



Digital**Preservation**Coalition

Digital Cultural Heritage Roadmap Project

# Preservation strategies for digital cultural heritage

Part One: Cultural Heritage and Digital Preservation

Part Two: Challenges for the future

Part Three: The DCHRP Roadmap

## Part One: Cultural Heritage and Digital Preservation

*Digital resources are extraordinarily flexible and surprisingly fragile...*

*... what need we do to ensure that our generation's digital creativity becomes a meaningful digital legacy for the next?*



Digital **Preservation** Coalition

# Digital Preservation make bleak reading: **a game we can all play**

*<Enter your details here>*

.....

.....

.....



## **What's the problem?**

Obsolete media

Obsolete software

Obsolete wrappers (file formats)

Bit rot

De-synchronisation

Poor representation information

Lack of funding

Loss of rights

Encryption and security

Physical loss (fire, theft, flood ...)

Virtual damage (malware, ...)

etc ...



Digital**Preservation**Coalition

## Digitization and Digital Preservation

# Digital Cultural Heritage

Access and engagement

Surrogacy

Crowdsourcing

Comparison

Etc ...

*But ...*

Sustainability

Synchronization

Technical obsolescence

Maintenance

Etc ...

## Preservation policies in digitisation

# 18 questions...

**Who** → are your users long and short term?  
maintains the intellectual integrity of the content?  
maintains the technical integrity of the content?

**What** → is the content and where did it come from?  
formats and metadata do you use?  
is the size of the collection?

**Where** → is the master copy of the metadata?  
is the master copy of the principle content?  
are other copies held?

**When** → How long do you expect content to be available for?  
(if) things go wrong what are the consequences?

**How** → will the collection be created?  
maintained?  
migrated or emulated?  
will the collection be updated ?  
Will the metadata be updated?  
do you track who did what?  
do you keep data synchronised?

Policy needs to be funded

# How much will this cost?

## 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



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

Archaeology Data Service

Glasgow Museums Resource Centre

**Not a direct  
comparison**

Setup:  
Tens of thousands

Setup:  
Tens of millions?



Assertion based on my experience:

**We are able to understand and assert the value of museums, libraries, archives, heritage ...**

But

**We are poor at understanding and asserting the value of digital.**

**(Is data the new oil?)**



Digital**Preservation**Coalition

**Digital preservation expensive ..?**

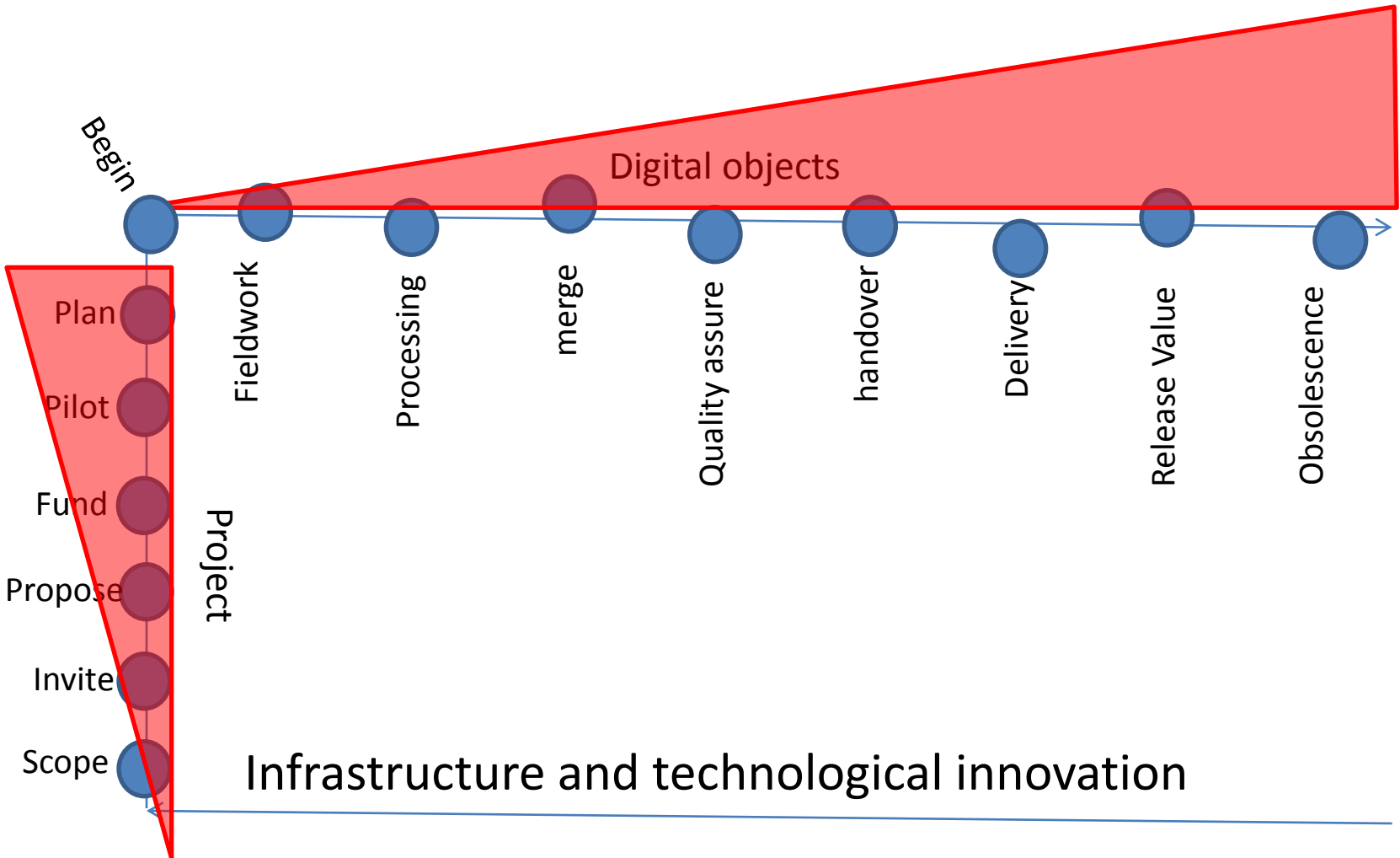
**Not necessarily: it's an  
unfunded mandate**

**Don't throw money at it:**

**Get the mandate properly incorporated**



# When should preservation start?



# Digitization and Digital Preservation

## Digital Cultural Heritage



### Preservation policies in digitisation

#### 18 questions...

- Who** → are your users long and short term?  
maintains the intellectual integrity of the content?  
maintains the technical integrity of the content?
- What** → is the content and where did it come from?  
formats and metadata do you use?  
is the size of the collection?
- Where** → is the master copy of the metadata?  
is the master copy of the principle content?  
are other copies held?
- When** → How long do you expect content to be available for?  
(if) things go wrong what are the consequences?  
will the collection be created?  
maintained?  
migrated or emulated?
- How** → will the collection be updated ?  
Will the metadata be updated?  
do you track who did what?  
do you keep data synchronised?

@williamkilbride

www.dpconline.org

Sustainability not just technical  
Sustainability not just money  
Sustainability not just planning

# Sustainability (also) about community



## *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, opportunities are lost



Digital **Preservation** Coalition

**asking the right questions**

Digital preservation is not just about 'data':

Digital preservation is not just about 'technology':

**it's about**

**people and**

**opportunity**

A bright yellow, multi-pointed starburst graphic with a black outline, containing the text 'Start here!' in white.

**Start  
here!**

## Part Two: Some thoughts in the future

The file is not necessarily the atomic unit of data. Preservation or records management approaches which fetishize files are never likely to be sufficient.



## Six (or Seven) Observations

1. **Data is growing**, budgets are not
2. **Big data complex data** as a metaphor for our future problems: does the cloud / GRID help?
3. Does the cloud / GRID make it **easier to engage** in digital preservation
4. Why is digital preservation **so hard**?
5. What is **data** anyway?
6. What is **trust** going to be like in the distributed world?

Does '**preservation as a service**' help?





## **Data is growing on three axes**

Scale

Complexity

Expectation

## **Capacities are not**

Data volumes: 60% pa

Storage capacities: 25% pa

Data centre budgets: 2% pa



Digital **Preservation** Coalition

... successful, practical digital preservation is going to be about **workflow and capacity** as much as about **obsolescence or authenticity**.

### Migration

Intervention at data layer to ensure information objects  
Based on significant properties of content and performance  
Quick start, low cost, ready quality assurance, focus on data/access  
loss of authenticity, poor with complex objects

### Emulation

Intervention at software / OS layer to ensure operation of software  
Based on significant properties of the environment and its behaviours  
Slow start, high technical threshold, access less transparent  
retains authenticity, geared towards complex objects

Migration has done all the running in the last 10 years (20 years)



Digital**Preservation**Coalition

## **Big data / complex data**

Web archives

Sound and vision

Digitised content

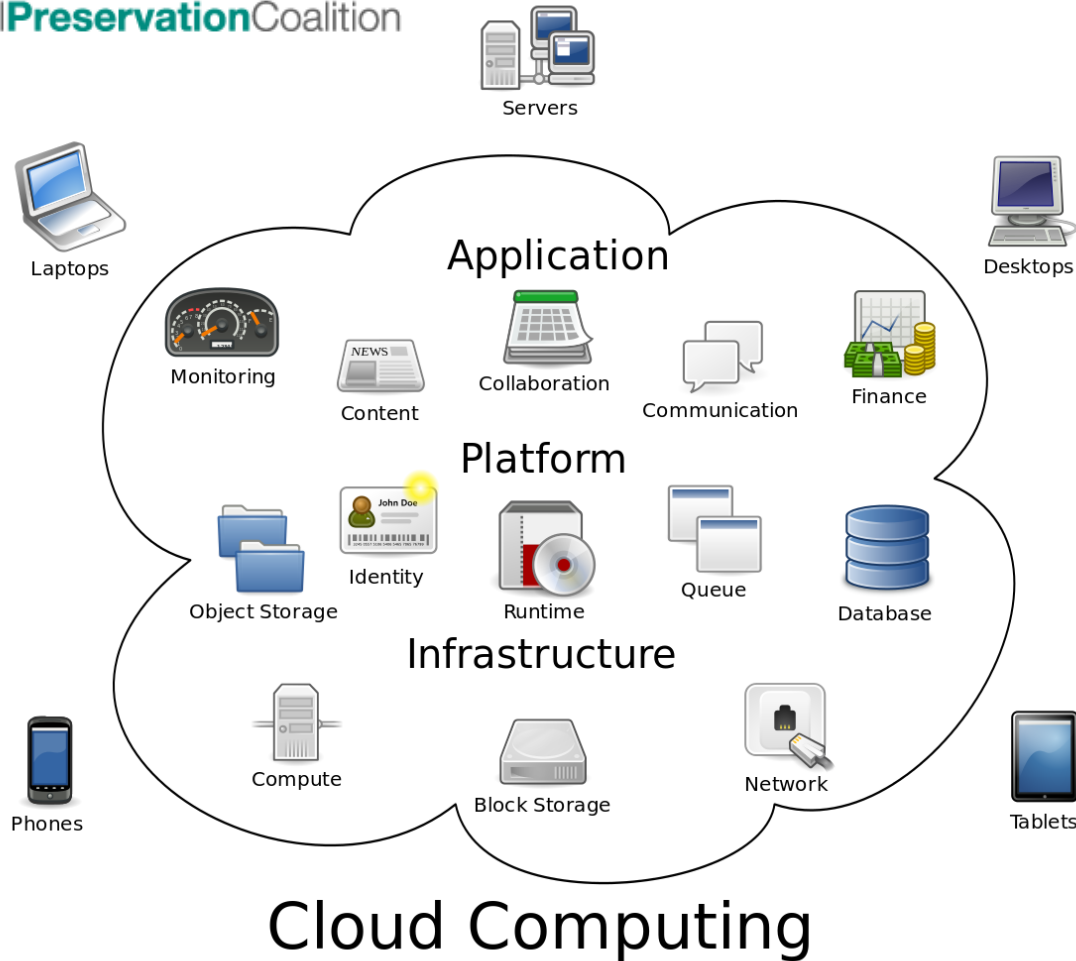
Email

Complex, vast, valuable,  
heterogeneous

Difficult to move

Difficult to access

Greater than the sum of its  
parts



.... the delivery of computing as a service rather than a product

## Scalable and Elastic

Services scale on demand to add or remove resources as needed

## Service based

The service could be considered "ready to use" or "off the shelf"

Offers IaaS, PaaS, and SaaS

## Economical

Services share a pool of resources to build economies of scale

Metered by Use : Pay-as-you-go

## Evolvability

Supports for migration and upgrades.

Services are configurable

### Virtualization

- Virtualisation puts a layer between physical hardware and controls access to that machine.
- Each guest machine (VM) that is built on top of the abstraction layer (hypervisor) is then provided access to the physical host's resources without modification.
- The hypervisor act as a traffic cop by allowing certain amount of the physical resources to be used by the guests, as well as manages resource sharing when more than on guest system try to access the resources.

- Emulation

- Duplication of functionality of systems, be it software, hardware parts, or legacy computer system as a whole, needed to display, access, or modify a certain contents.



## Does virtualisation make DP easier?

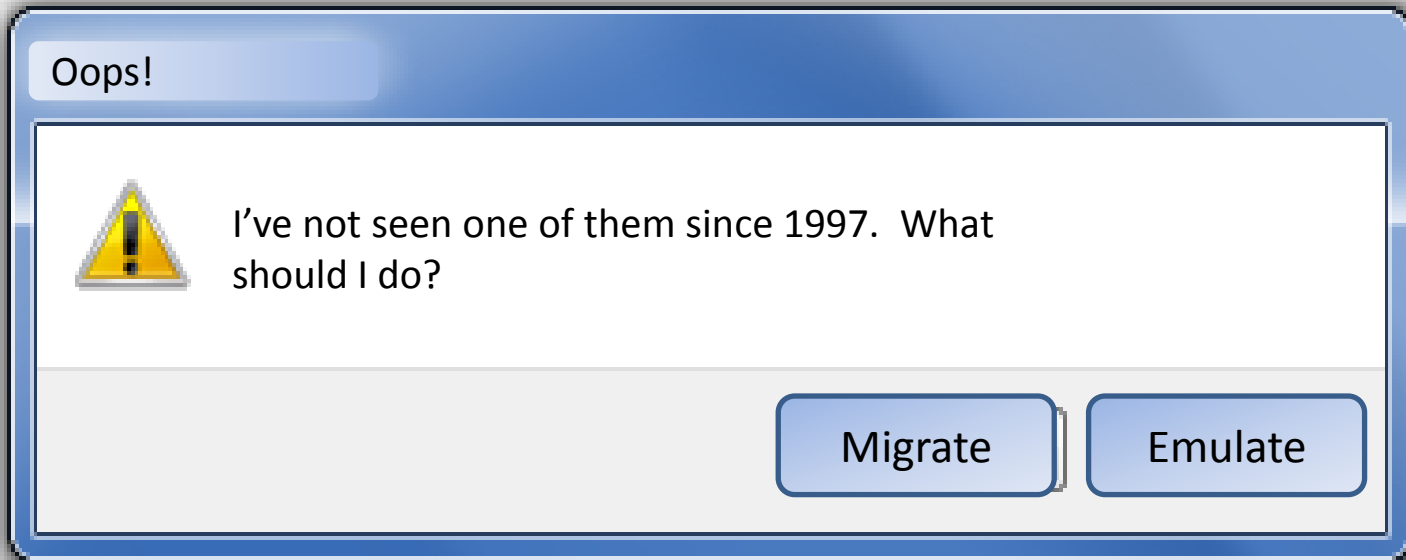
- Elastic
- Bypass (generic) corporate IT
- Industrial scale preservation?
- Highly dependent on services
- Pret a porter?
- Highly dependent on location and configuration of services
- Easier deployment of complex solutions – virtualisation





Digital **Preservation** Coalition

**We've made this all very hard.**





# What is data any way??

❄️ ◻ ♦️ 🌀 ◻ ⚡️ ♦️ 🌀  
 † ◼ ✕ ✦ ♀ ◻ ♦️ 🌀 ●  
 ⚡️ ♀ ♀ ● 🌀 ◻ 🌀 ♦️ ✕ ◻ ◼ ◻ ♂  
 ✕ ◼ ♦️ ♀ ◻ ⚡️ ♀ ◻ ♀ ◼ ⚡️ ♀ ◼ ♀ ♀ 🖱️

or

## Towards a Universal declaration of interdependence.



Digital **Preservation** Coalition

## From Trusted Digital Repository to untrusted digital commons

It's not the repository we trust – it never was.

the **people**, the **tools**, the **services**, the **policies**, the **business plan**, the **mission**, the **organisation**, the **context**, the **user-focus**.

How is this different?



Can infrastructural approaches help core DP issues?

- Storage – yes!
- Compute – yes!
- Costs – maybe (maybe not)
- Skills – to some extent
- Making emulation (virtualisation) realistic?

### Part three: The DCH Roadmap

*Common procedures and workflows would reduce the cost in terms of time and money and would contribute to the interoperability and open-ness of data.*

*Existing e-infrastructures could be efficient channels for the delivery of advanced services that can be used by the digital cultural heritage sector for digital preservation.*



## **Core assumptions**

- Digital cultural heritage collections need preservation
- Expensive and tricky to accomplish individually
- Significant economies of scale and scope are possible
- E- infrastructures (esp Grid) have proven their value in the hard sciences
- E-infrastructures are flexible and have capacity to assist



## Core challenges

- Considerable diversity in workflows in cultural heritage sector
- Skills gap in cultural heritage agencies
- E- infrastructures (esp Grid) have limited preservation experience
- How to develop trust in distributed preservation infrastructure
- How to ensure sustainability
- How to respond to other emerging technologies





Digital**Preservation**Coalition

## Practical actions

- Harmonizing storage and preservation practices in DCH institutions
- Facilitate dialogue between DCH institutions and E-infrastructure providers
- Understand and create the conditions in which these two sectors can integrate their efforts
- Examine and agree models for the governance, maintenance and sustainability of infrastructure



## Timescales

- Short Term (2014-5) Define topic
  - Improve interoperability
  - Harmonise storage and preservation
  - Establish cross-sector integration
  - Establish governance model
- Mid Term (2016-7) next steps
  - Enhanced data storage and preservation
  - Enhanced interoperability
  - Programme of integration
  - Redesign and implement governance
- Long term (2018 onwards)
  - Review, update, consolidate



## Audiences

# Who is paying attention?

## Policy makers

- Create a legal and institutional setting which makes progress possible

## Digital Cultural Heritage

- Harmonise (simplify?) practice and identify priorities (eg training)

## E-Infrastructure providers

- Provide access to infrastructure in trusted but simple ways

In conclusion: Some Questions



What this means ....

## Some questions?

What will success look like?

Why chose E-infrastructures over Cloud?

How will we train our staff?

How realistic is it to harmonise workflows?

How will we track and respond to user needs?

What other sorts of partnership are needed?

Where will the money come from?



Digital**Preservation**Coalition

Digital Cultural Heritage Roadmap Project

# Preservation strategies for digital cultural heritage

Part One: Cultural Heritage and Digital Preservation

Part Two: Challenges for the future

Part Three: The DCHRP Roadmap