DigitalPreservationCoalition

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

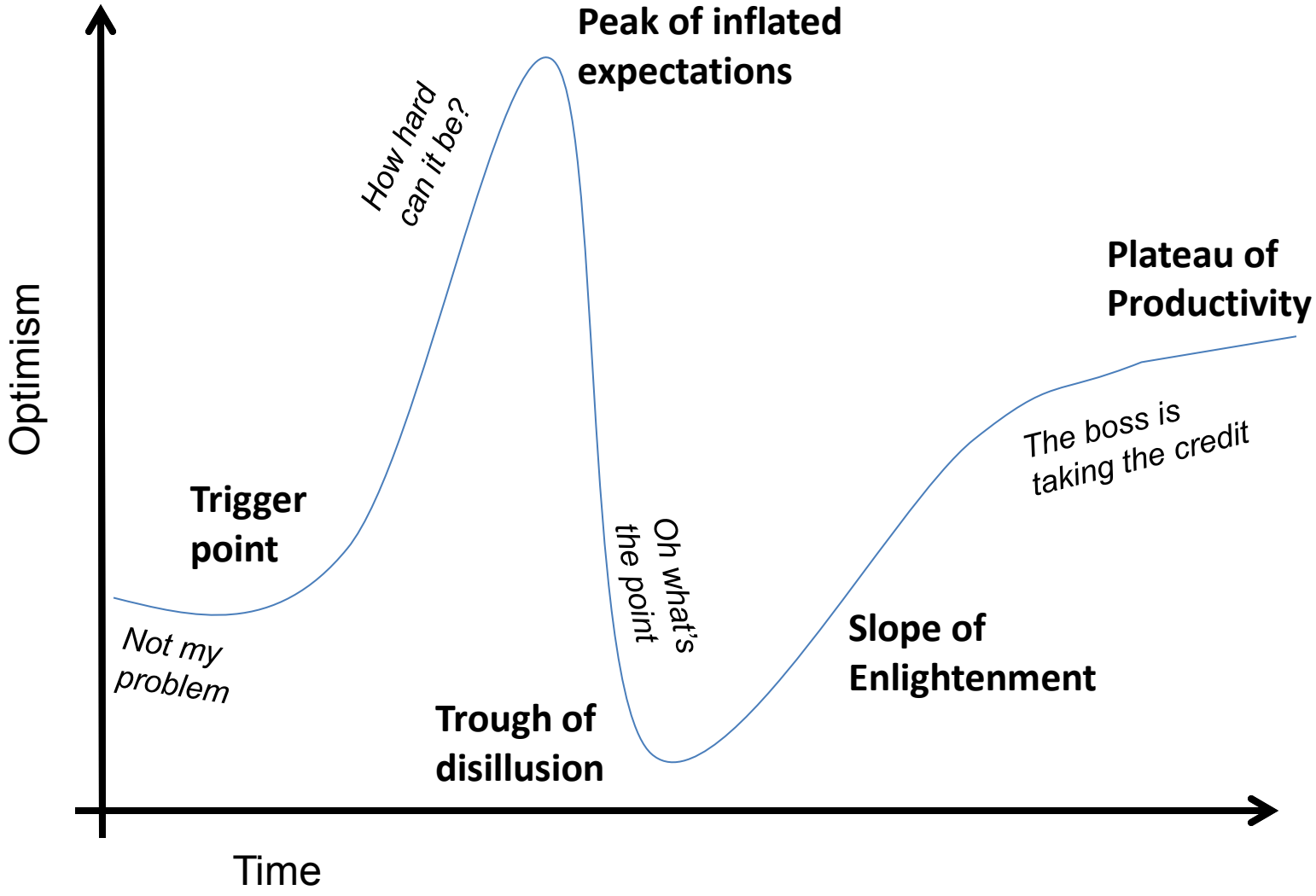
It won't go away
It won't do itself

You already have many of the skills you need!

William Kilbride
william @dpconline.org

DigitalPreservationCoalition

With apologies to Gartners' Hype Cycle



DigitalPreservationCoalition

What's the problem?

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

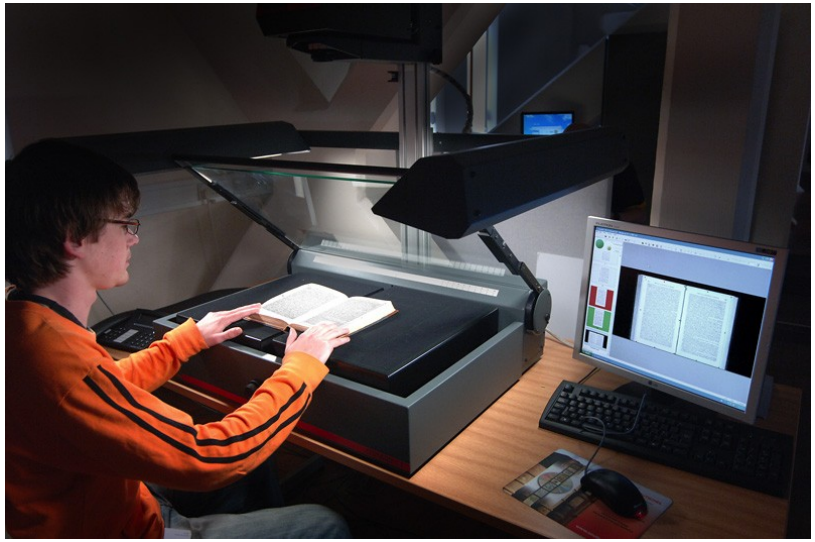
...but...

- 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

DigitalPreservationCoalition



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

‘In perpetuity’

‘Indefinitely’

‘50 years’

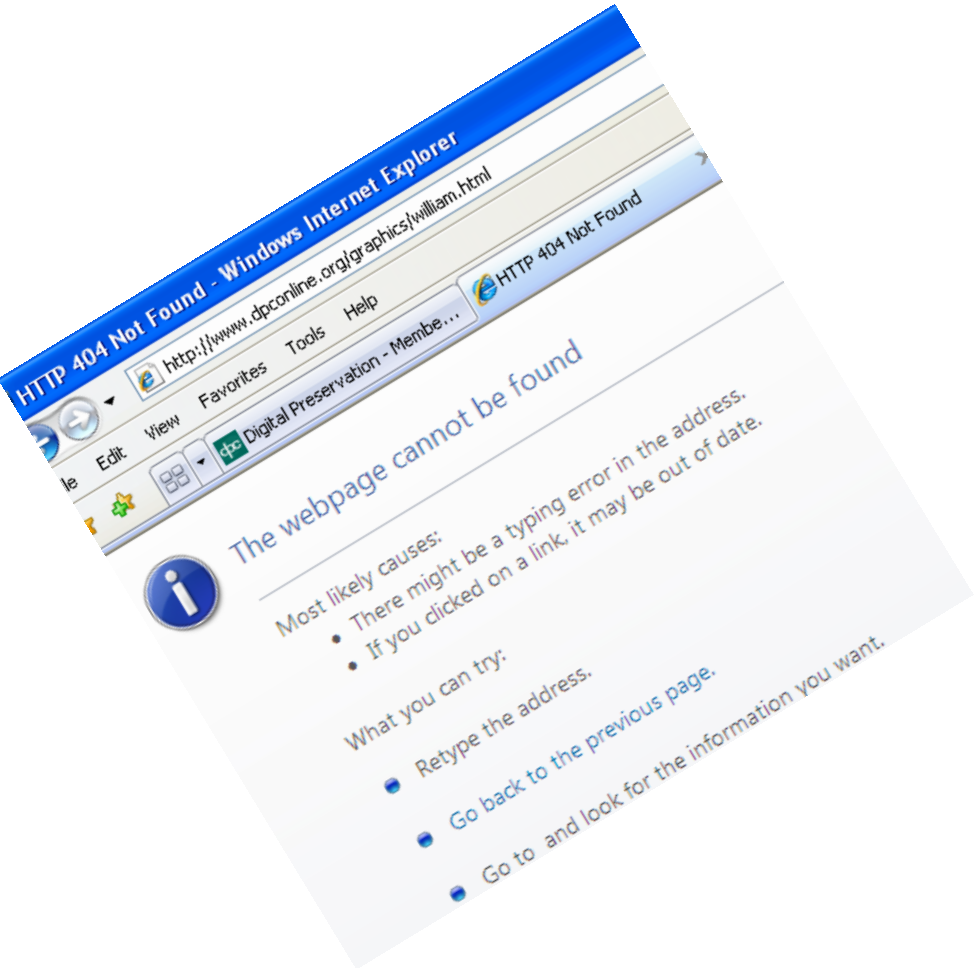
‘10 years then elsewhere’

‘until 2014’

‘forever or for three years’

DPC/Portico/ULCC 2010

DigitalPreservationCoalition



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

(Spencer et al 2009)

DigitalPreservationCoalition

Challenge 1

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

Documentation can capture configurations

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

DigitalPreservationCoalition

Challenge 2

Technology continues to change creating the conditions for obsolescence.

Technology watch services can give advanced notice of obsolescence.

Migration and emulation reduce the impact of changes in technology.

DigitalPreservationCoalition

Challenge 3

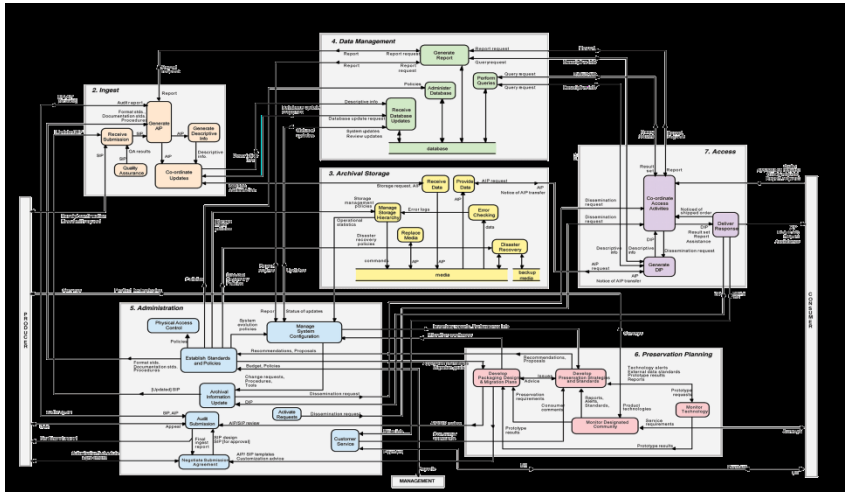
Storage media have a short life. Storage devices are subject to obsolescence.

Storage media can be refreshed and can self-check.

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

DigitalPreservationCoalition

Challenge 4

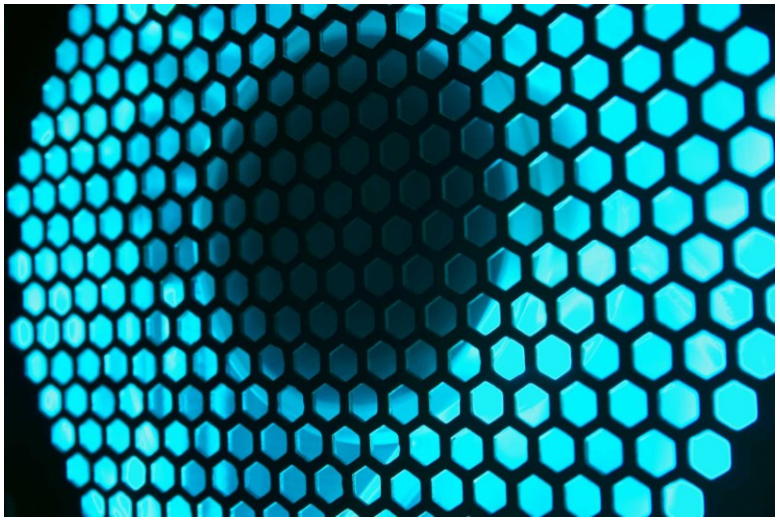


Digital preservation systems are subject to the same obsolescence as the objects they safeguard.

Systems can be modular and conform to standards.

Fitness for purpose can be monitored through time.

DigitalPreservationCoalition



Challenge 5

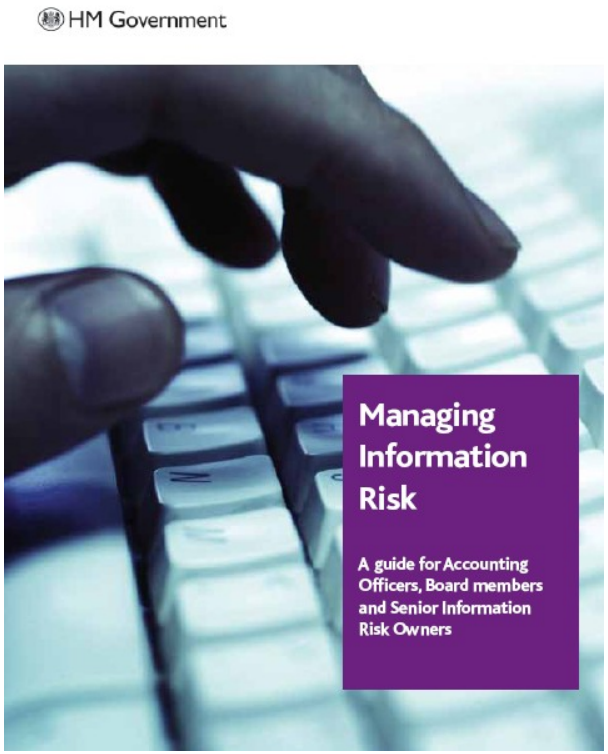
Digital resources can be altered, corrupted or deleted without obvious detection.

Signatures and wrappers can safeguard authenticity

Security can control access.

Copies are perfect replicas with no degradation.

DigitalPreservationCoalition



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.

DigitalPreservationCoalition

Challenge 7

The necessary skills are badly distributed.

Six core functions of a DP service ...

1. Determine what content is appropriate and negotiate for it
2. Obtain appropriate control over content received
3. Understand and survey the needs of the user community
4. Ensure independent utility of data within the system
5. Follow procedures for preservation
6. Disseminate data to users

... you probably have many of the skills and policies already

DigitalPreservationCoalition

Challenge 8

We have limited experience.

Rapid churn in technology accelerates our research

Transformed in last decade

This is a shared problem

DigitalPreservationCoalition

Key Approaches

1. Migration

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

2. Emulation

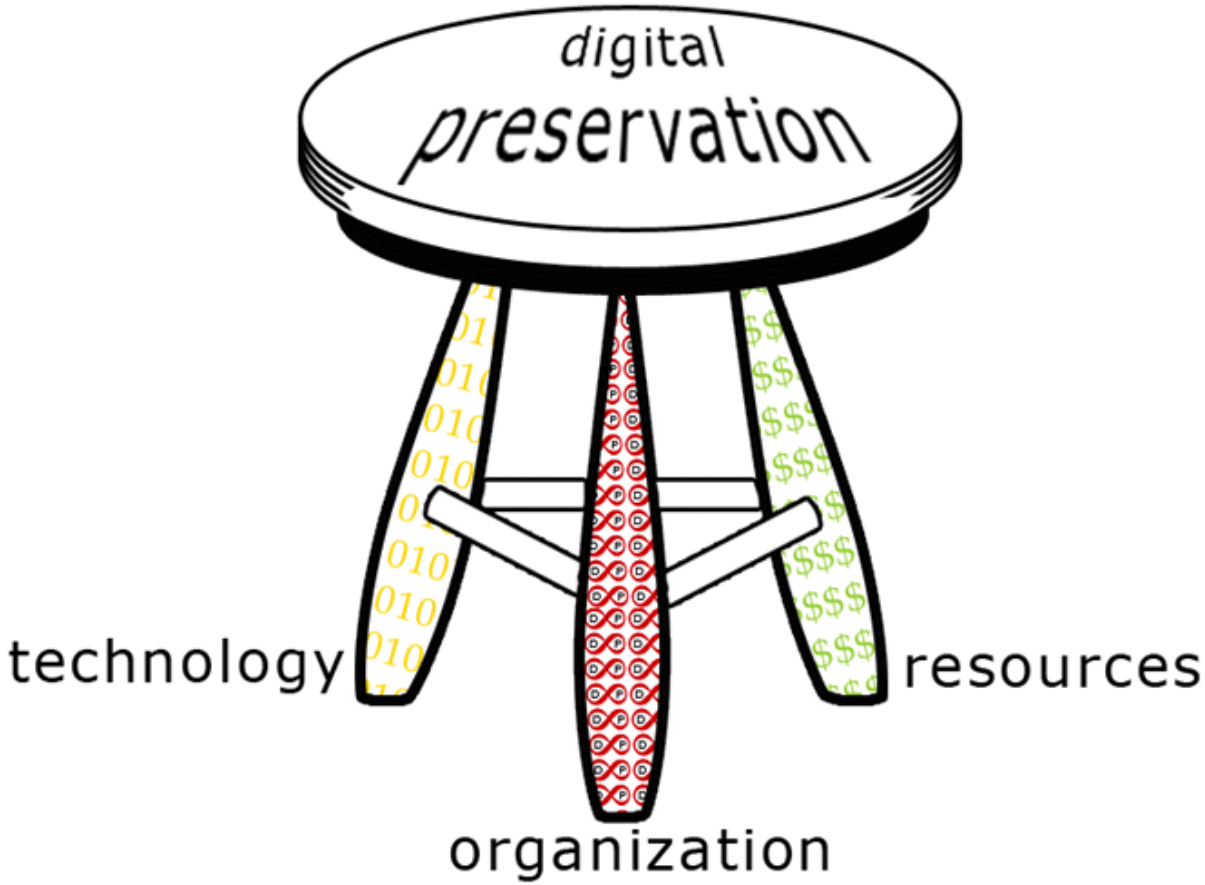
Intervening in the operating system to ensure that old software can function and information content can be read

3. Hardware preservation

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

4. Exhumation

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


digital
preservation
technology
resources
organization

DigitalPreservationCoalition

Key Terms and Functional Relationships

Reference Model for an Open Archival Information System 'OAIS'

Archive **Disseminate**

Submit

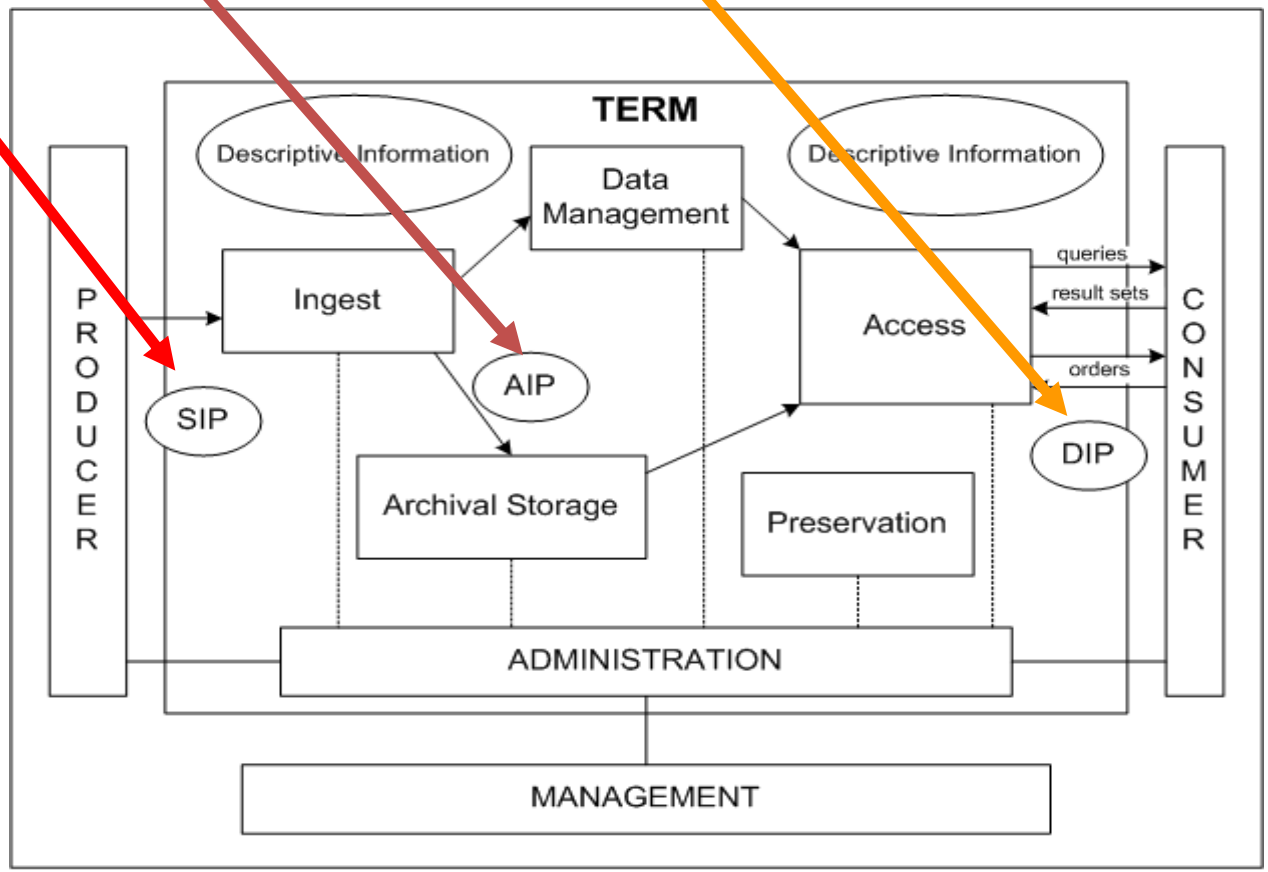
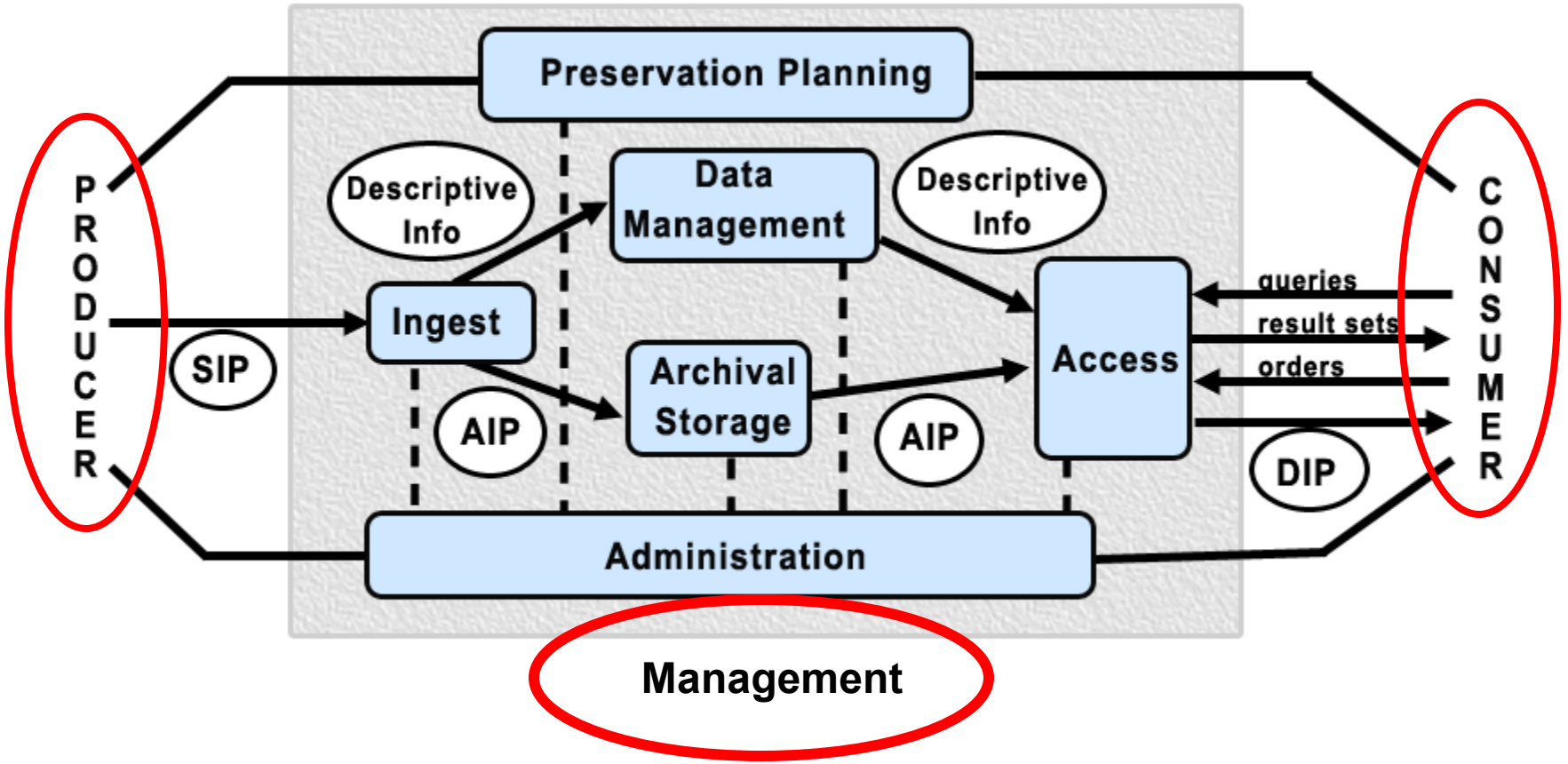


Fig. 1. Major functions of the OAIS Reference Model from Consultative Committee for Space Data Systems (CCSDS), CCSDS 650.0-W-1, Producer-Archive Interface Methodology Abstract Standard. (OAIS). White Book. Issue 1. Draft Recommendation for Space Data System Standards.

Picture from DLib

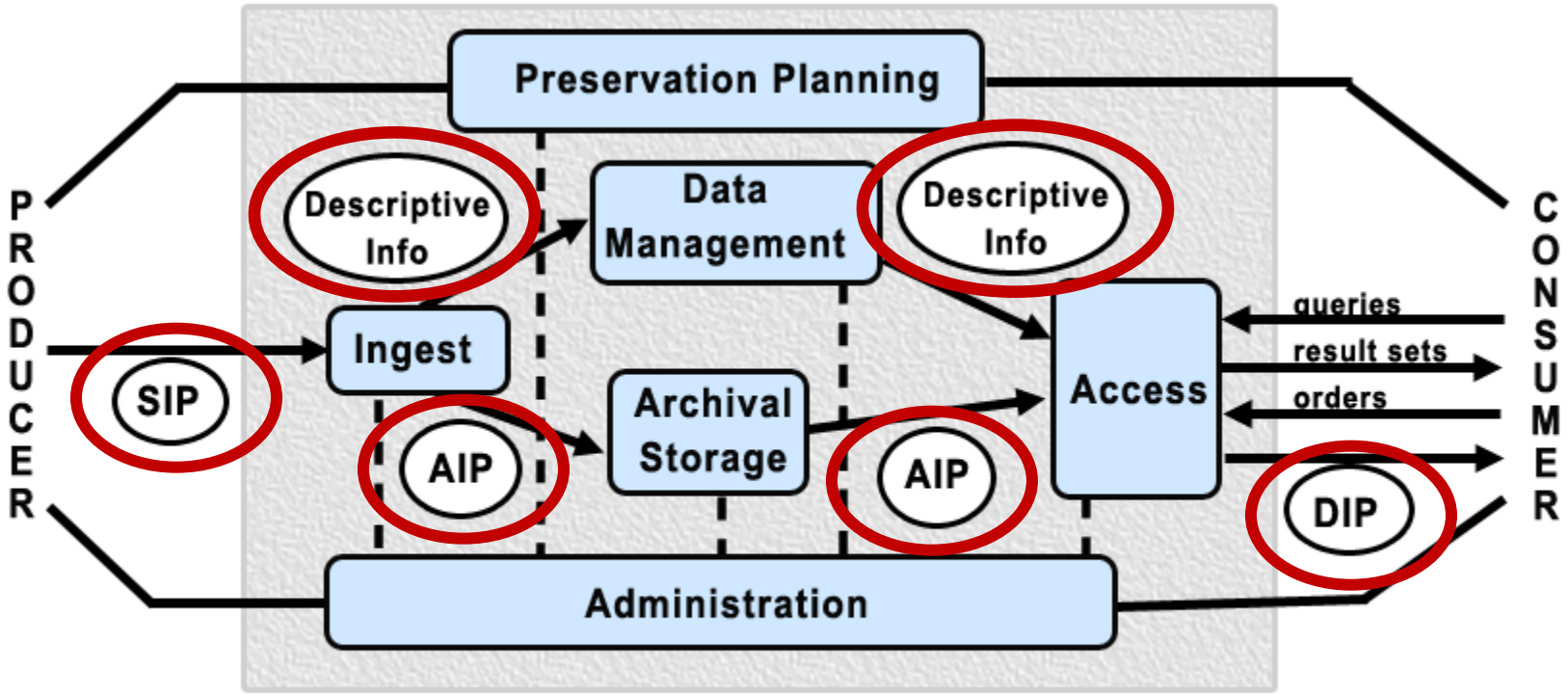
Key Vocabulary: the actors



Who are your producers?

Who are your consumers?

Key Vocabulary: the objects

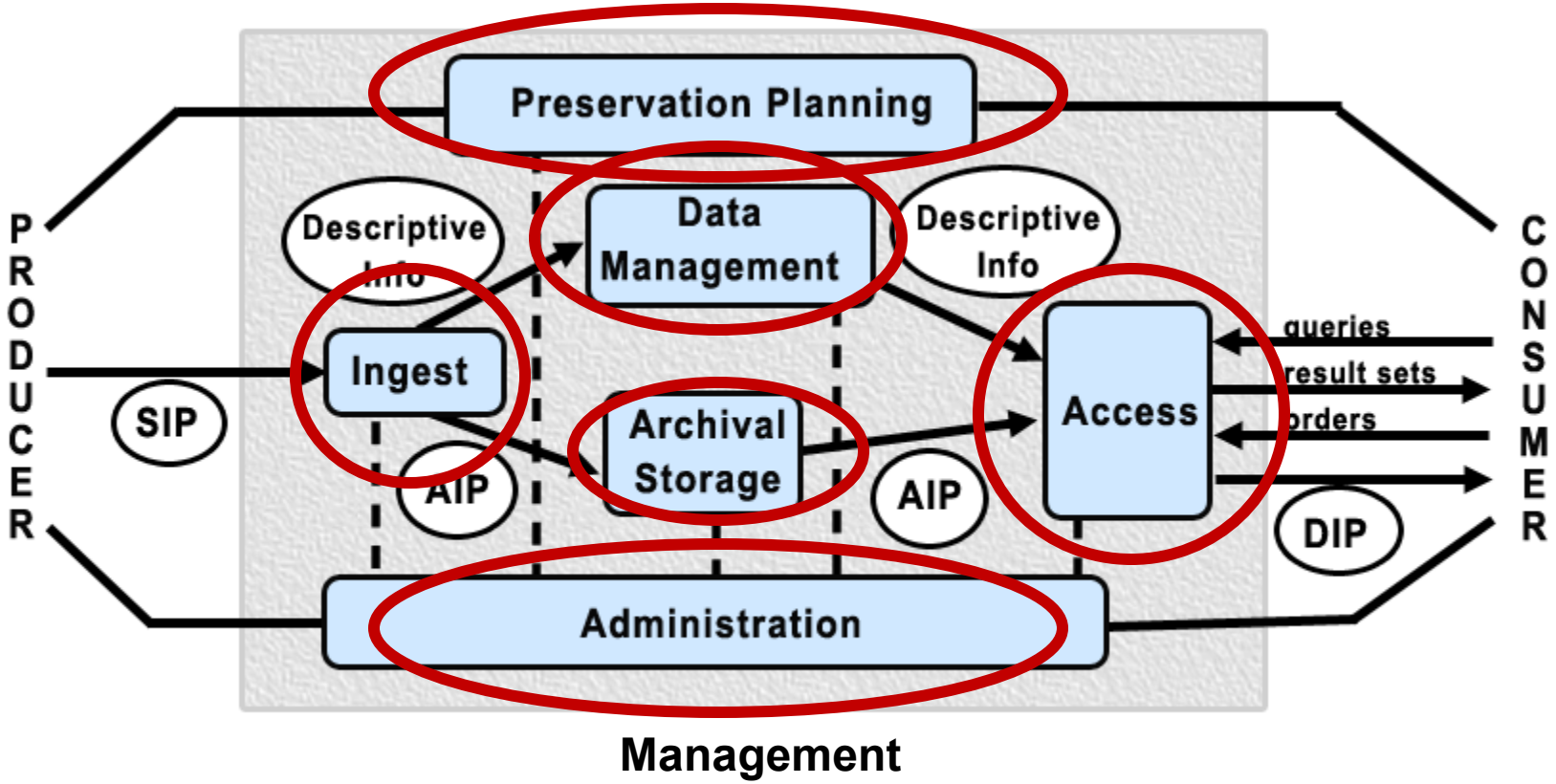


Management

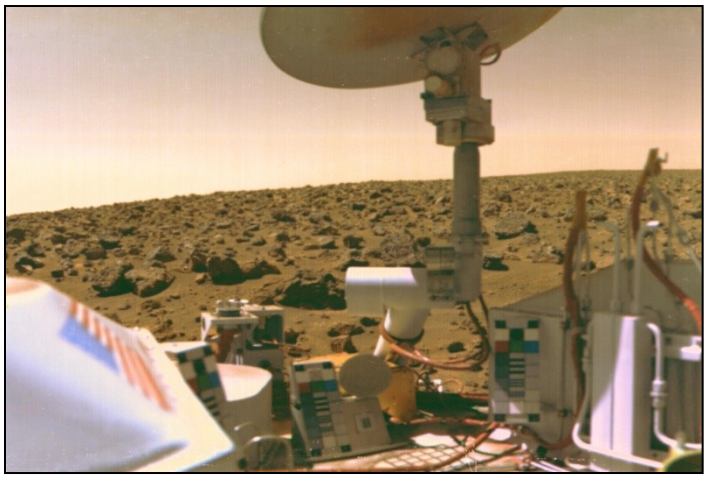
What do your producers produce?

What do your consumers consume?

Key Vocabulary: actions



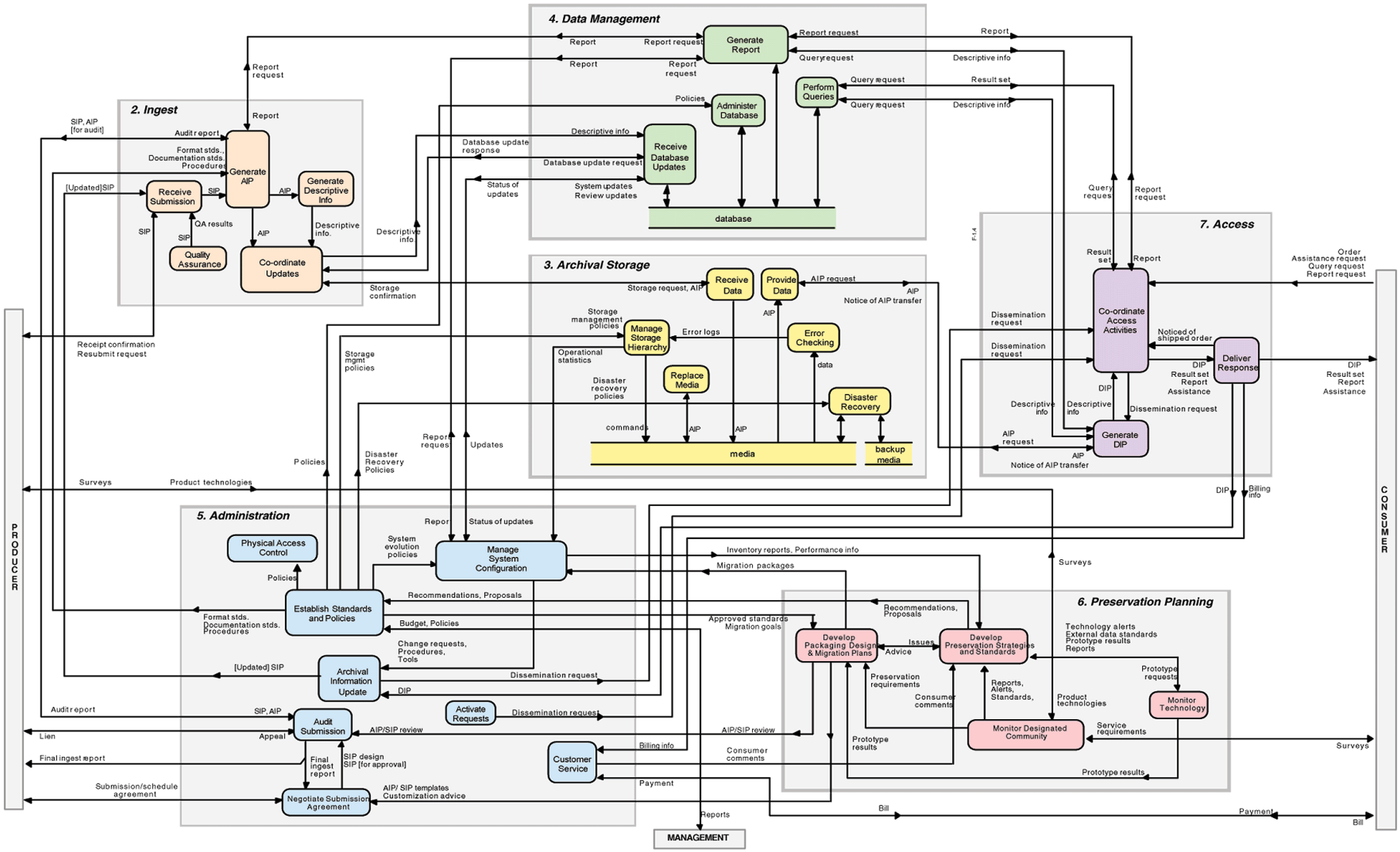
NB: not a production-line process …



Courtesy NASA/JPL-Caltech

OAIS : you need to know ...
Consultative Committee for Space Data Systems
Inadvertent comparison with NASA
Scales up really well
Scales down?
Gets very complicated!

DigitalPreservationCoalition



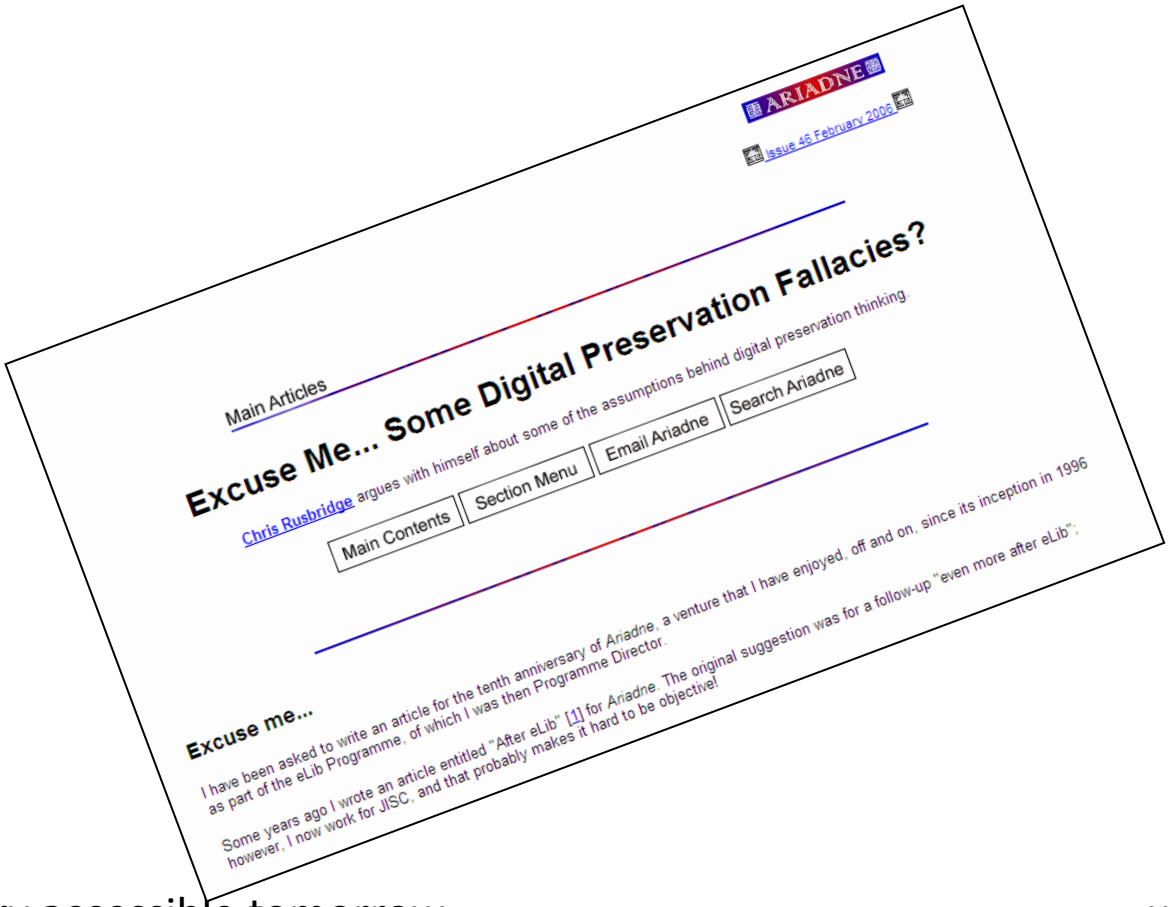
In reality?

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

Rusbridge, CR 2006 'Excuse me ... Some Digital Preservation Falacies?' in Ariadne 46, online at http://www.ariadne.ac.uk/issue46/rusbridge/



DigitalPreservationCoalition

updated for today ...

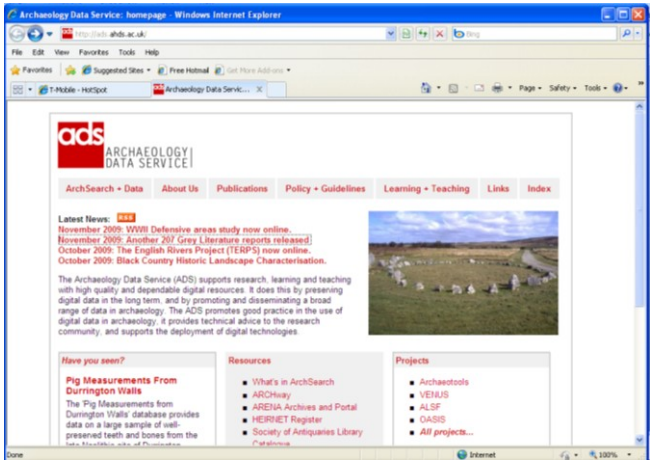
1. Providing long term access is ***expensive*** because ...
2. File ***formats are always*** *changing* and ...
3. We need to ***keep everything*** which means ...
4. You need to **do *lots of work*** that is ...
5. Closer to ***rocket science*** than conservation
6. It's ***your job*** to fix it, even if
7. ***No one will use the stuff any way***

Expensive ... eh?

Lifecycle costs of digital objects
vs
Lifecycle costs of transparencies
vs
Lifecycle costs of books
vs
Lifecycle costs of objects
vs
Lifecycle costs of archives
Vs
Lifecycles costs of historic environment

DigitalPreservationCoalition

How much does a repository cost
Here's two I prepared earlier ...



Setup:
Tens of thousands?

Setup:
Tens of millions?

DigitalPreservationCoalition

Digital preservation expensive ..?

No: it's an unfunded mandate

Don't throw money at it: get the mandate properly incorporated

DigitalPreservationCoalition

And it's your job to fix it all?

In a busy museum or archive or library ...

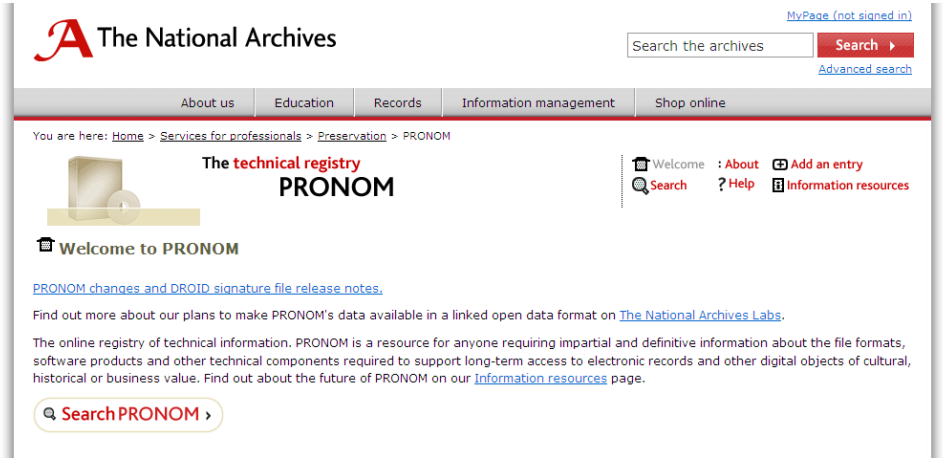
Records management
Business continuity
Communications
Photography
Oral History
Digital Art
AV interactives
Research data ...

Friends and allies ...

Data centres, DPC, others?
IT services?
Other types of alliance?
Trust and accreditation?

DigitalPreservationCoalition

And it's your job to fix it all?



http://www.nationalarchives.gov.uk/PRONOM/

For example: PRONOM
File format identification and behaviour

PLANETS Tools
PLATO Preservation Plans ...
Others ...

There but for the grace of God ....

DigitalPreservationCoalition

It's your job to fix it all ...?

No: but responsibilities need planned

There are friends, allies and tools

It might not be your job.

But it is somebody's job.

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!

William Kilbride
william @dpconline.org