



Digital **Preservation** Coalition

BRITISH
LIBRARY

Preservation
Advisory Centre

Getting Started in Digital Preservation

Wellcome Collection - 4th February 2011

#starting_dp



Digital **Preservation** Coalition

BRITISH
LIBRARY

Preservation
Advisory Centre

Obsolescence and migration

Joining up the dots between partners

Making sense of storage requirements

Persuading people that it matters

Business case – and funding development

Access to funds at the right time – sustainability

Preserving stuff that turns up



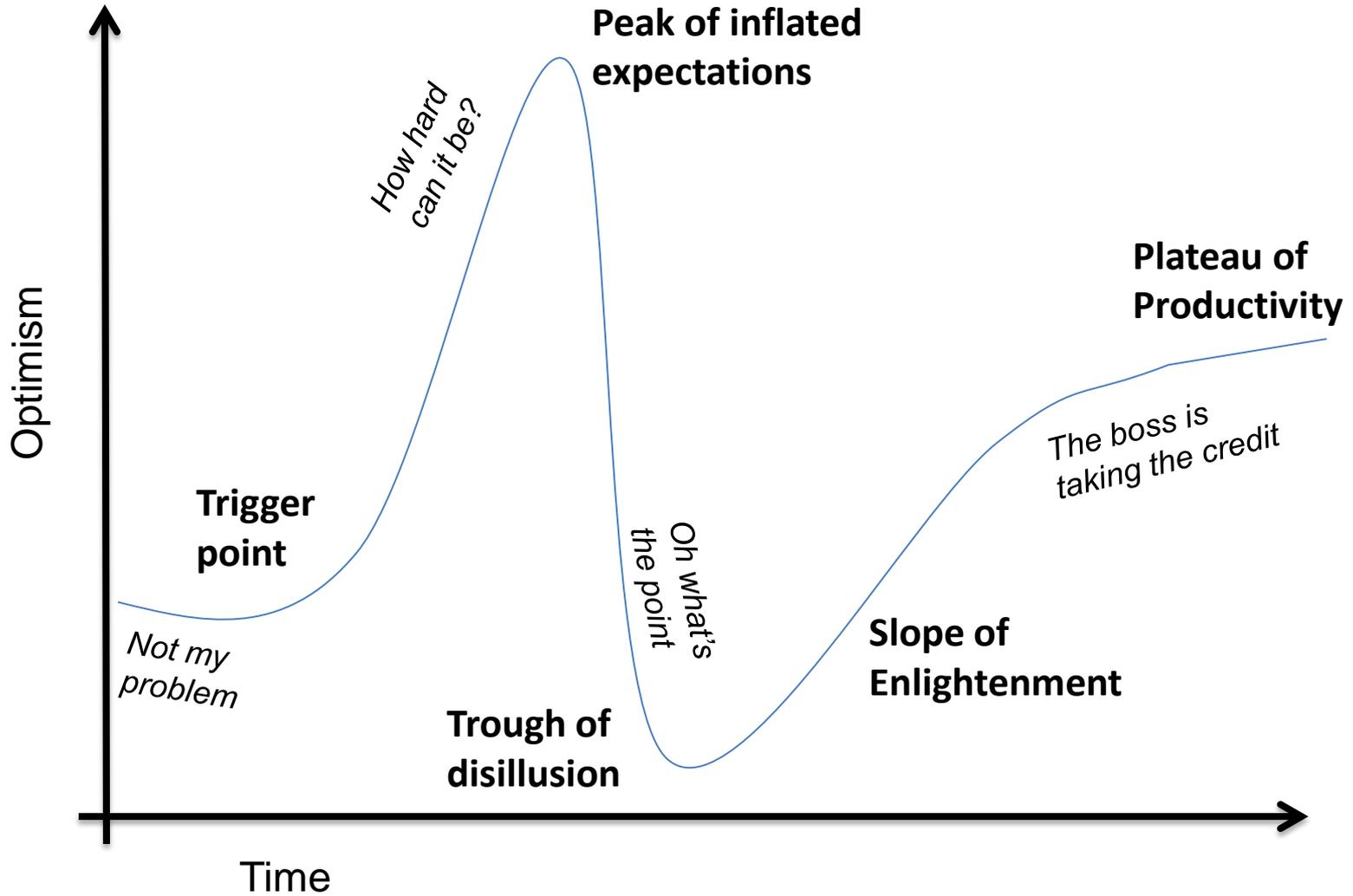
Digital **Preservation** Coalition

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!

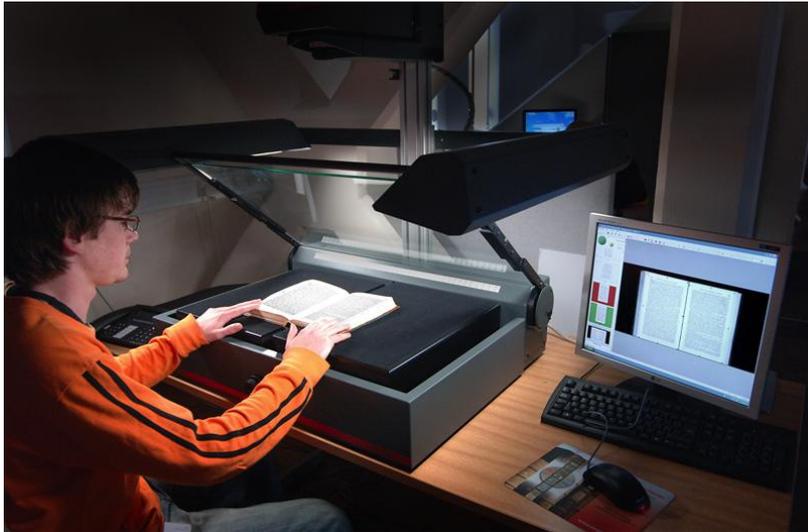
William Kilbride
william @dpconline.org





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



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

‘In perpetuity’

‘Indefinitely’

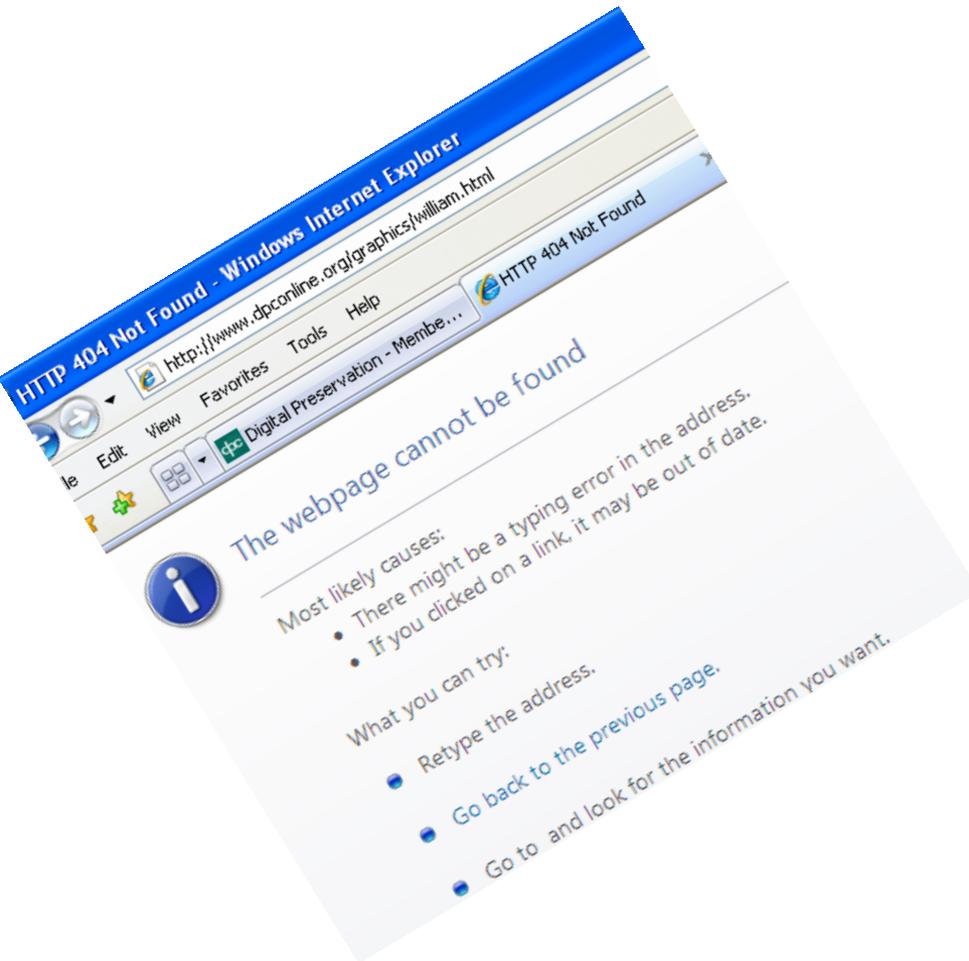
‘50 years’

‘10 years then elsewhere’

‘until 2014’

‘forever or for three years’

DPC/Portico/ULCC 2010



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

(Spencer et al 2009)



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.



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.





Challenge 3

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

Storage media can be refreshed and can self-check.

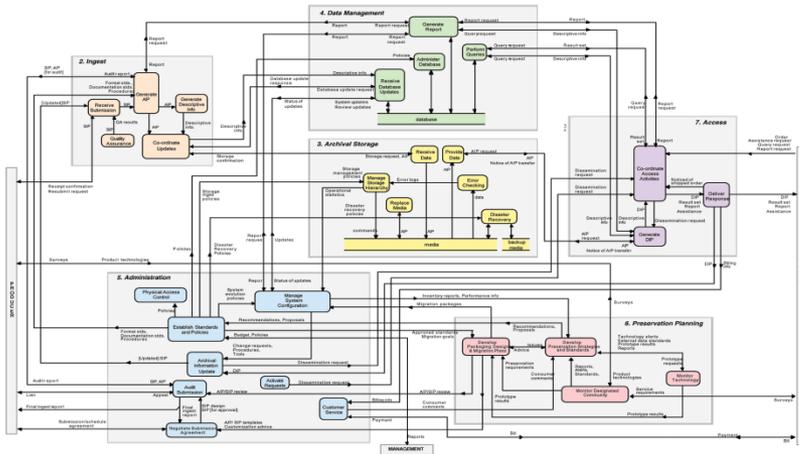
Storage densities continue to improve offering greater capacity at less cost.

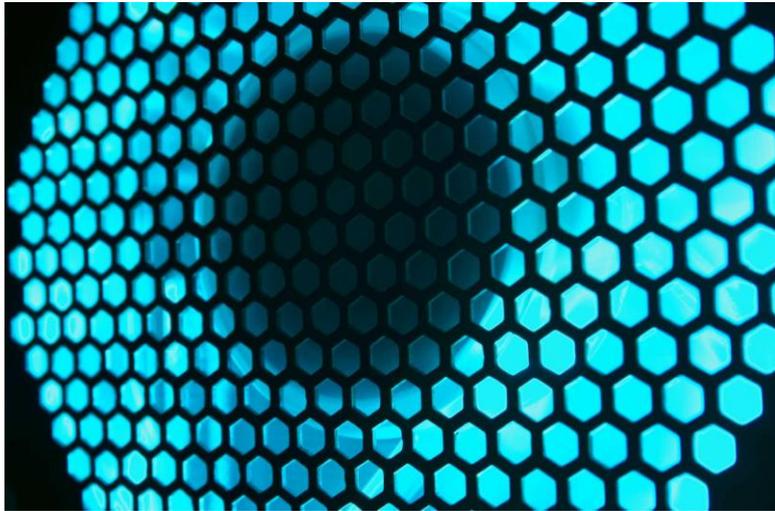
Challenge 4

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.





Challenge 5

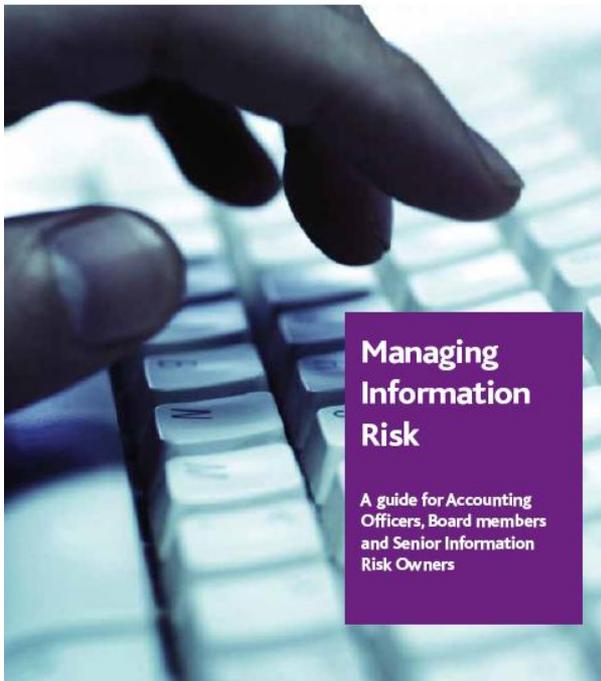
Digital resources can be altered, corrupted or deleted without obvious detection.

Signatures and wrappers can safeguard authenticity

Security can control access.

Copies are perfect replicas with no degradation.

HM Government



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.

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

Challenge 8

We have limited experience.

Rapid churn in technology
accelerates our research

Transformed in last decade

This is a shared problem





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** Coalition



Key Terms and Functional Relationships

Reference Model for an Open Archival Information System 'OAIS'

Submit

Archive

Disseminate

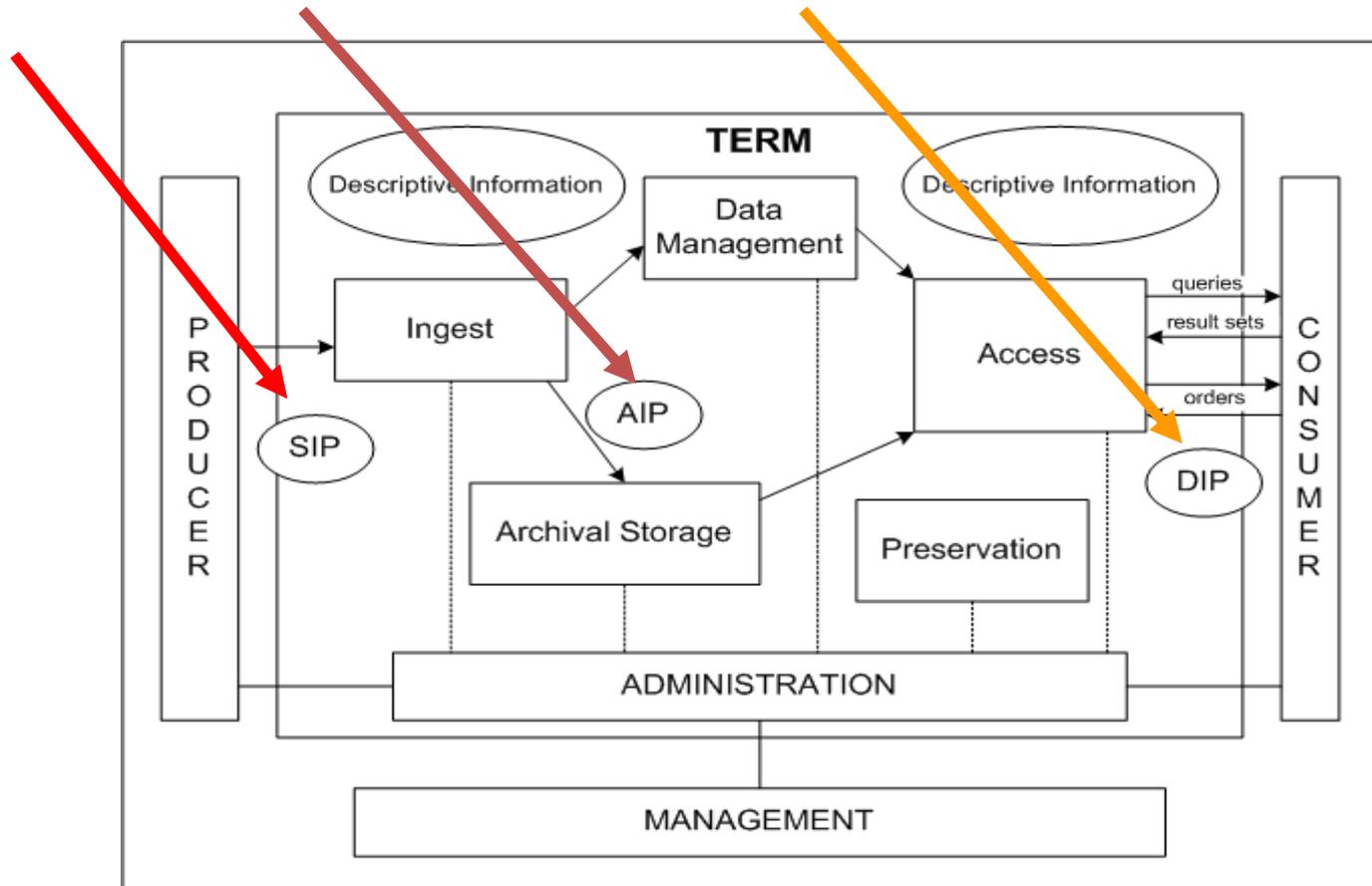
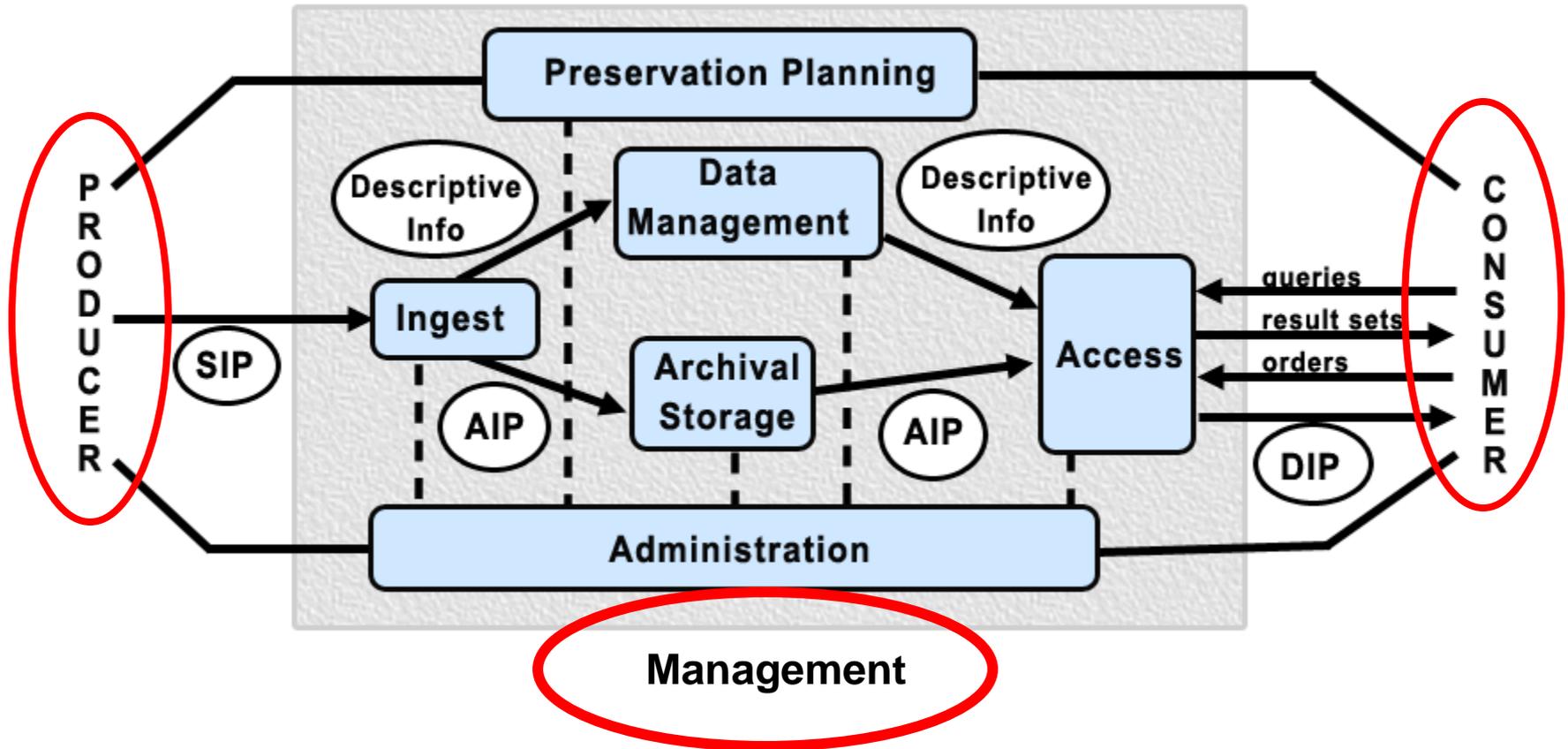


Fig. 1. Major functions of the OAIS Reference Model from Consultative Committee for Space Data Systems (CCSDS), CCSDS 650.0-W-1, Producer-Archive Interface Methodology Abstract Standard, (OAIS), White Book, Issue 1, Draft Recommendation for Space Data System Standards.

Picture from DLib

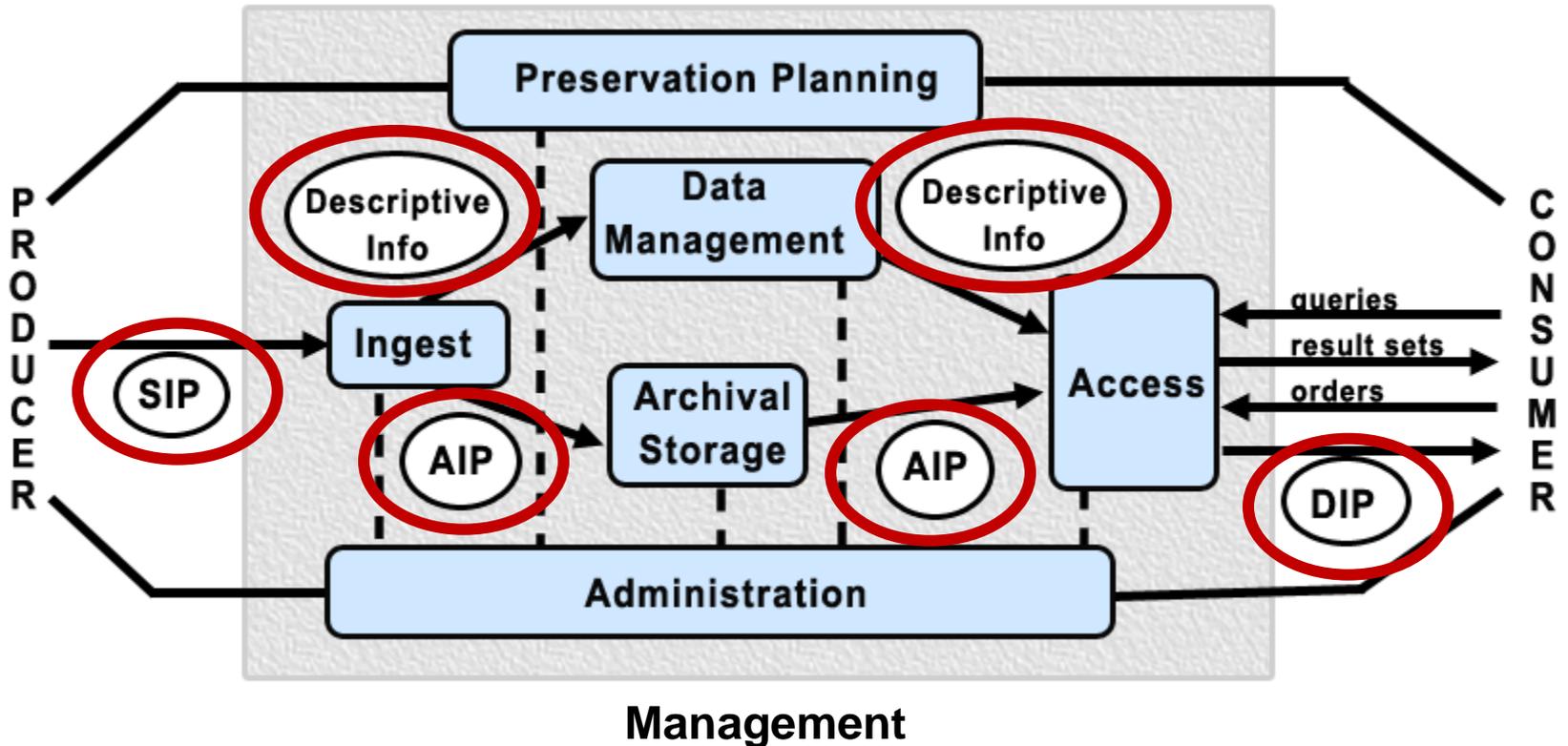
Key Vocabulary: the actors



Who are your producers?

Who are your consumers?

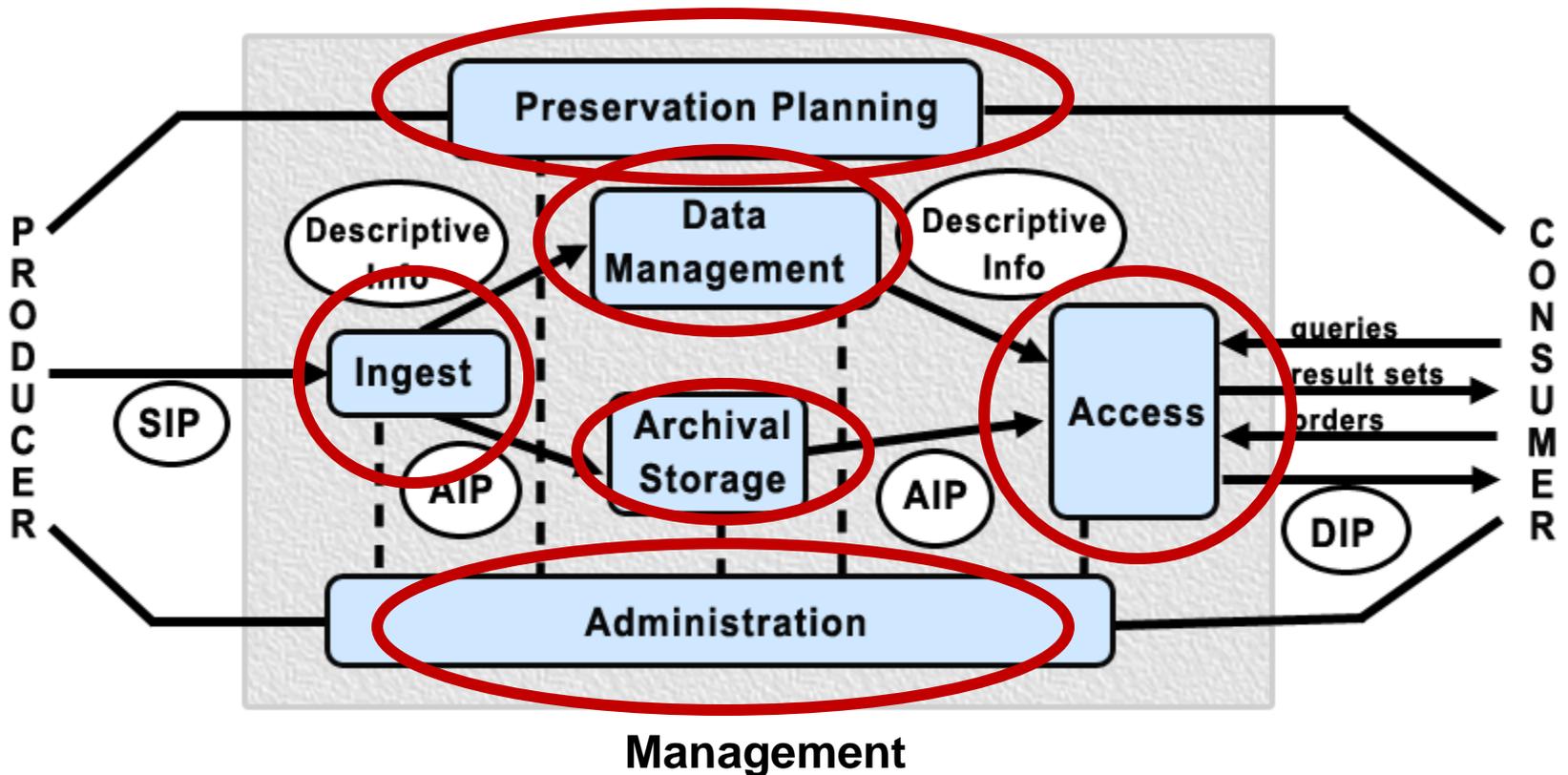
Key Vocabulary: the objects



What do your producers produce?

What do your consumers consume?

Key Vocabulary: actions

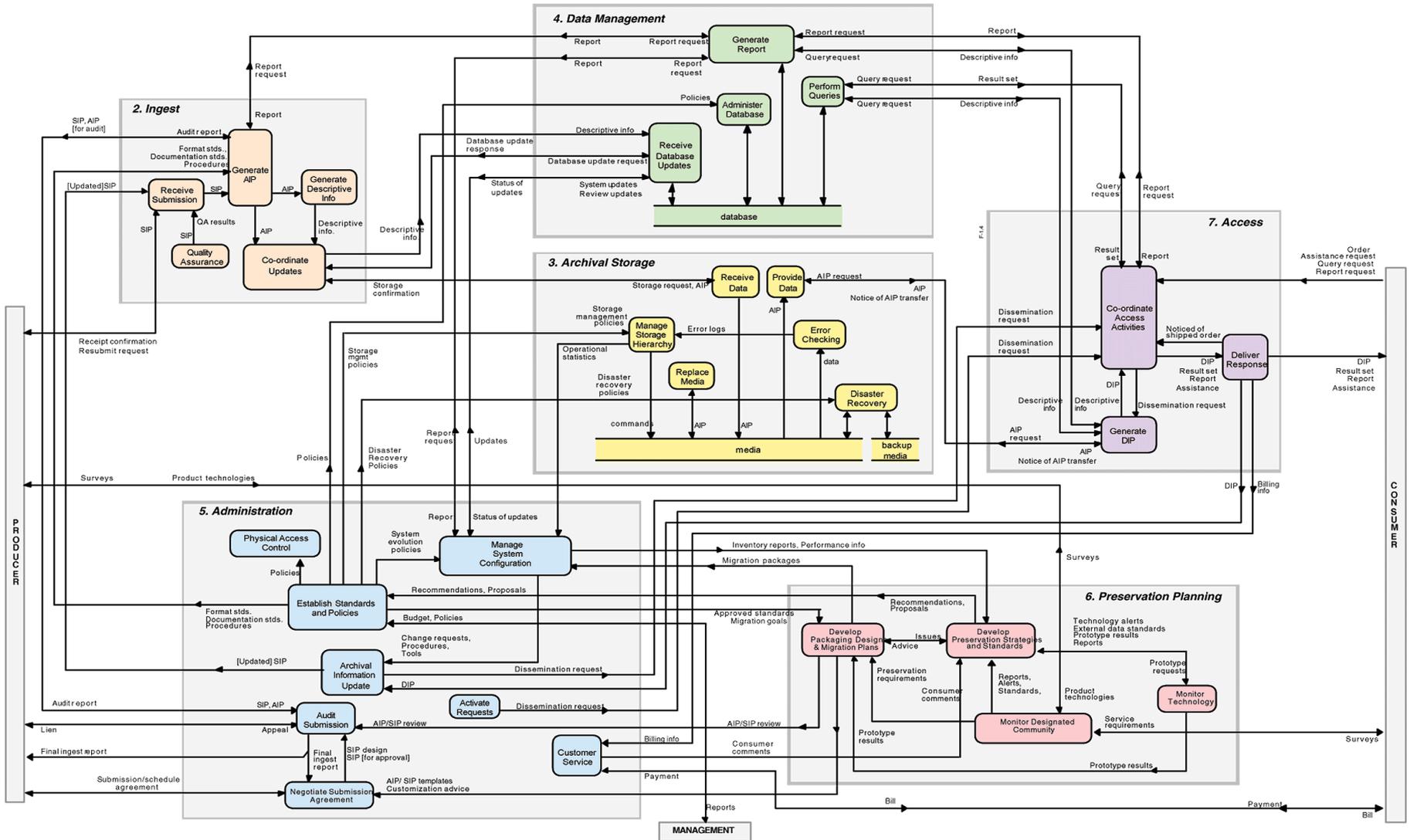


NB: not a production-line process ...



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!

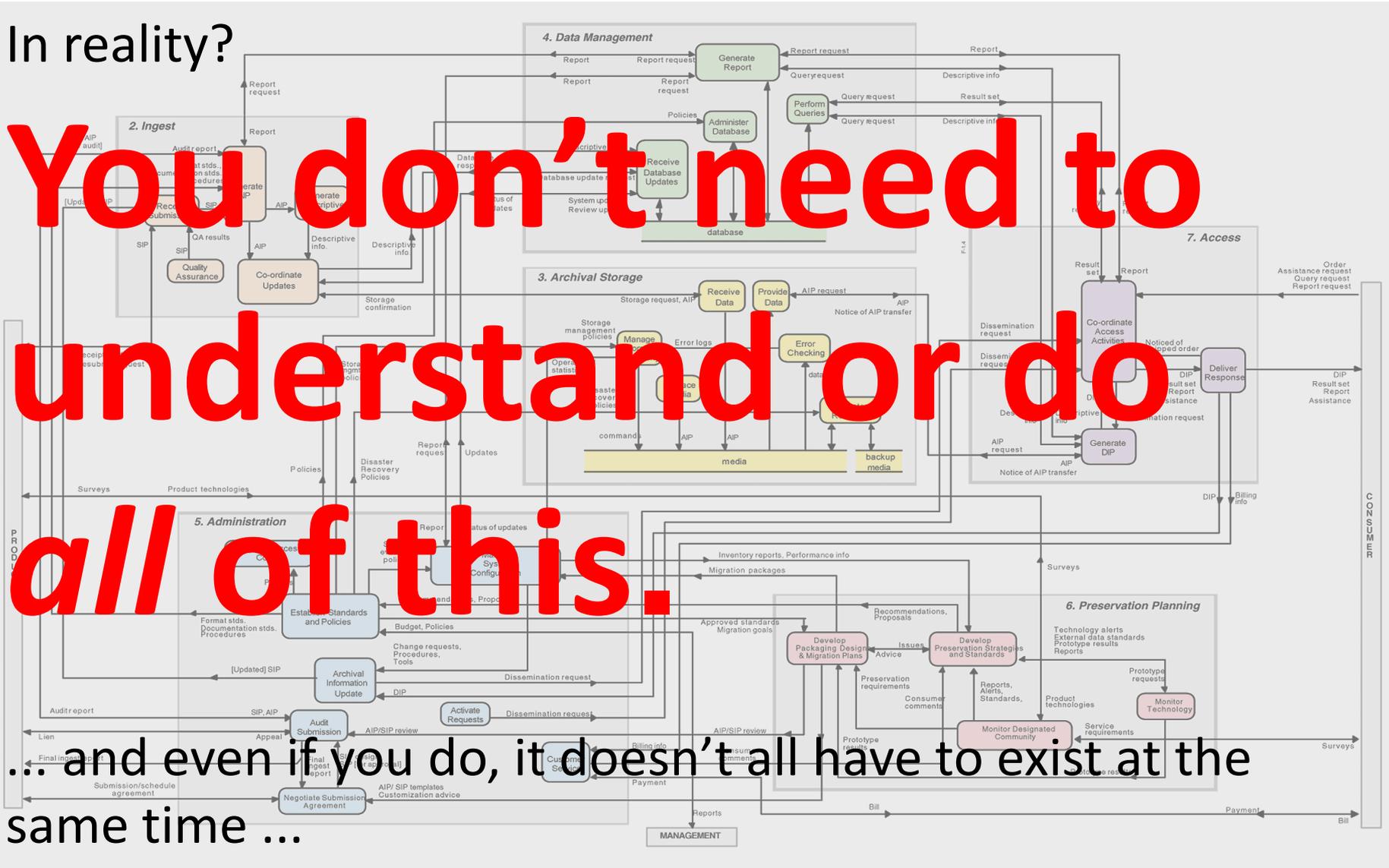




In reality?

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 ...





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

The screenshot shows the article page for 'Excuse Me... Some Digital Preservation Falacies?' by Chris Rusbridge, published in Ariadne 46, February 2006. The page includes a navigation menu with links for 'Main Contents', 'Section Menu', 'Email Ariadne', and 'Search Ariadne'. The article text begins with 'Excuse me...' and discusses the author's involvement with the eLib Programme and the journal Ariadne.

Excuse Me... Some Digital Preservation Falacies?

[Chris Rusbridge](#) argues with himself about some of the assumptions behind digital preservation thinking.

[Main Contents](#) | [Section Menu](#) | [Email Ariadne](#) | [Search Ariadne](#)

Excuse me...

I have been asked to write an article for the tenth anniversary of Ariadne, a venture that I have enjoyed, off and on, since its inception in 1996 as part of the eLib Programme, of which I was then Programme Director.

Some years ago I wrote an article entitled "After eLib" [1] for Ariadne. The original suggestion was for a follow-up "even more after eLib"; however, I now work for JISC, and that probably makes it hard to be objective!



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

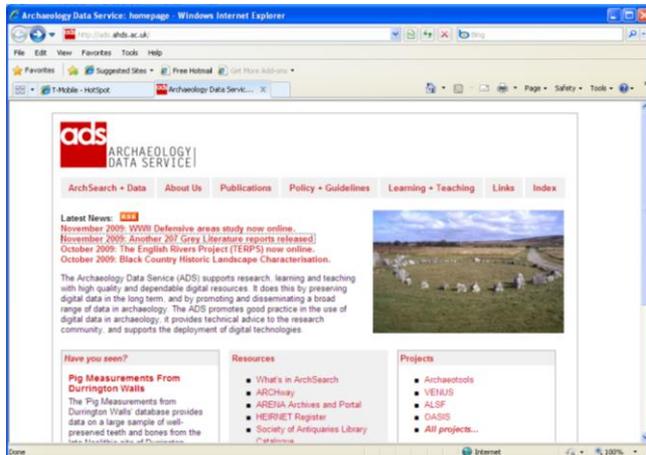
vs

Lifecycle costs of archives

Vs

Lifecycles costs of historic environment

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



**Setup:
Tens of thousands?**



**Setup:
Tens of millions?**



Digital **Preservation** Coalition

Digital preservation expensive ..?

No: it's an unfunded mandate

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



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 ...

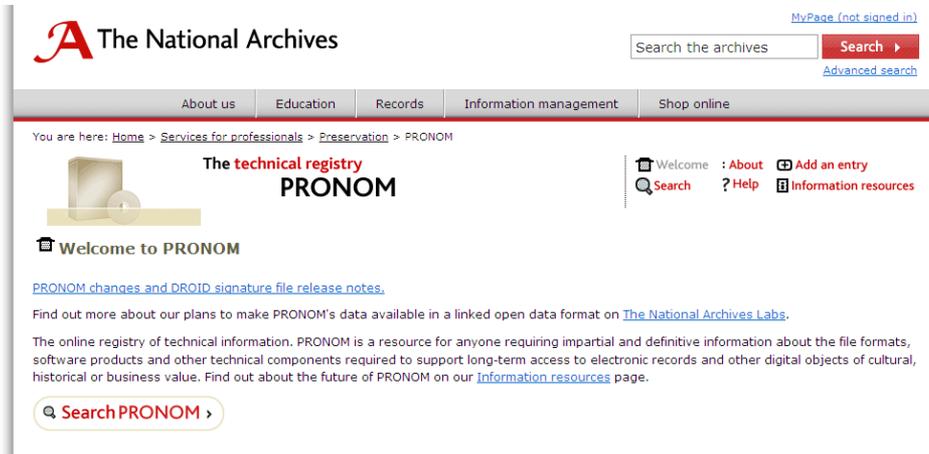
Data centres, DPC, others?

IT services?

Other types of alliance?

Trust and accreditation?

And it's your job to fix it all?



The screenshot shows the PRONOM page on the National Archives website. The page title is "The technical registry PRONOM". It includes a search bar, navigation links, and a "Search PRONOM" button. The page content mentions "PRONOM changes and DROID signature file release notes" and "The online registry of technical information. PRONOM is a resource for anyone requiring impartial and definitive information about the file formats, software products and other technical components required to support long-term access to electronic records and other digital objects of cultural, historical or business value."

<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



Digital **Preservation** Coalition

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.



Digital **Preservation** Coalition



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