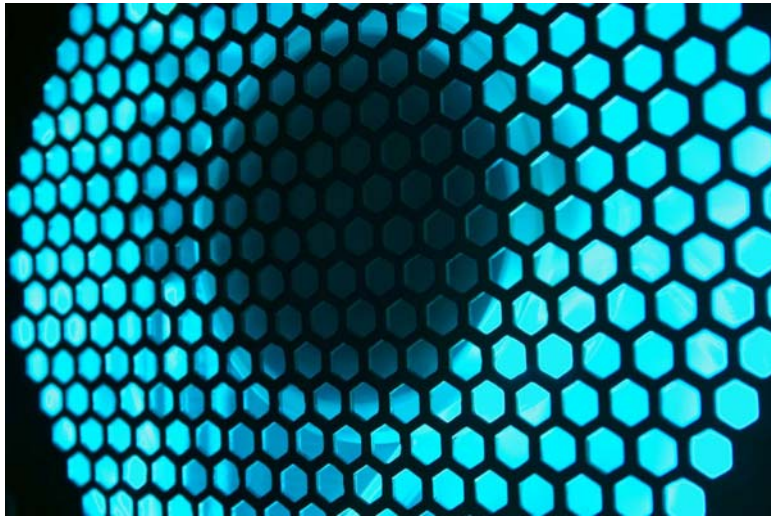*Digital resources can be altered, corrupted or deleted without obvious detection.*

Signatures and wrappers can safeguard authenticity

Security can control access.

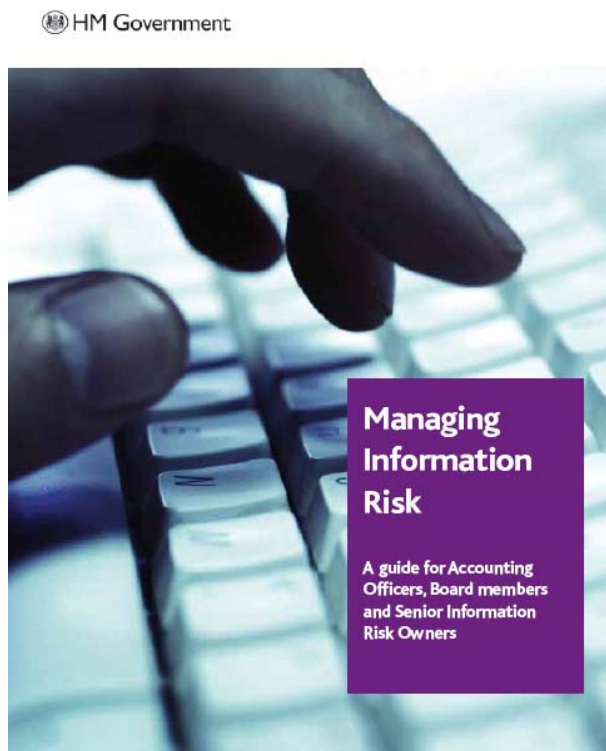Copies are perfect replicas with no degradation.

*Challenge 6*

*Digital resources are intolerant of gaps in preservation.*

Ongoing risk management can provide monitoring.

There are significant economies of scale

Many processes can be automated.

*Challenge 7*

*The necessary skills are badly distributed.*

That's why we're here…

Leadership
Training
Experience
Learning by doing

**Digital**Preservation**Coalition**



# Key Approaches
# 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 so that information content can be rendered*

# 3. Hardware preservation

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

# 4. Exhumation

*Maintaining access to an execution environment or software services so that processes can be re-run with new data*

www.dpconline.org

*Three ways to think about Digital Preservation*

*1. A three legged stool*
*2. Digital lifecycles*
*3. Archival information systems*
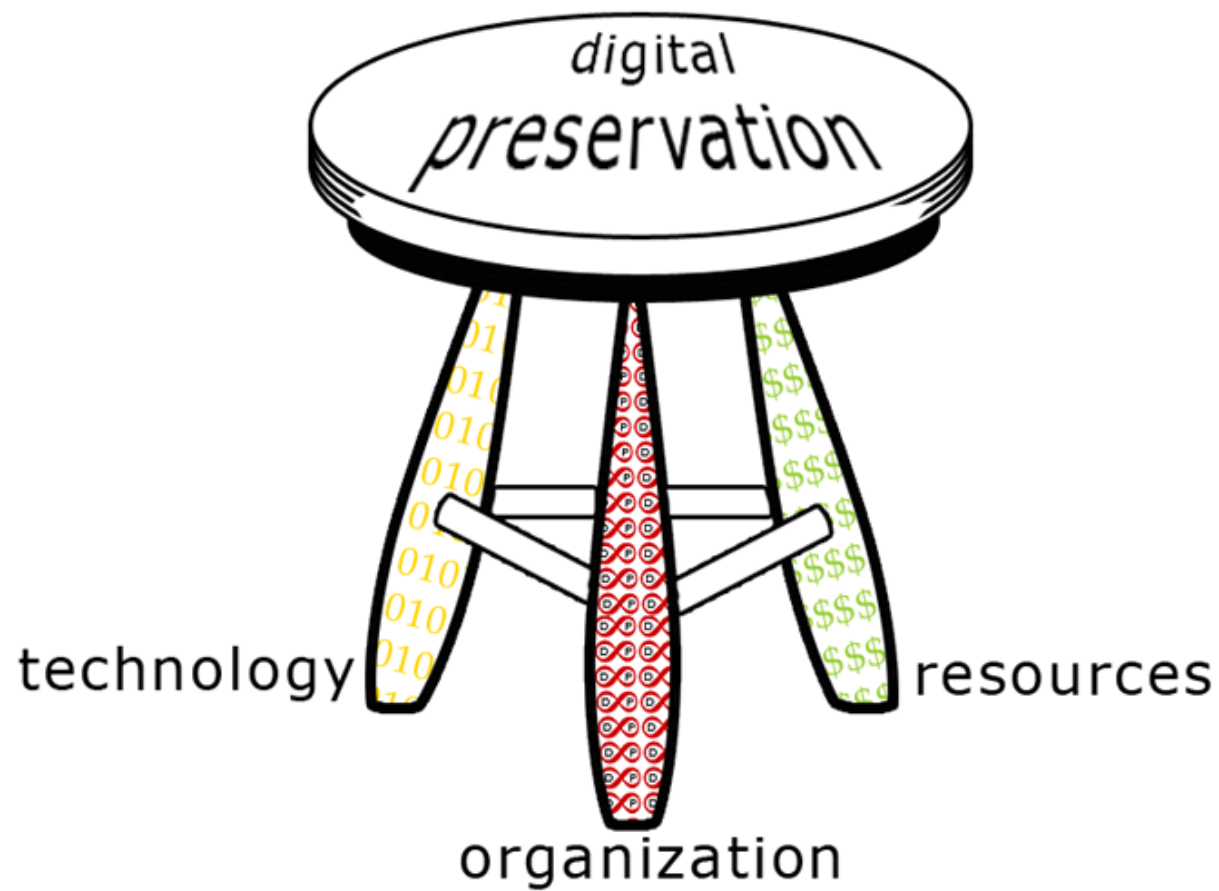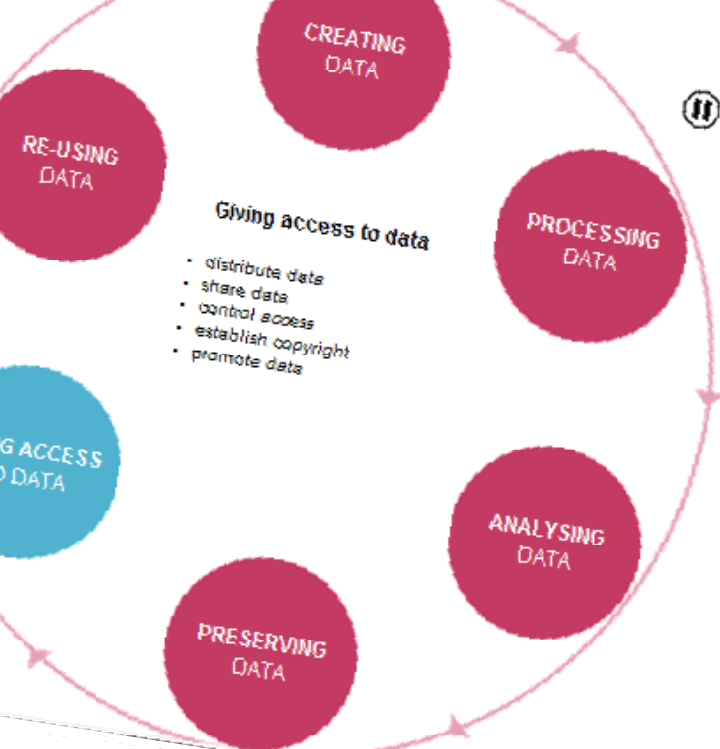
*Digital lifecycles*

*Think of a research question*
*Gather some relevant data*
*Process the data*
*Refine the data*
*Draw some conclusions*
*Publish your findings and data*
*Start again*
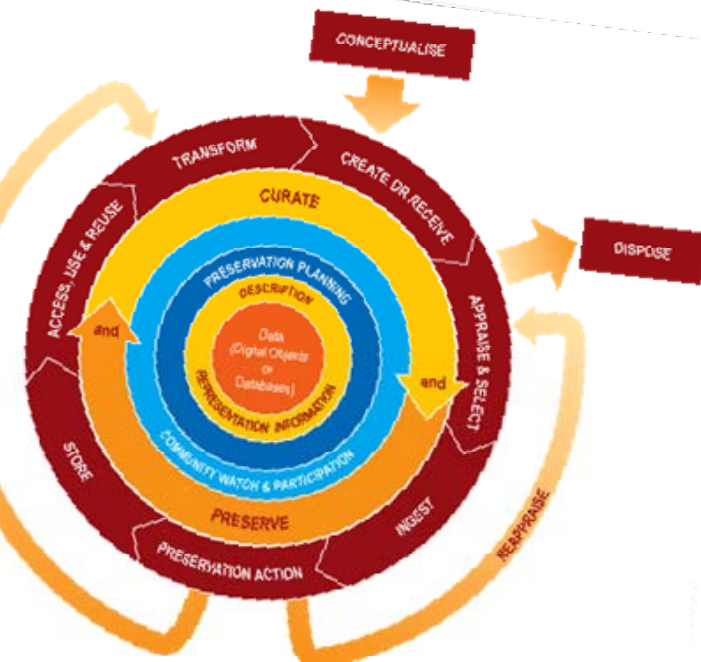
# Reference Model for an Open Archival Information System 'OAIS'
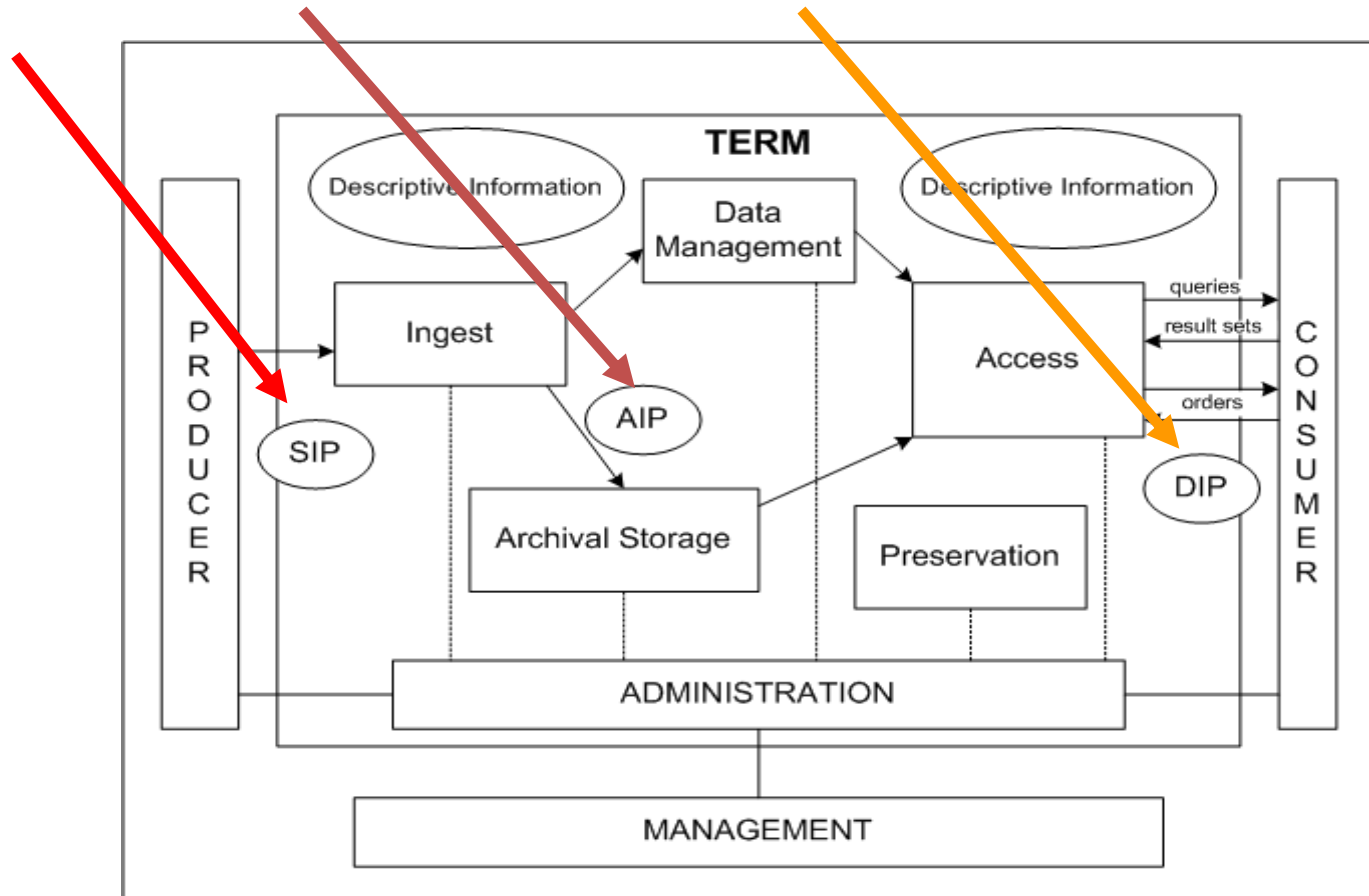
Fig. 1. Major functions of the OAIS Reference Model from Consultative Committee for Space Data Systems (CCSDS), CCSDS 650.0-W-1, Producer-Archive Interface Methodology Abstract Standard. (OAIS). White Book. Issue 1. Draft Recommendation for Space Data System Standards.

Picture from DLib

www.dpconline.org

Digital**Preservation**Coalition



Courtesy NASA/JPL-Caltech

Consultative Committee for Space Data Systems

Inadvertently comparing yourself to NASA ...

Scalability? It scales up really well ...

www.dpconline.org

The reality?

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

… and it doesn't all have to exist at the same time

www.dpconline.org

# *some* tools ....

Knowing what you've got
PRONOM+DROID
Planning what to do with it
PLATO
Putting it somewhere safe
LOCKSS

# file formats and their characteristics

**Digital**Preservation**Coalition**

**The National Archives**

About us | Education | Records | Information management | Shop online

You are here: Home > Services for professionals > Preservation > PRONOM
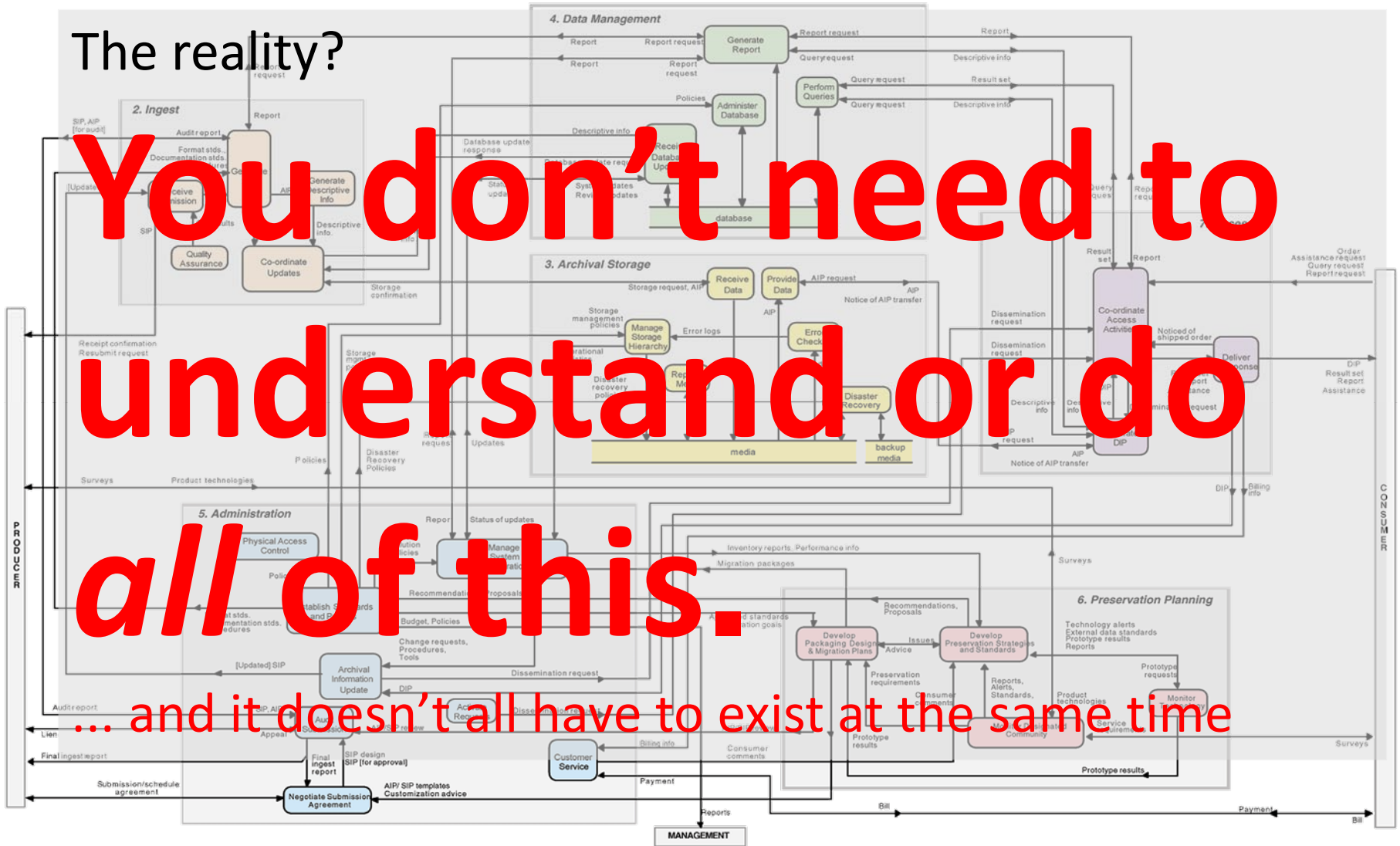
The **technical registry**
**PRONOM**

🏠 Welcome   : **About**   ⊕ **Add an entry**
🔍 **Search**   **? Help**   🗐 **Information resources**

🏠 **Welcome to PRONOM**

PRONOM changes and DROID signature file release notes.

Find out more about our plans to make PRONOM's data avai

The online registry of technical information. PRONOM is a res
software products and other technical components required
historical or business value. Find out about the future of PRC

🔍 **Search PRONOM ›**

Simple search | File format | PRONOM Unique Identifier | Software | Vendor | Lifecycles | Migration Pathways

**Details for: Tagged Image File Format 3**   🖫 Save as... **XML | CSV**   🖨 Print

Go to: **Summary** | **Documentation** ❯ | **Signatures** ❯ | **Compression** ❯ | **Character encoding** ❯ | **Rights** ❯ | **Reference files** ❯ **Properties** ❯

## Summary

| | |
|---|---|
| **Name** | Tagged Image File Format |
| **Version** | 3 |
| **Other names** | TIFF (3) |
| **Identifiers** | PUID: fmt/7<br>MIME: image/tiff<br>Apple Uniform Type Identifier: public.tiff |
| **Family** | |
| **Classification** | Image (Raster) |
| **Disclosure** | Full |
| **Description** | The Tagged Image File Format (TIFF) is a raster image format originally developed by the Aldus Corporation, primarily for use in scanning and desk-top publishing. When Adobe Systems Incorporated purchased Aldus in 1994, they acquired the rights to the TIFF format and have maintained it since then. TIFF files comprise three sections: an Image File Header (IFH), an Image F Directory (IFD), and the image data. TIFF files can contain multiple images (multi-page TIFF), and ea image has a separate IFD. The IFH always appears at the beginning of the file, and is immediately followed by a pointer to the first IFD. The IFD contains metadata which describes the associated image, stored as a series of tags. The IFD also contains a pointer to the actual image data. TIFF 3.C supports colour depths from 1 bit to 24 bit (e.g. monochrome to true colour), and a range of compression types (RLE, and CCITT Group 3 and Group 4). |

DROID: Search and report on files from an entire network

Identify files by extension
Identify files by contents
Report errors and concerns
Provides 'checksum' signatures

Part of a preservation architecture

**http://www.nationalarchives.gov.uk/pronom/**

www.dpconline.org

PLATO – digital preservation planning tool

From the PLANETS suite of tools
Planning a shared function
Typically ad hoc

PLATO offers three things:
- Methodology for planning
- Online planning toolkit
- Library of plans

**http://www.ifs.tuwien.ac.at/dp/plato/intro.html**

**Digital**PreservationCoalition

Team Digital Preservation and the Arctic Mountain Adventure
wepreserve  9 videos  Subscribe

PRESERVATION PLANNING

1:07 / 4:23   CC   360p

Like   Add to   Share   2,984

Uploaded by wepreserve on 14 Jun 2010

8 likes, 0 dislikes

**Define requirements**

Define basis → Choose records → Identify requirements

**Evaluate alternatives**

Go/No-Go ← Define alternatives

Develop experiment → Run experiment → Evaluate experiment

**Analyse results**

Analyse results ← Set importance factors ← Transform measured values

Preservation Action Recommendation

**Build preservation plan**

Create executable preservation plan → Define preservation plan → Validate preservation plan

Preservation Plan

**http://www.ifs.tuwien.ac.at/dp/plato/intro.html**

www.dpconline.org

LOCKSS – Lots Of Copies Keeps Stuff Safe



Illustration courtesy of LOCKSS

Cooperative replication
Dark archive
Self-fixing

Originally E-journals for libraries
Post cancellation access

Now also 'Private LOCKSS Networks' Eg MetaArchive

http://lockss.stanford.edu/
http://www.lockssalliance.ac.uk/
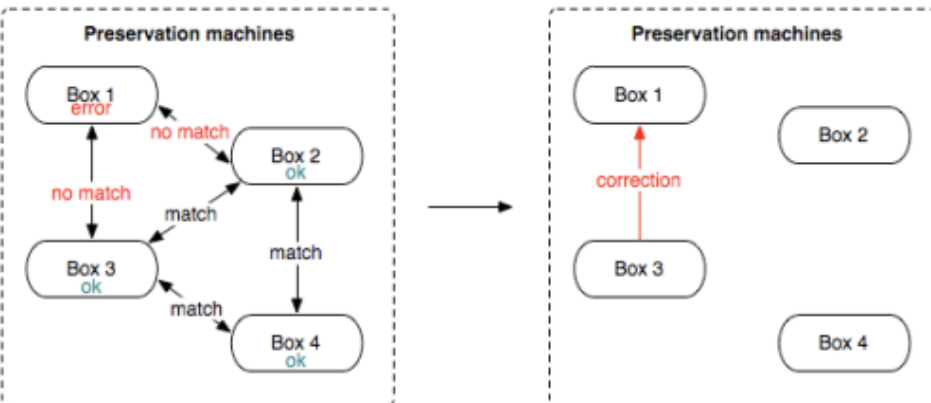http://www.metaarchive.org/

Part of a preservation architecture

www.dpconline.org

PRONOM and DROID
PLATO
LOCKSS

= Parts of a preservation architecture

Not just at the end of the process but all the way through creation

Many more tools ...

# Oh and ... the Digital Preservation Coalition

*...to make our digital memory accessible tomorrow ...*

Enabling
Agenda-setting

www.dpconline.org

PORTICO

Glasgow University

Public Records Office
of Northern Ireland

Trinity College Dublin
National Library of
Ireland

National Library of Wales
RCAHMW

Archives and Records
Association

English Heritage
RCUK

Portsmouth
University

Bodleian
Libraries

OCLC
St Andrews University

National Library of Scotland
National Archives of Scotland
Digital Curation Centre
Creative Scotland
RCAHMS
UK LOCKSS Alliance

British Library
Archaeology Data Service

Open University Library

Cambridge University Library

UK Data Archive

Wellcome Library
King's College London
Publishers Licensing Society
RLUK
The National Archives
Tate
Natural History Museum
MLA
ULCC
BBC
Parliamentary Archives
RIN
JISC
Metropolitan Police

www.dpconline.org

*... first and foremost a coalition...*

Shared challenge
Cross-sector
Cross-discipline
Policy and practice

www.dpconline.org

*... Enabling*
DP Handbook
Email lists
Leadership Programme
What's new
Workshops
Conferences
Technology watch reports

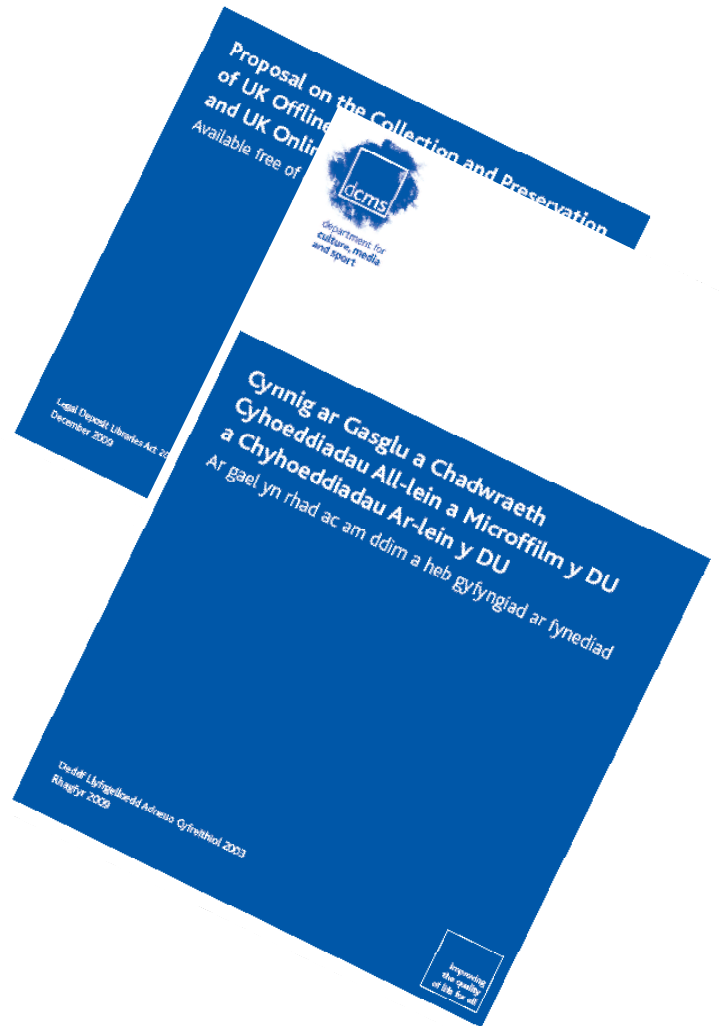www.dpconline.org

*... agenda setting*

Digital Britain
Archives for the 21$^{st}$Century
Electronic Legal Deposit
Heritage Science Strategy
Preservation Exceptions

Web Archiving and
Preservation Task Force

www.dpconline.org

*Join us!*

Associate members
Full members
Personal members

Next – archive and
information schools

# Digital Preservation in byte sized chunks:

## why we should be careful what we wish for ....

biographical, idiosyncratic observations

**William Kilbride**
**william@dpconline.org**

www.dpconline.org