DigitalPreservationCoalition

Our digital memory tomorrow

what we've learned about digital access and preservation and how it's useful now and in the future

Access and preservation are almost the same thing:

No long term access without preservation

No use to preservation without some access

What is the question?

6 basic challenges and some solutions

3 lessons from experience

Some emerging trends …

www.dpconline.org

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

Digital preservation typically makes bleak reading …

ads ARCHAEOLOGY DATA SERVICE

Digital Archives

Strategies for Digi

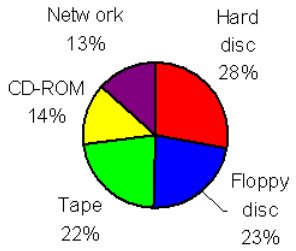
Findings and Recommendations f
Archaeology: A Survey of

Cadw DoENI English Heritage The Heritage Council Historic Scotland RC

ads ARCHAEOLOGY DATA SERVICE

Digital Archives

The present state of digital archiv



Th
ce

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

Digital preservation typically makes bleak reading 2

<Enter your details here>

……

……

……

DigitalPreservationCoalition

Let's restate the problem …

- 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

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

What that means for business

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

DigitalPreservationCoalition

What that means for the public sector

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, social knowledge

4. Wealthier

e.g. safe markets, efficient business, management of IP, employment, planning

5. Healthier

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

6. Greener

e.g. environmental policy development, efficient retention

DigitalPreservationCoalition

Key responses

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

6 Challenges and how they have been tackled

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

DigitalPreservationCoalition

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

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

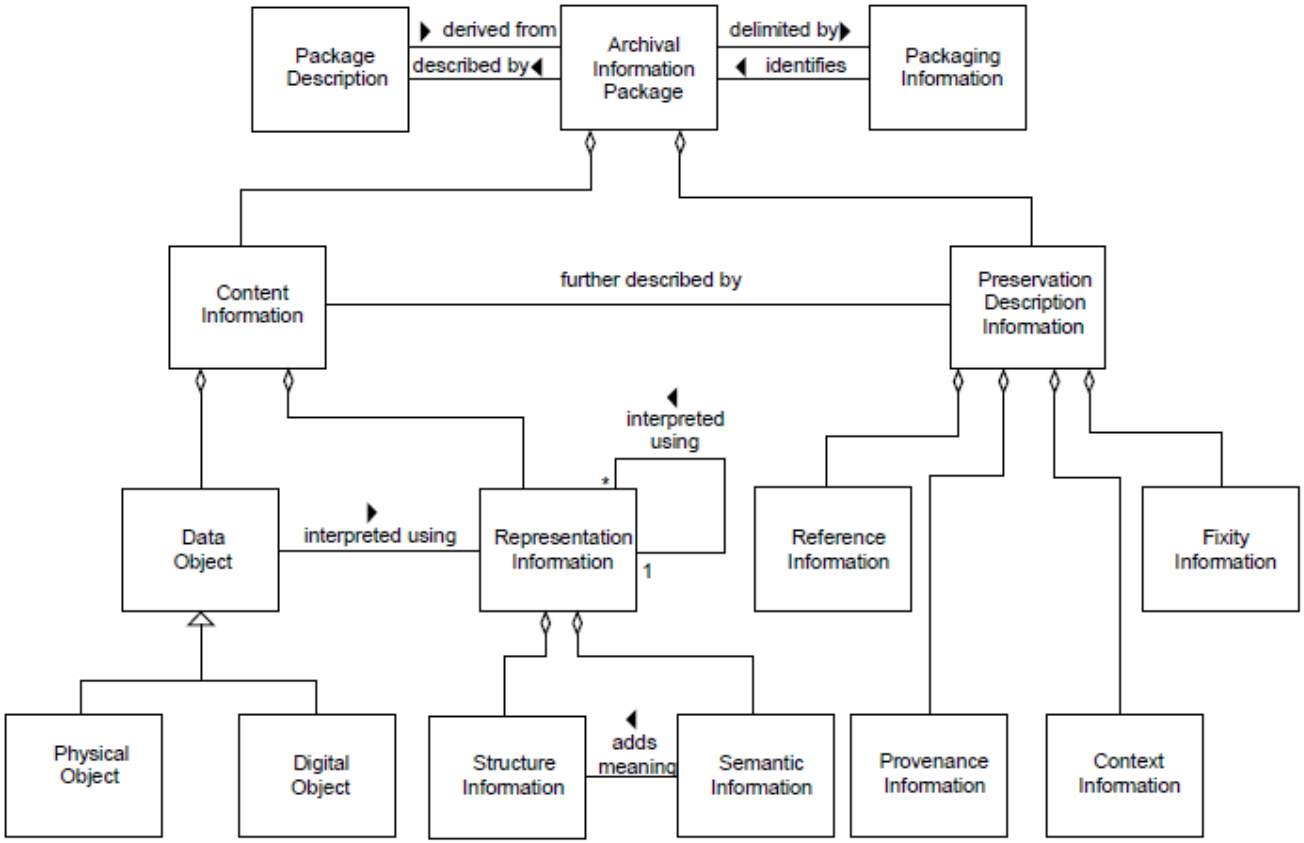
Digital preservation systems are subject to the same obsolescence as the objects they safeguard.

… so we need systems which are modular, based on standards and which are tested

DigitalPreservationCoalition

Challenge 4

Standards … OAIS



DigitalPreservationCoalition

Archive **Disseminate**

Submit

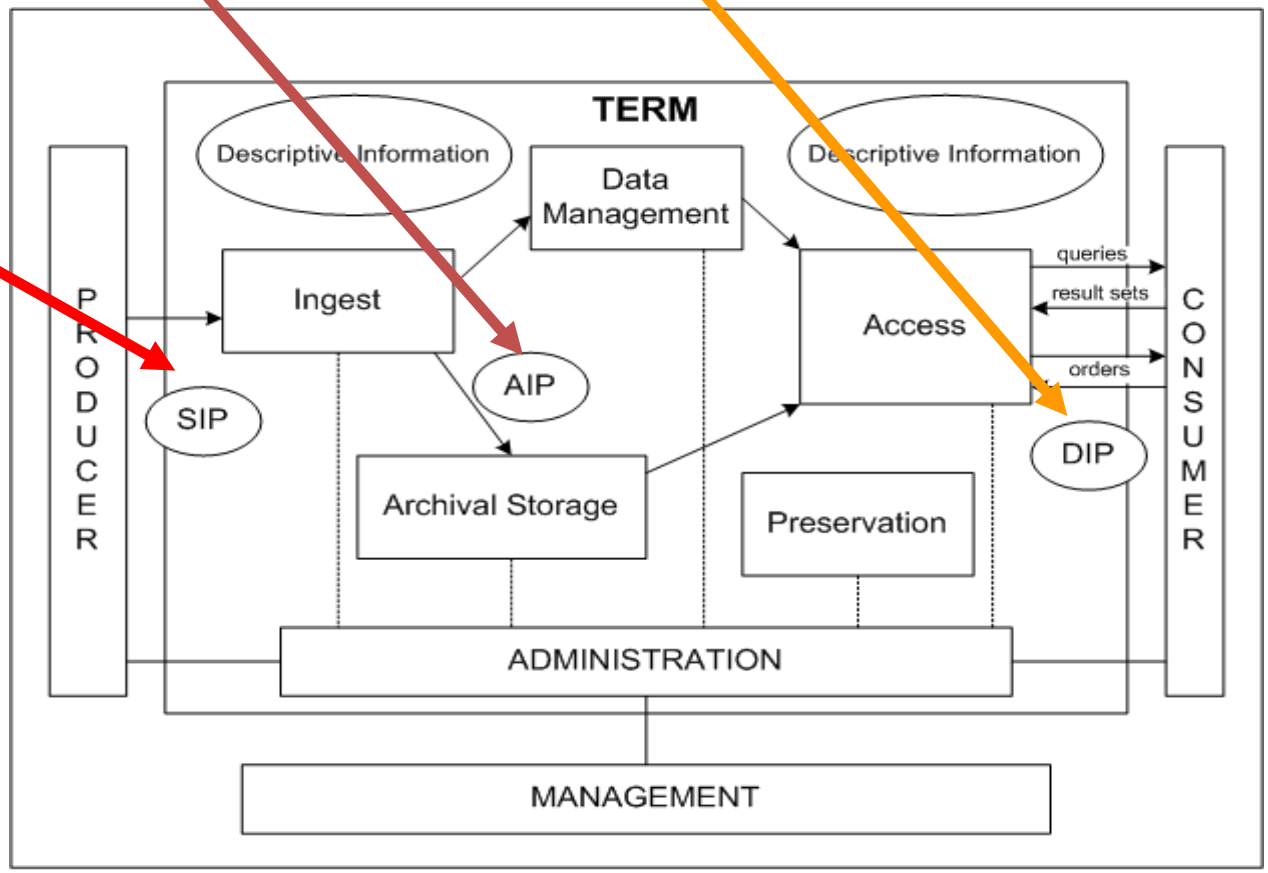


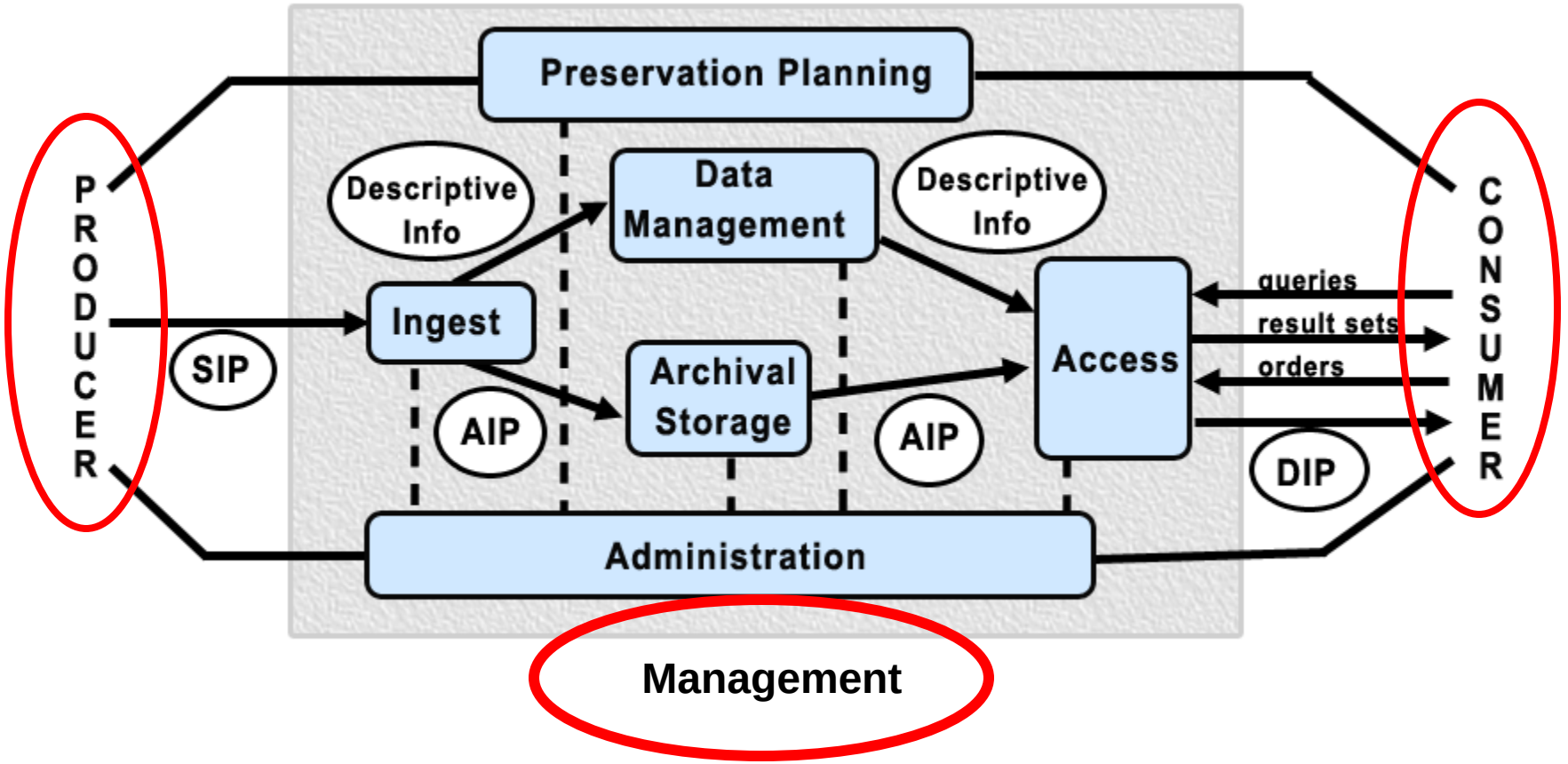
Fig. 1. Major functions of the OAIS Reference Model from Consultative Committee for Space Data Systems (CCSDS), CCSDS 650.0-W-1, Producer-Archive Interface Methodology Abstract Standard. (OAIS). White Book. Issue 1. Draft Recommendation for Space Data System Standards.

Picture from DLib

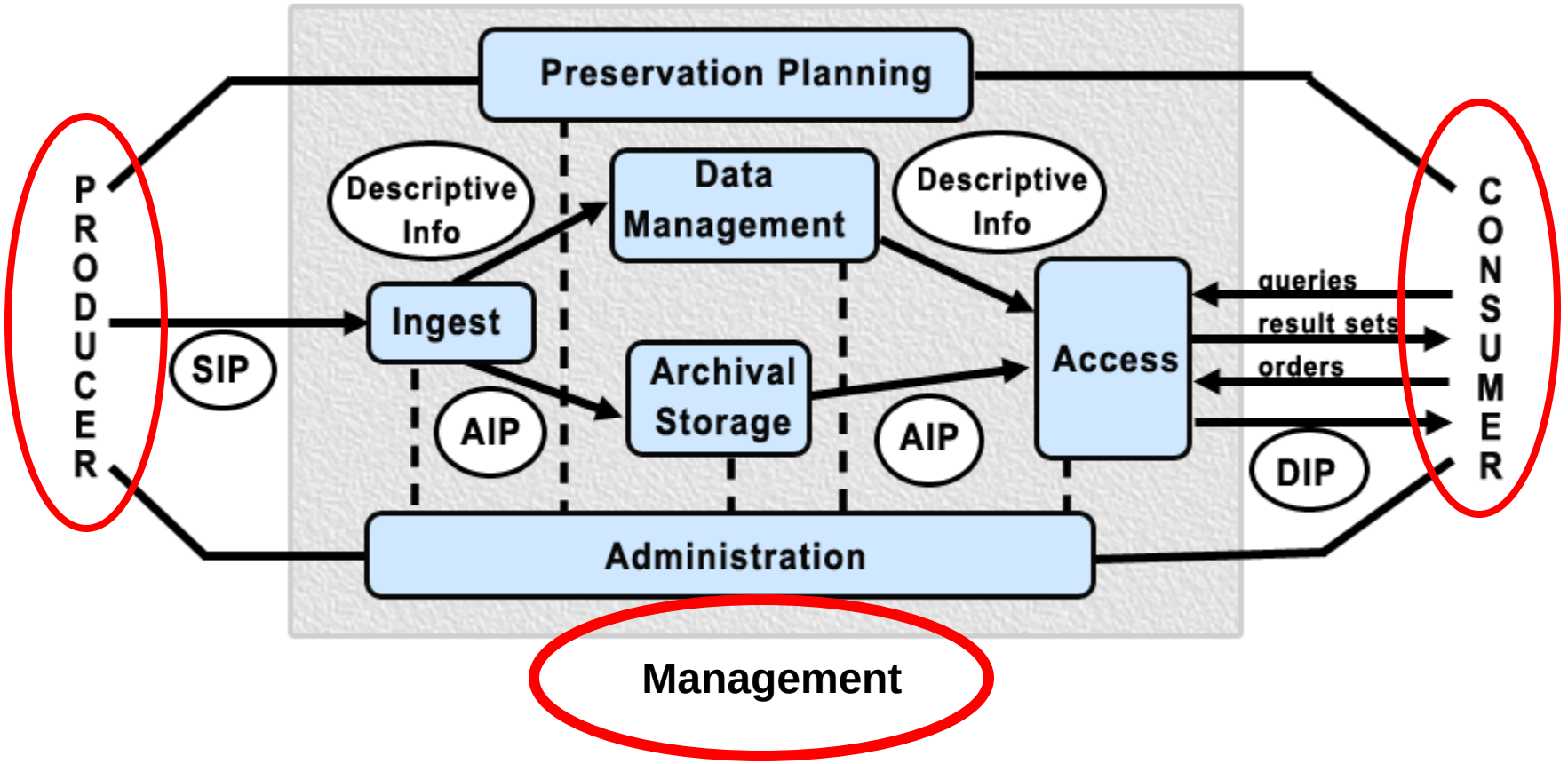
OAIS: Core Functions

- **Negotiate with for appropriate content**
- **Obtain sufficient control**
- **Determine the scope of the community**
- **Ensure independent utility of data**
- **Follow procedures for preservation**
- **Disseminate data to community**

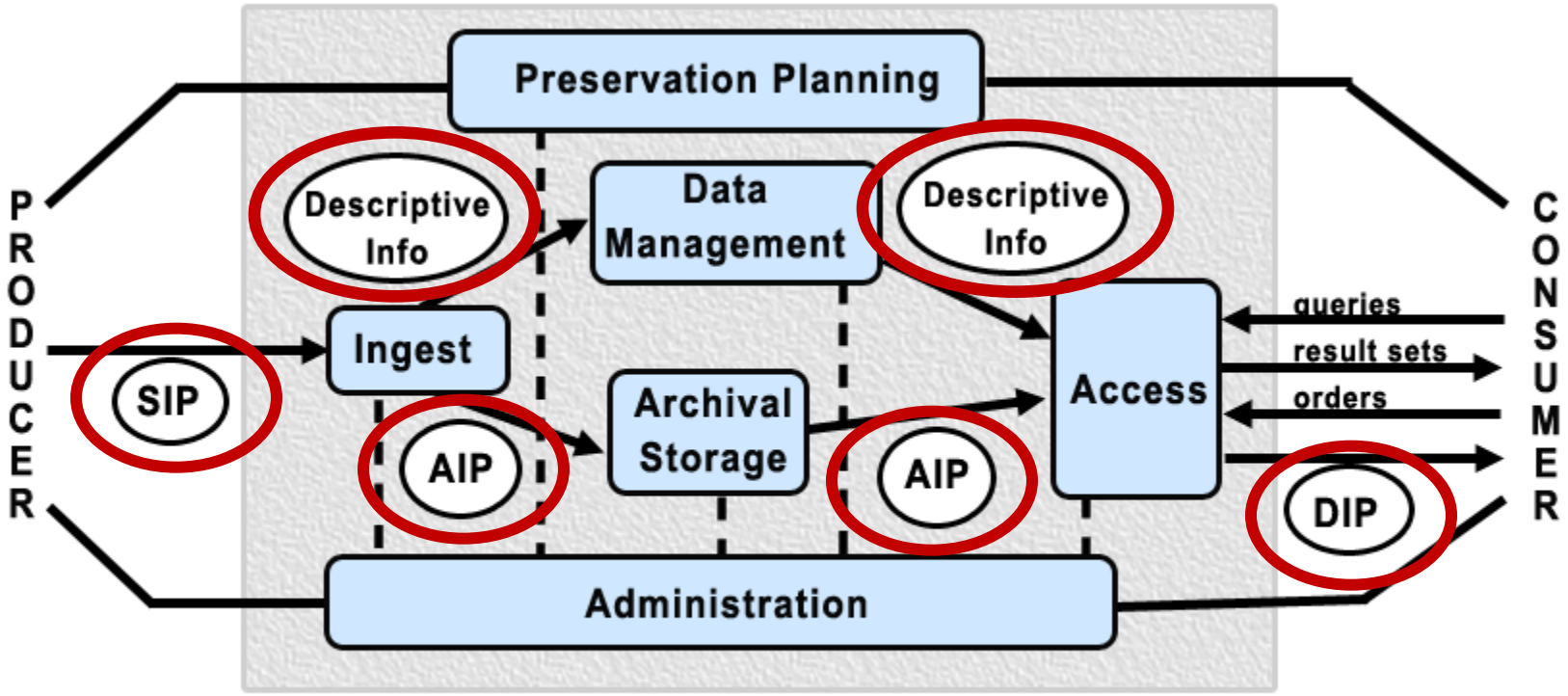
Diagramarama: the actors



Diagramarama: the actors

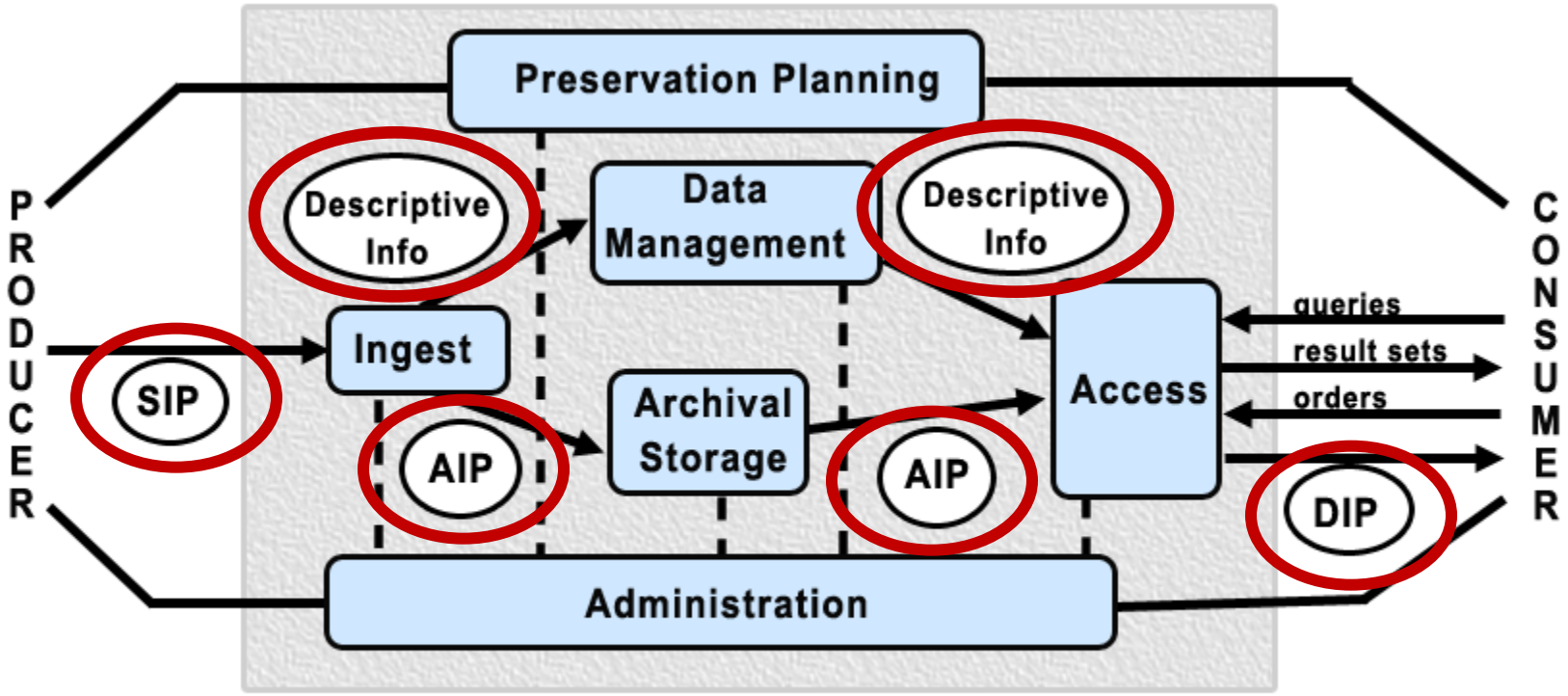


Diagramarama: the objects



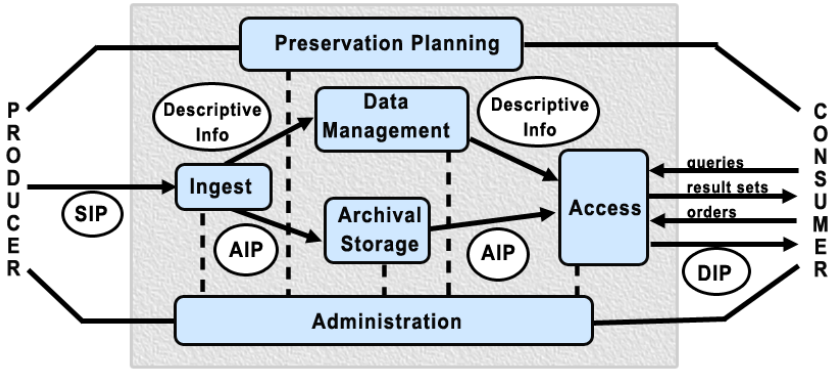
Management

Diagramarama: the objects



Management

Diagramarama: 4 things not necessarily obvious from the diagrams…



1. not a production-line process …
2. SIP:AIP:DIP not necessarily 1:1:1 relationship
3. not necessarily all one agency
4. not all of this has to exist simultaneously

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 to fixity and authenticity checks

DigitalPreservationCoalition

Challenge 6

Fixity and authenticity

A variety of solutions:

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

DigitalPreservationCoalition

3 things we've learned

DigitalPreservationCoalition

The issues are more subtle than we realised a decade ago… three examples

- File format obsolescence?
- Costs of preservation?
- How hard is preservation?

DigitalPreservationCoalition

File formats can be a problem

- Changing file formats?
- Conformant containers?
- Real life is messy

- Format registries and characterisation tools (e.g. PRONOM + DROID)
- Forensic tools and digital archaeology

DigitalPreservationCoalition

How to pick a winner ...

Adoption
Dependency
Disclosure
Transparency
Metadata support
Interoperability
Complexity
Stability
Rights management

Todd, M 2009 'File formats for preservation', DPC Technology Watch Report 02/09, online at http://www.dpconline.org/advice/technology-watch-reports.html

DigitalPreservationCoalition

How to pick a winner ...

beyond and **potentially over-writing the criteria** ... repository managers should **align** the recognition and weighting of criteria with a **clear preservation strategy** that articulates the **purpose** of the repository and the **needs of its designated community**;

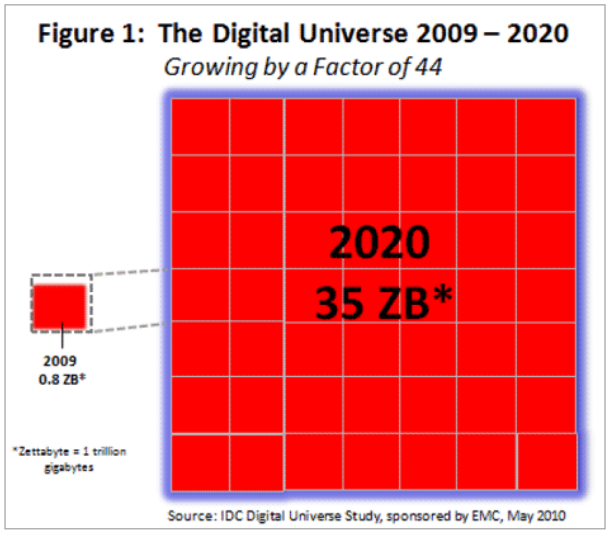
Todd, M 2009 'File formats for preservation', DPC Technology Watch Report 02/09, online at http://www.dpconline.org/advice/technology-watch-reports.html

DigitalPreservationCoalition

How to pick a winner ...

‘Digital Universe’ Nears A Zettabyte

May 4th, 2010 : Rich Miller



The Great Recession hasn’t slowed the breakneck growth of the Digital Universe. In 2010 the volume of digital information created and duplicated in a year will reach 1.2 zettabytes, according to new data from IDC.

You ain’t seen nothing yet

Data growth on 3 axes

- volume
- complexity
- Value

The ‘file’ is not the only unit of information

... it’s not going to be about obsolescence so much as workflow, capacity and dependencies

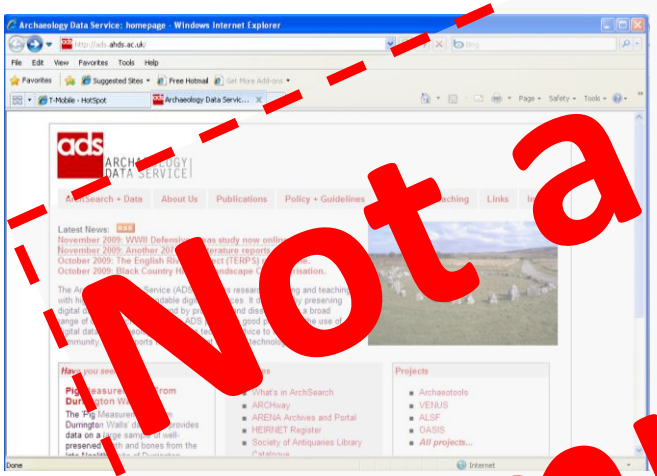
DigitalPreservationCoalition

How much does preservation cost?

Lifecycle costs of digital objects
vs
Lifecycle costs of books
vs
Lifecycle costs of museum 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 ...



Not a direct comparison

Setup:
Tens of thousands?

Setup:
Tens of millions?

http://www.life.ac.uk/

L is the complete lifecycle cost over time 0 to T. Other categories are

Aq – Acquisition,

I – Ingest,

M – Metadata,

Ac – Access,

S – Storage,

P – Preservation

http://www.life.ac.uk/

$$L_T = Aq + I_T + M_T + Ac_T + S_T + P_T$$

Lifecycle element	Acquisition	Ingest	Metadata	Access	Storage	Preservation
Element 1	Selection (Aq1)	QA (I1)	Characterisation (M1)	Reference linking (Ac1)	Bit-stream storage costs (S1)	Technology watch (P1)
Element 2	IPR (Aq2)	Deposit (I2)	Descriptive (M2)	User support (Ac2)		Preservation tool cost (P2)
Element 3	Licensing (Aq3)	Holdings update (I3)	Administrative (M3)	Access Mechanism (Ac3)		Preservation metadata (P3)
Element 4	Ordering and invoicing (Aq4)					Preservation action (P4)
Element 5	Obtaining					Quality assurance (P5)
Element 6	Check-in (Aq6)					

DigitalPreservationCoalition

Basic assumption:

Society understand the value of museums, heritage etc.

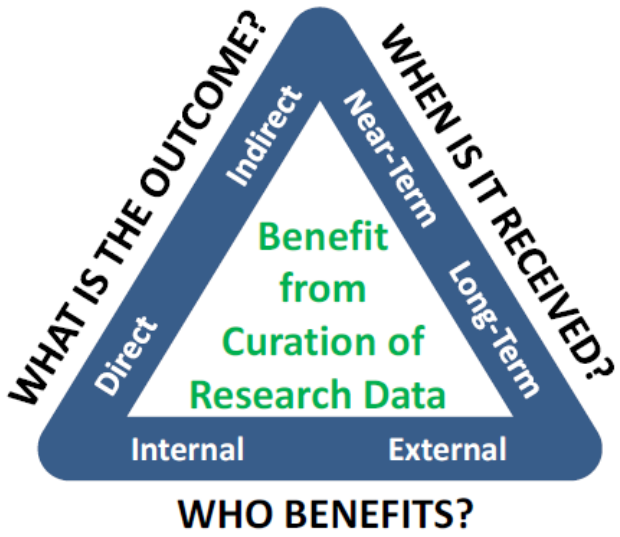
But

Don’t yet understand the value proposition of digital

DigitalPreservationCoalition

Keeping Research Data Safe 2

Anatomy of a Benefit:



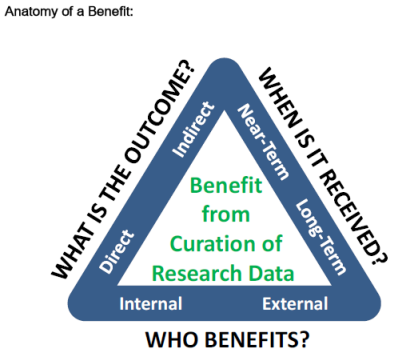
Relatively simple process:

- Identify benefits
- Categorise them
- Identify potential measures
- Illustrate value / impact

Measures benefits on three axes
Aided by list of generic benefits …

DigitalPreservationCoalition

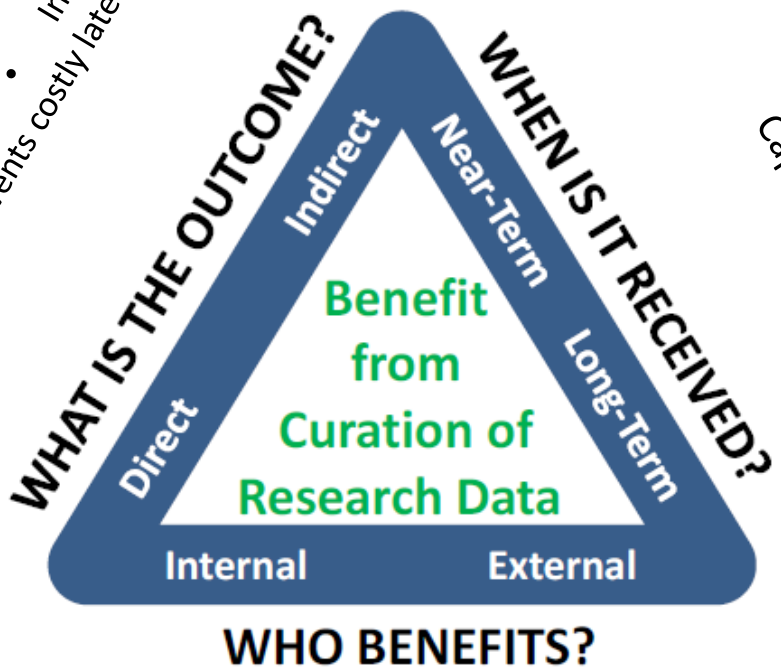
Typical examples of benefits



Reduction in costs through federated storage
Cheap early intervention prevents costly later intervention
Ensure data is accessible to other researchers
Capacity to do diachronic analyses
Reputation and citability of data
Ability of others to assess and develop your work
…

DigitalPreservationCoalition

- Direct benefits:
  Actual reductions in costs through federated storage
- Indirect benefits
  Cheap early intervention prevents costly late intervention
- Near Term Benefits:
  Ensure current data available to others
- Long term Benefits:
  Capacity to do diachronic analyses



- Internal Benefits:
  Reputation and citation of research

- External Benefits
  Ability of others to test and develop your research

www.dpconline.org

DigitalPreservationCoalition

Experience:

Organisation that have an internal digital preservation policy are more effective

than

Organisations that spend a lot of money fitfully or in responsive mode

DigitalPreservationCoalition

Digital preservation expensive ..?

No: it's an unfunded mandate

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

DigitalPreservationCoalition

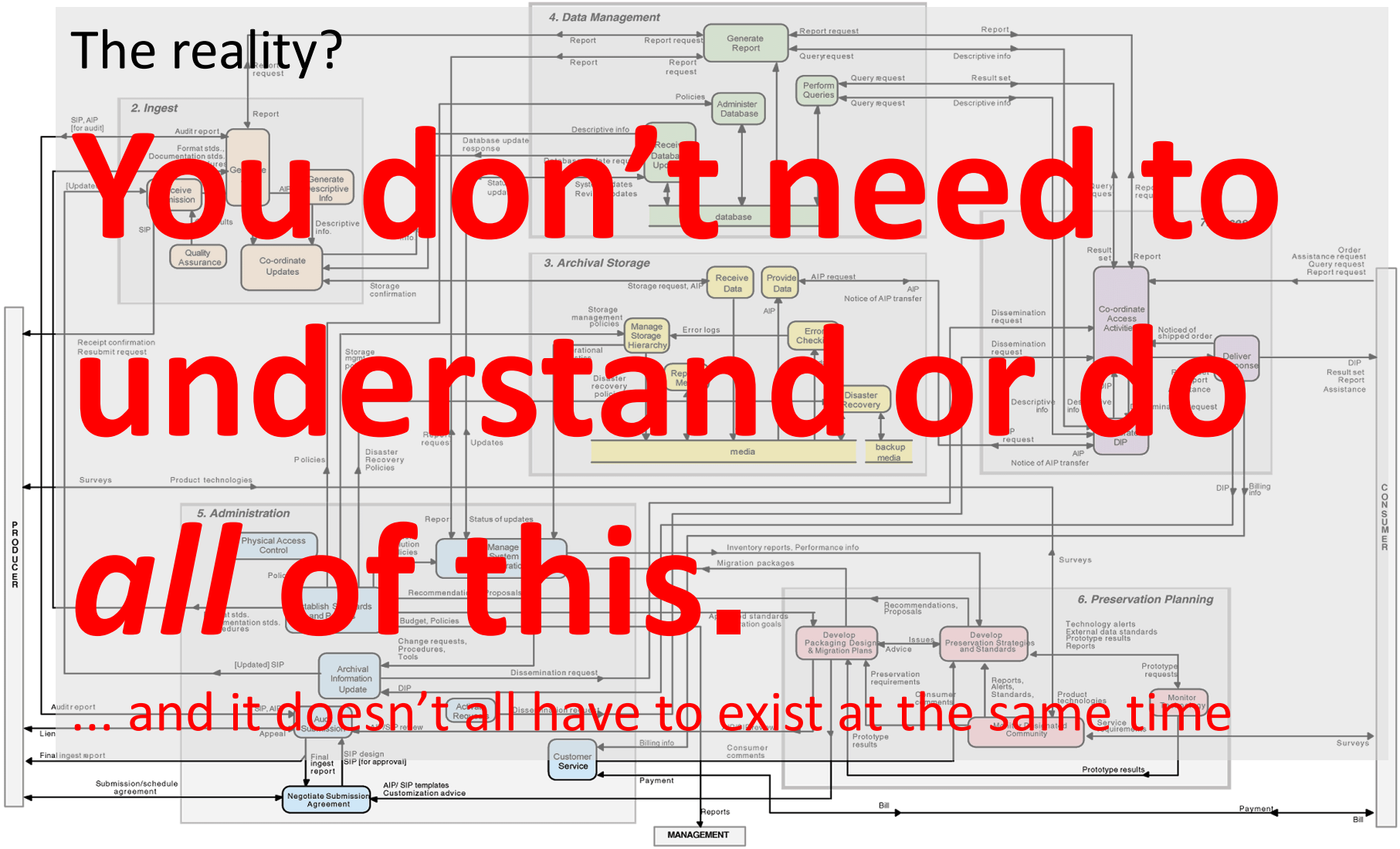
How hard can it be?

Preservation or *curation* or *longevity* or *continuity* or *resilience* or *legacy* or *permanent access* or *long-term access* or *inheritance* or *whatever term is fashionable this week* …

Daunting challenge
Decade of research and development
Replete with jargon and acronyms
Turf war between professions?
Disconnected from practice?
A whole new barrier

DigitalPreservationCoalition

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

DigitalPreservationCoalition

The last decade has shown definitively that creating complexity is not the same as solving problems

But you can (must) get started:

- *Preservation planning*
- *Risk assessment*
- *Partnerships, collaboration*
- *Policy and business case*
- *Training and awareness*

DigitalPreservationCoalition

Things to watch for …

DigitalPreservationCoalition

The blue horizon: things to watch for

- *Preservation at scale*
- *Preservation and the cloud*
- *Preservation of processes*
- *Preservation-ready objects*

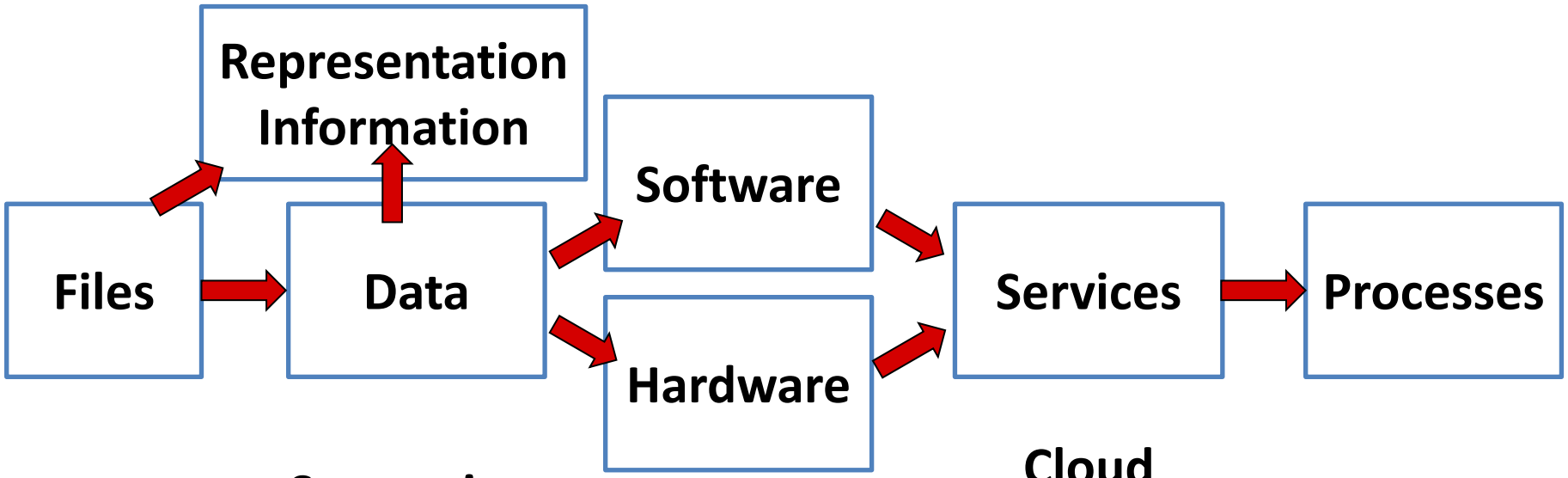
For example, TIMBUS …

TIMBUS

TIMELESS BUSINESS

Digital Preservation

Risk and Business Continuity Management



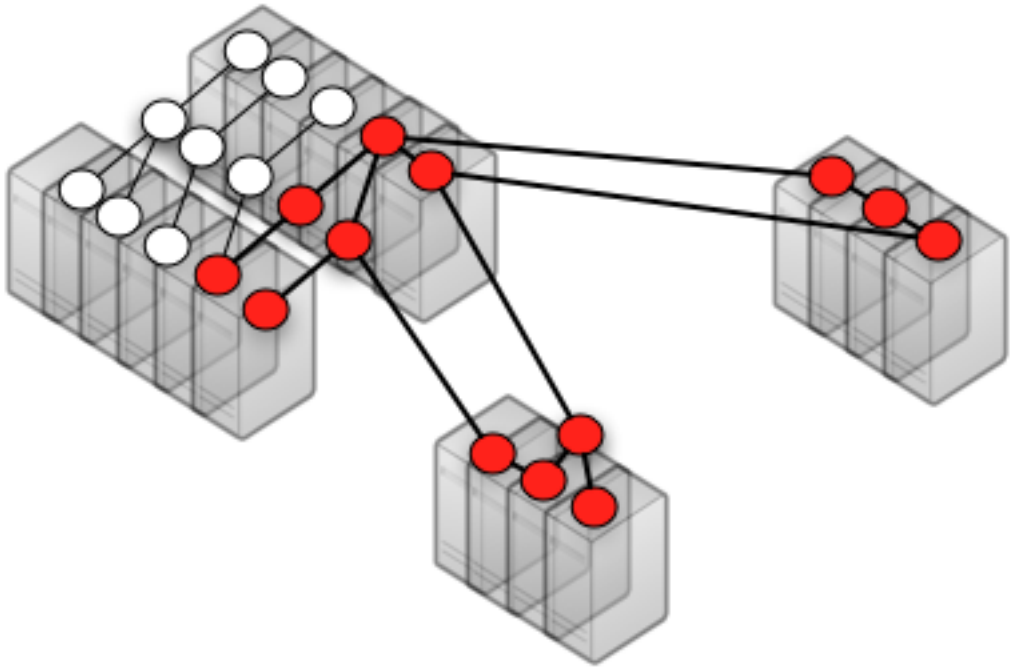
SEVENTH FRAMEWORK PROGRAMME

DigitalPreservationCoalition

TIMBUS

TIMELESS BUSINESS

1. Related services in a Business Process
2. Distributed – Public Cloud / Private Cloud / On Premise etc.



3. Risk analysis determines subset of business process must be available in >= 30 years
4. TIMBUS methods and tools used to preserve business process

SEVENTH FRAMEWORK PROGRAMME

www.dpconline.org

DigitalPreservationCoalition

TIMBUS

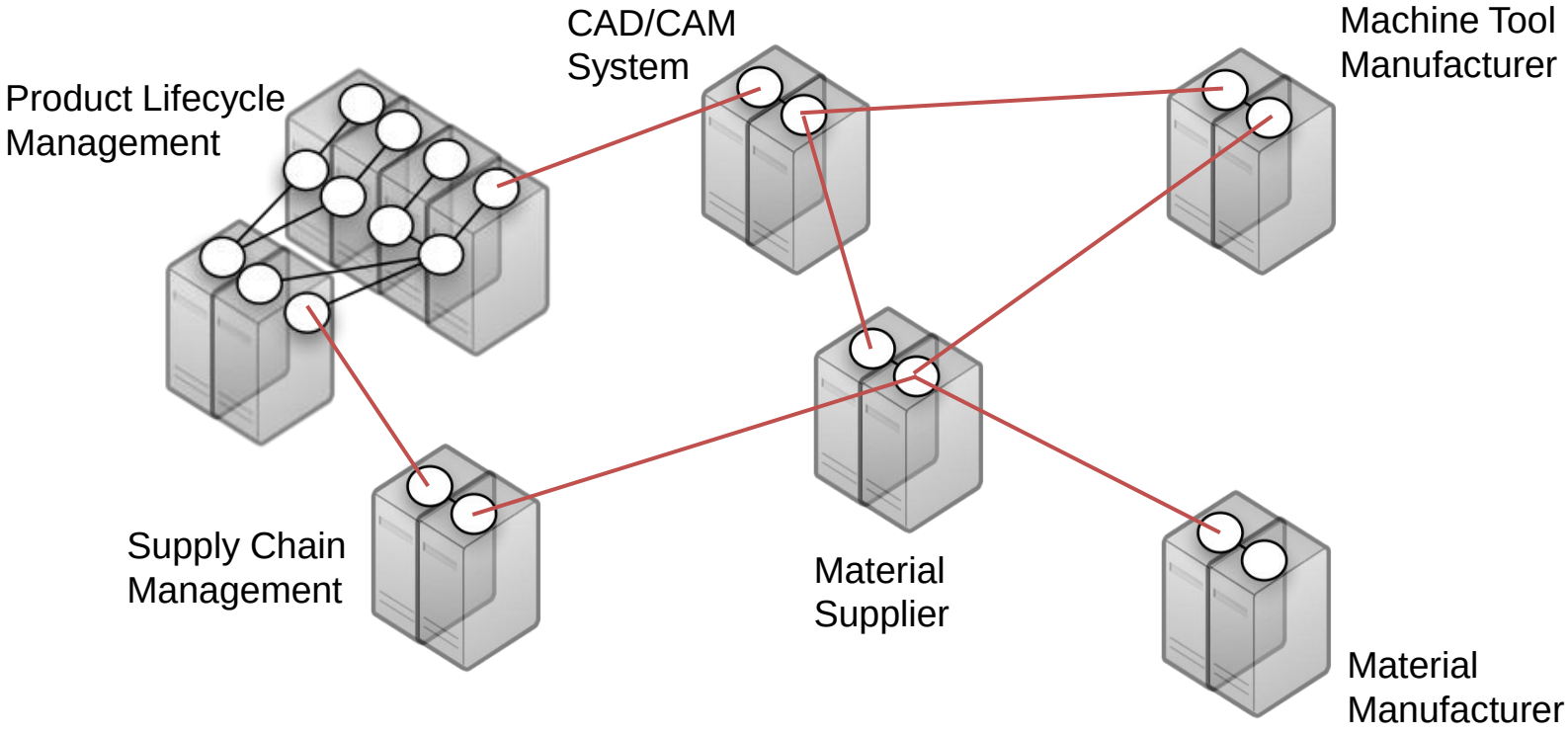
TIMELESS BUSINESS

1. Aerospace company BigAirTransportation (BAT) is involved in a project with an estimated lifecycle of 50 years.
2. This spans the initial vision through to product decommissioning and recycling.
3. The vision is to produce a lighter passenger jet that will use approximately 20% less fuel than their current product and use key components (e.g. wings, fuselage, and empennage) that require 10% less maintenance inspection.
4. After initial design and cost engineering the passenger jet will be constructed from 52% reinforced plastic composite.

SEVENTH FRAMEWORK PROGRAMME

Concrete TIMBUS Example

DigitalPreservationCoalition



DigitalPreservationCoalition

TIMBUS

TIMELESS BUSINESS

Planning

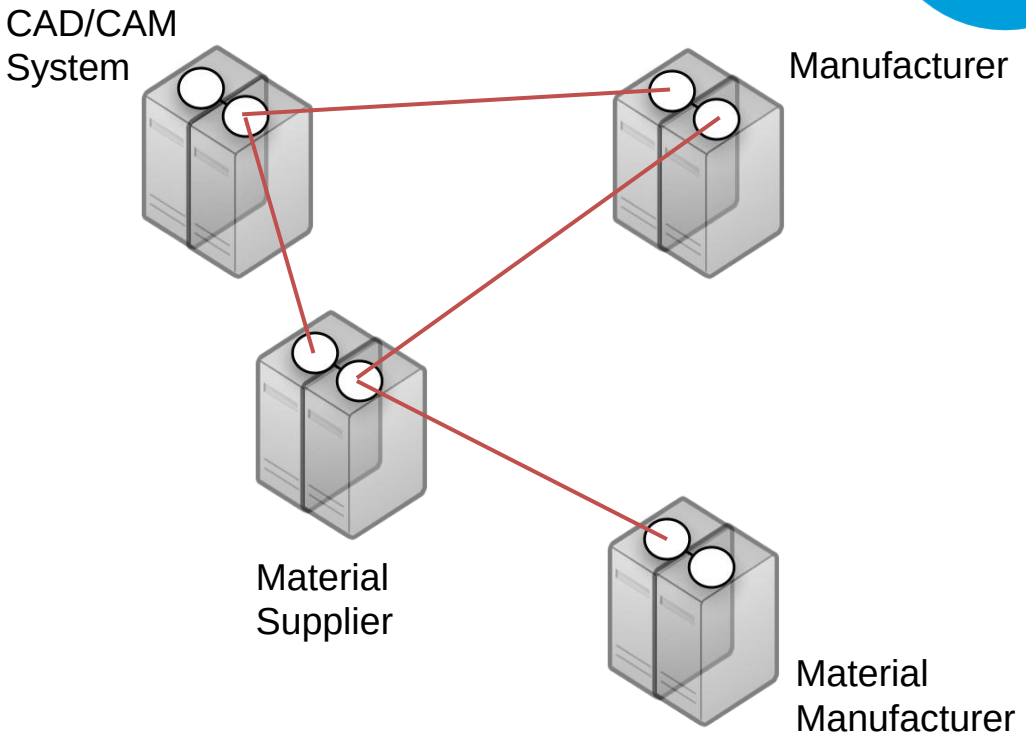
- Risk analysis performed
- PLM -> SCM *not* deemed expedient for preservation
- CAD/CAM -> Material Supplier -> Material Manufacturer -> Machine Tool Manufacturer *critically* expedient for preservation

Preserve – *performing the preservation of business processes*

- Legalities Lifecycle Management
- Software Service Engineering for Preservation
- Business Process Virtualization and Storage
- Processes and Standards for Digital Preservation of Business Processes

Re-deploy– *reruning/extending a business process at a future date*

- Business Process Exhumation and Integration Support (with Future Simulated Test Bed)



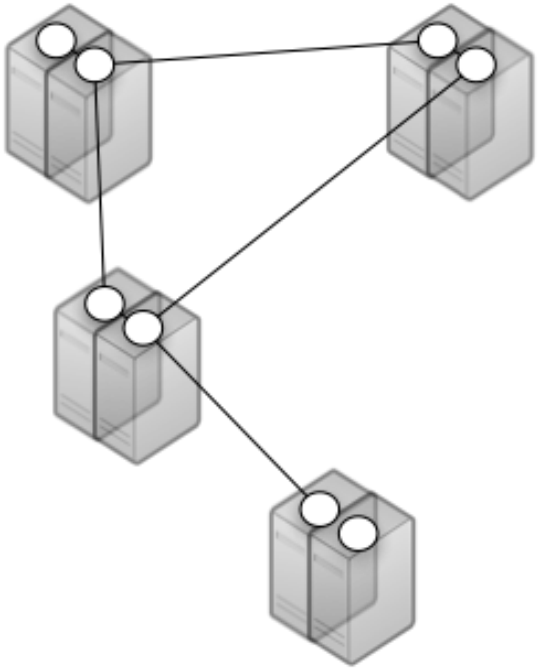
SEVENTH FRAMEWORK PROGRAMME

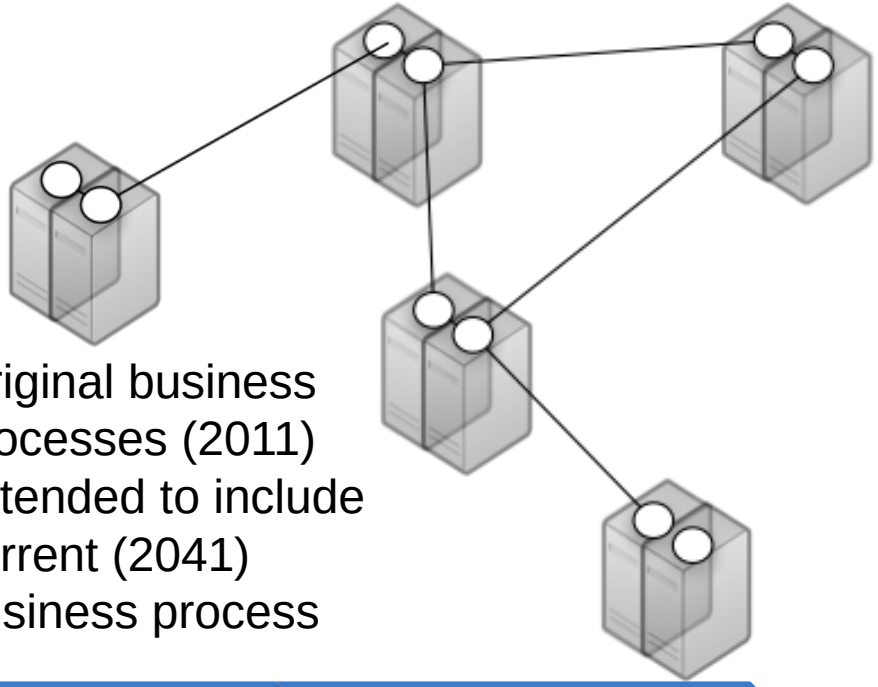
www.dpconline.org

DigitalPreservationCoalition

TIMBUS

TIMELESS BUSINESS





Original business processes (2011) extended to include current (2041) business process

2011 ... 2041

SEVENTH FRAMEWORK PROGRAMME

www.dpconline.org

TIMBUS

TIMELESS BUSINESS

- Planning Innovations
  - Service Dependency Analysis
  - Business Process Context Capture
- Preservation Innovations
  - Legalities Lifecycle Management
  - Software Service Engineering for Preservation
  - Business Process Virtualization and Storage
  - Processes and Standards for Digital Preservation of Business Processes
- Redeploy Innovations
  - Business Process Exhumation and Integration Support (with Future Simulated Test Bed)

SEVENTH FRAMEWORK PROGRAMME

DigitalPreservationCoalition

Oh and ... the Digital Preservation Coalition

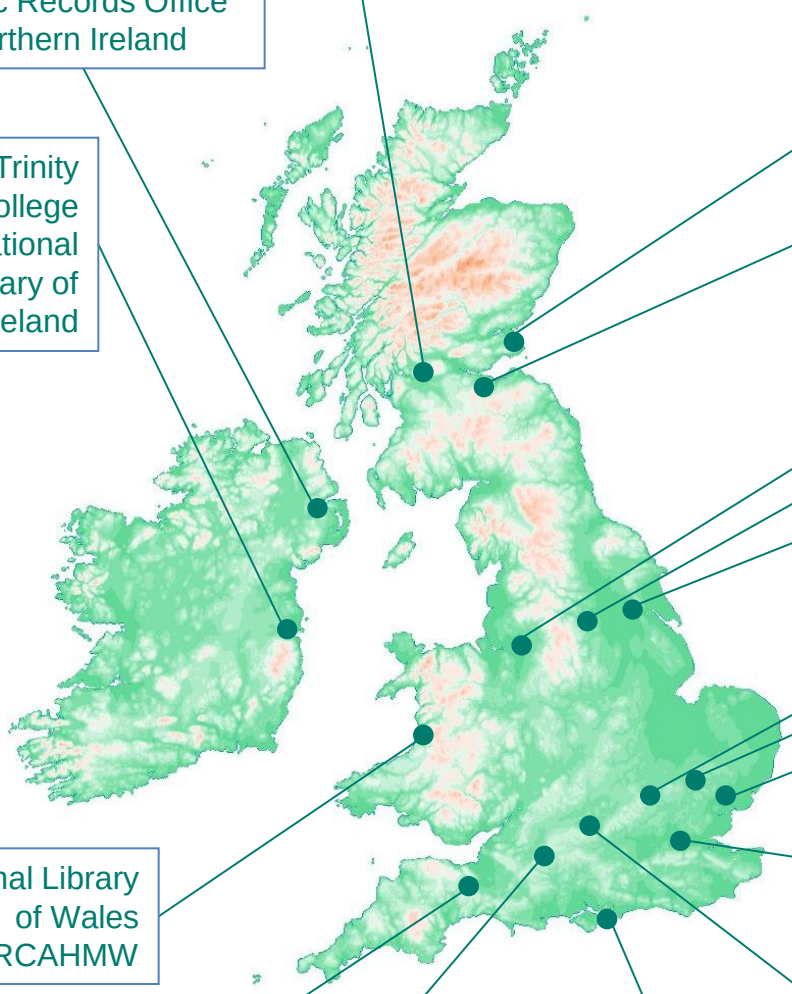
...to make our digital memory accessible tomorrow ...

- Workforce development
- Advocacy
- Knowledge Exchange
- Assurance and Practice
- Partnership

UKLA

Trinity College Library Dublin
University of Dublin

w.dpconline.org



PORTICO

Glasgow University

Public Records Office of Northern Ireland

Trinity College 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

Manchester University Library

Leeds University Library

British Library
Archaeology Data Service
York University

Open University Library

Cambridge University Library

UK Data Archive

Wellcome Library
King's College London
Publishers Licensing Society
The National Archives
Tate
London School of Economics
ULCC
BBC
Parliamentary Archives
RIN
JISC
LSE
Royal Museums Greenwich

www.dpconline.org

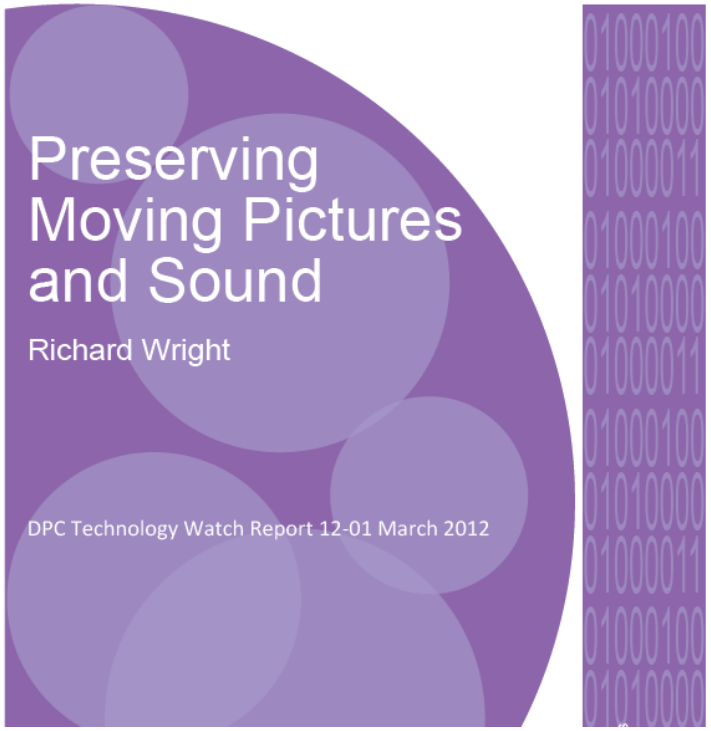
DigitalPreservationCoalition

Workforce Development

- Leadership Programme
- Expert Briefing
- ‘Roadshows’
- Peer networking

DigitalPreservationCoalition

Knowledge Exchange



- Technology Watch
- What's New
- Briefing Days
- Email list
- Case studies
- Conference reports

www.dpconline.org

DigitalPreservationCoalition

Assurance and Practice

- Standards Watch
- Working parties:
  - CERT
  - WAPTF
  - ‘Bedern Group’
  - CDT
- Internal consultancy

DigitalPreservationCoalition

Advocacy

- Direct Advocacy
- Knowledge Base
- Executive Briefings
- Awareness Raising
- Preservation Award

DigitalPreservationCoalition

Partnership

- Planning Day
- Directors' Group
- Co-funded projects
- Awareness Raising
- Preservation Award

DigitalPreservationCoalition

DPC as a partner and friend!

(join us)

www.dpconline.org

DigitalPreservationCoalition

Our digital memory tomorrow

what we've learned about digital access and preservation and how it's useful now and in the future

Access and preservation are almost the same thing:

No long term access without preservation

No use to preservation without some access

What is the question?

6 basic challenges and some solutions

Three lessons from experience

Some emerging trends …

www.dpconline.org