Two minute exercise before we get started:

What do you want to know?
What are your most important digital collections?
What data do you want to have available in 10 years time?
What has brought you to the digital preservation challenge?

@williamkilbride

www.dpconline.org

Digital Preservation: technology, information and content

1. Digital preservation: what's the problem?
2. Digital preservation and digitization: planning preservation from the start
3. Practical wisdom: what I wish someone had told me before I started
4. Emerging challenges in digital preservation: planning for an uncertain future
5. The digital preservation community: a welcome and encouragement

@williamkilbride

www.dpconline.org

1. Digital Preservation: What's the Problem?

Traditional, 'analogue'
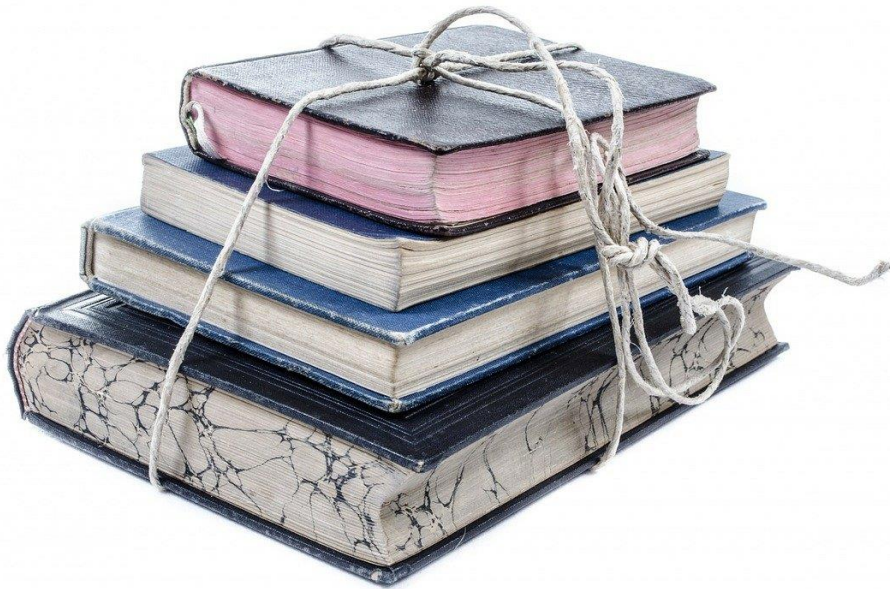
- Robust
- Tangible
- Independently understandable
- Well-developed approaches to preservation
- Experienced in assigning value

Digital

- Ephemeral
- Need tech to interpret
- Obsolescence
- How to estimate value?
- New skills and solutions required
- But also new opportunities!

What's at stake?

By the way that means …

Structured and unstructured data

Born digital and digitised content

From Selfie stick to space probe

Office Documents	Email
Digital Photographs	Web pages
Digital Video and Audio	Twitter
Grey literature	Laser Scans
CAD Plans	MRI and CT scans
Geo Mapping and GIS	Databases
Transactional Records	Financial data
Spreadsheets	Scientific outputs
eBooks and eJournals	Text and WhatsApp

It includes…

What's your favourite digital object?

What's the most important?

What digital output can you not live without?

What would you like to have available in 10 years time?

…

@williamkilbride

www.dpconline.org

What is Digital Preservation?

"the **series** of **managed activities** necessary to ensure **continued access** to digital materials for **as long as necessary**"

(Digital Preservation Handbook)

@williamkilbride

www.dpconline.org

What's the problem?

Media failure

Media obsolescence

Software failure

Software obsolescence

Inaccessible file format

Format migration errors

Loss of media

Disaster

Broken dependency

Failure of systems integration

Data Theft

Virus or malware

Poor documentation

Poor rights management

Lost encryption key

Poor version control

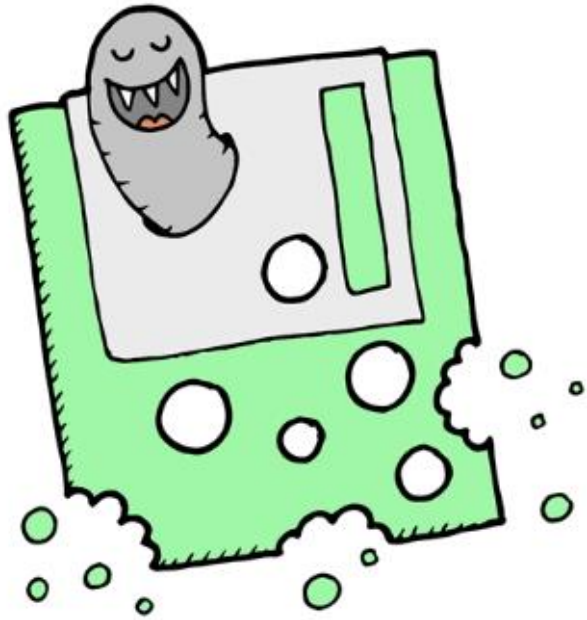Inadvertent deletion

Malicious deletion

Poor file-naming

Corruption

Link rot

…

@williamkilbride

www.dpconline.org

So digital preservation makes bleak reading

<Enter your details here>

....

....

....

In Summary

Digital materials (images, documents etc.) have value

They create opportunities

...but...

Access depends on software, hardware, and people

...and...

Technology and people change, creating barriers to reuse

...therefore...

We want to protect access to the data

So we can protects and maintain the opportunities for reuse and impact in the real world

Where to start?

Digital preservation not just about technology:
Digital preservation not just about data
Digital preservation not just about

it's about

people and opportunity

Don't start with the technology: start with the people

Pause for response, questions and answers.
Also sharing experience from the audience…

2. Digitization and Digital Preservation: Planning for a preservable future

Digitization: The process of creating digital files by scanning or otherwise converting analogue materials.

For the avoidance of doubt: Digitization _is not_ Digital Preservation
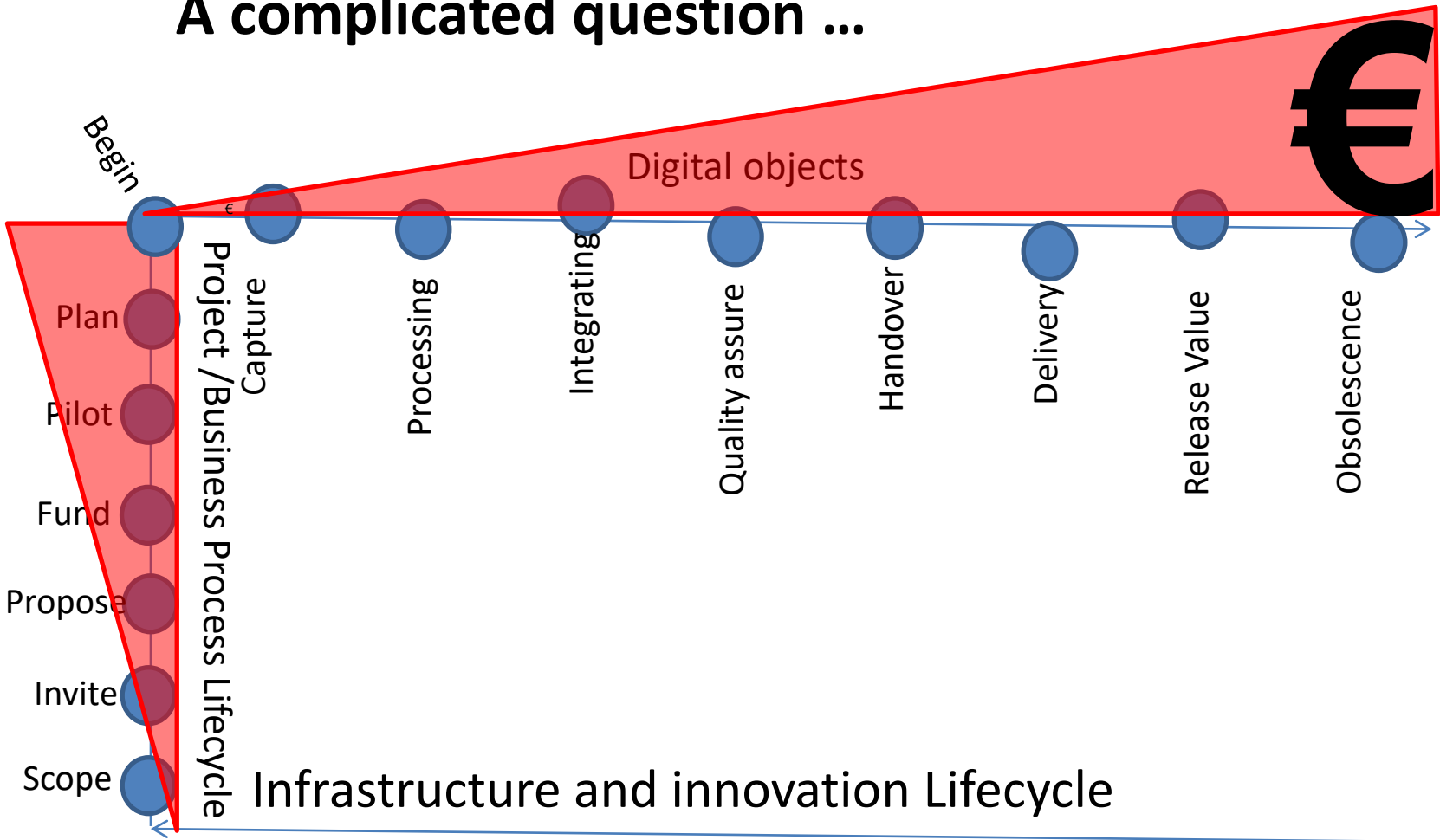(You may need to repeat this frequently)

Digital Preservation: the series of managed activities necessary to ensure continued access to digital materials for as long as necessary …beyond the limits of media failure or technological and organisational change.
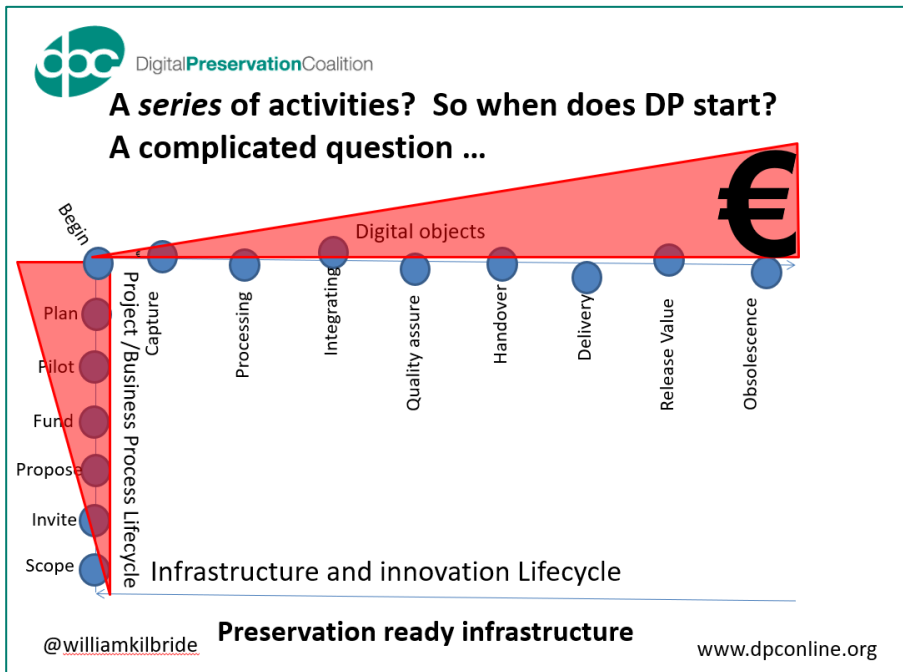
@williamkilbride

www.dpconline.org

A *series* of activities? So when does DP start?
A complicated question …

Digital objects

Begin

€

Plan

Pilot

Fund

Propose

Invite

Scope

Project /Business Process Lifecycle

Capture

Processing

Integrating

Quality assure

Handover

Delivery

Release Value

Obsolescence

Infrastructure and innovation Lifecycle

Preservation ready infrastructure

@williamkilbride

www.dpconline.org

Look more closely:
Who is involved?

dpc Digital**Preservation**Coalition

A *series* of activities? So when does DP start?
A complicated question …

€

Begin

Digital objects

Plan

Pilot

Fund

Propose

Invite

Scope

Project /Business Process Lifecycle

Capture

Processing

Integrating

Quality assure

Handover

Delivery

Release Value

Obsolescence

Infrastructure and innovation Lifecycle

Preservation ready infrastructure

@williamkilbride

www.dpconline.org

It starts early: so it involves record creators and managers as well as archivists.

In fact: if we leave all the work to archivists then it's probably already too late

Hard message to hear: it's likely to be a whole organization change

@williamkilbride

www.dpconline.org

...plan digital preservation into your digitization workflows

What can you do in the context of a digitization workflow to make digital preservation possible?

Documentation: capture metadata about the workflow and packaging it with the digital objects.

Format: pay attention to the subtle challenges that arise from different types of file format.

Succession: plan from the outset who will look after the data once the active project phase is over

Use case: talk to users and track their changing needs and expectations

Interoperate: the more open you can be with the project the greater the long term prospect

Deeper dive: the file formats issue…

Changing file formats

vs

'Robust' formats

vs

Proliferating formats

vs

Conformant data containers

How to pick a winner ...
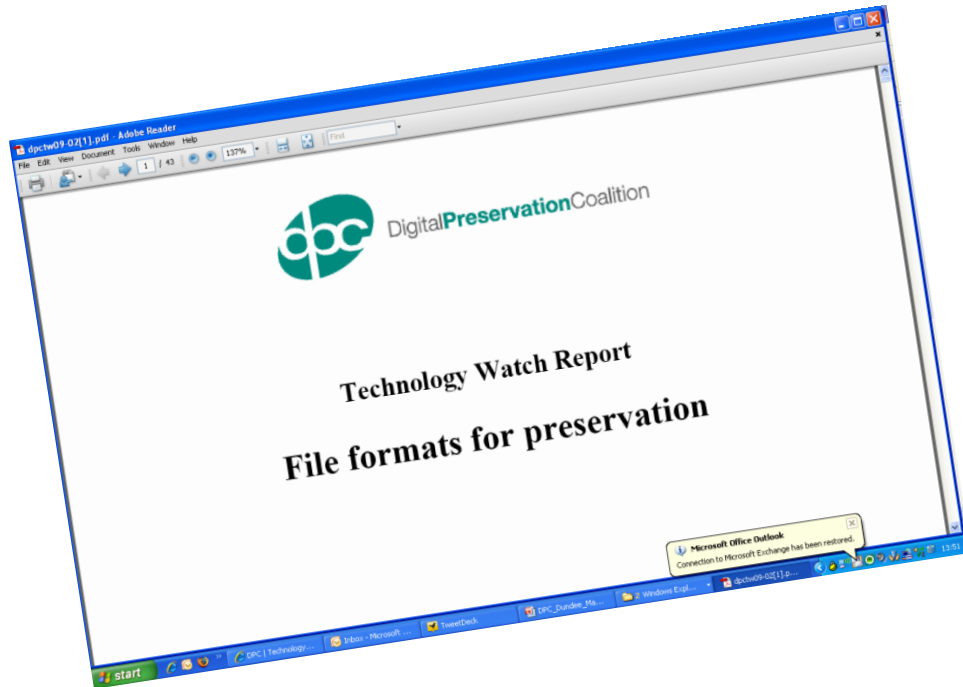
Adoption
Dependency
Disclosure
Transparency
Metadata support
Interoperability
Complexity
Stability
Rights management

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

@williamkilbride   www.dpconline.org

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

@williamkilbride

www.dpconline.org

How to pick a winner …

'Digital Universe' Nears A Zettabyte

May 4th, 2010 : Rich Miller

Figure 1: The Digital Universe 2009 – 2020
Growing by a Factor of 44

2020
35 ZB*

2009
0.8 ZB*

*Zettabyte = 1 trillion gigabytes

Source: IDC Digital Universe Study, sponsored by EMC, May 2010
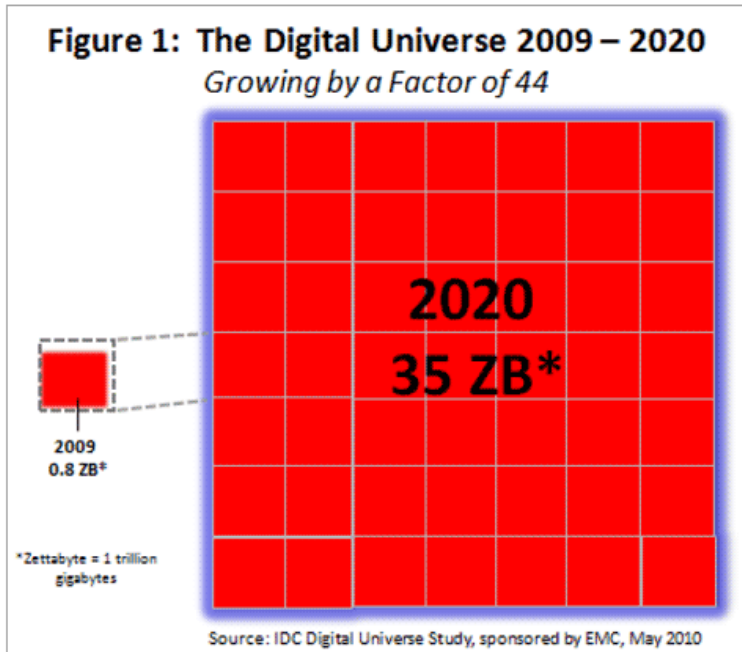
The Great Recession hasn't slowed the breakneck growth of the Digital Universe. In 2010 the volume of digital information created and duplicated in a year will reach 1.2 zettabytes, according to new data from IDC.
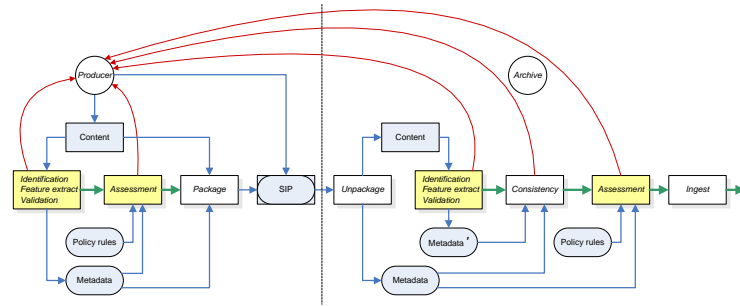
Format obsolescence is not the exact problem, though it's connected

Chaotic workflows and lack of capacity

The file is not the atomic unit of information

Pause for response, questions and answers.
Also sharing experience from the audience…

3. What I wish someone had to me:
Digital preservation in a nutshell?

Do you want to preserve the content yourself?

Do you want someone else to preserve the content?

Approaches:

migration, emulation, virtualisation …

Standards:

METS, OAIS, PREMIS, Trusted Repository…

Technologies:

Microservices, Repositories, End to End solutions

Approaches

Migration:

Intervention at the data layer to ensure the packaging remains current

Emulation:

Intervention at the software or operating system layer to maintain access to applications

Virtualization:

Assembling tools on demand in a cloud environment to access applications and data

Hardware Museum:

Retaining old equipment to access data and applications

(etc!)

It's a research topic and there may yet be new things

@williamkilbride
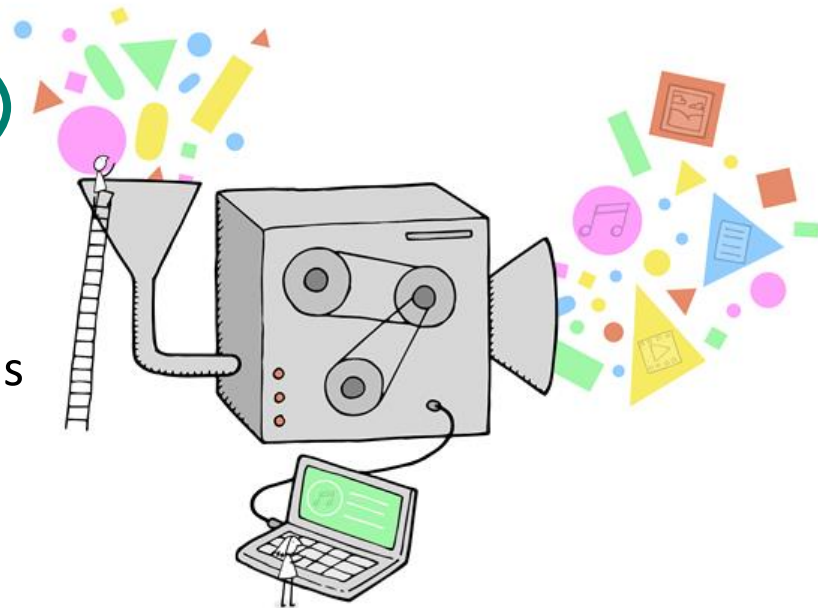
Standards

Open Archival Information System:
Information model, functional model, environment model
PREMIS and METS:
Preservation Metadata and Packaging Standard
Trusted repository:
Validation standard

Three parts to any solution:

digital preservation

technology

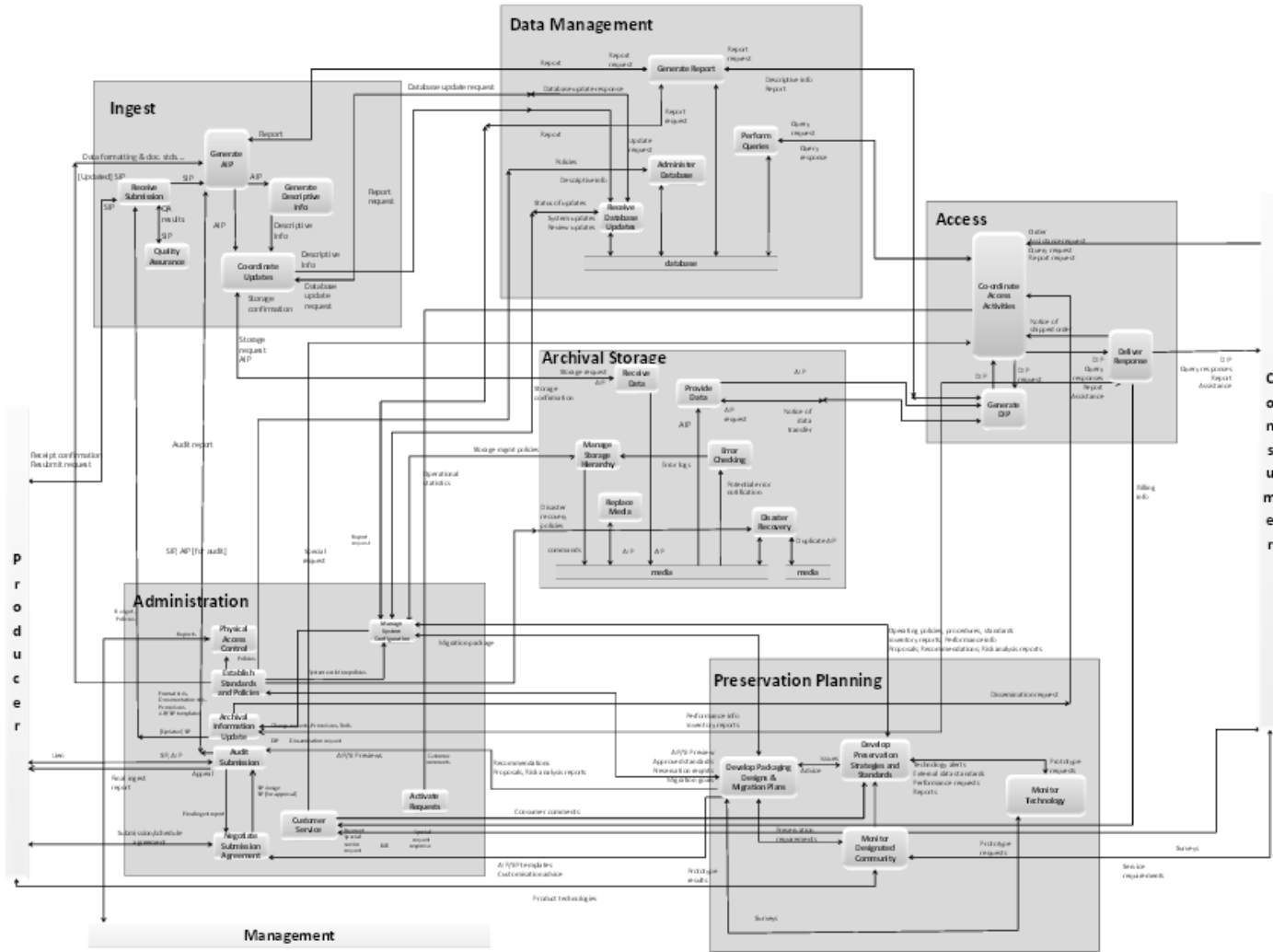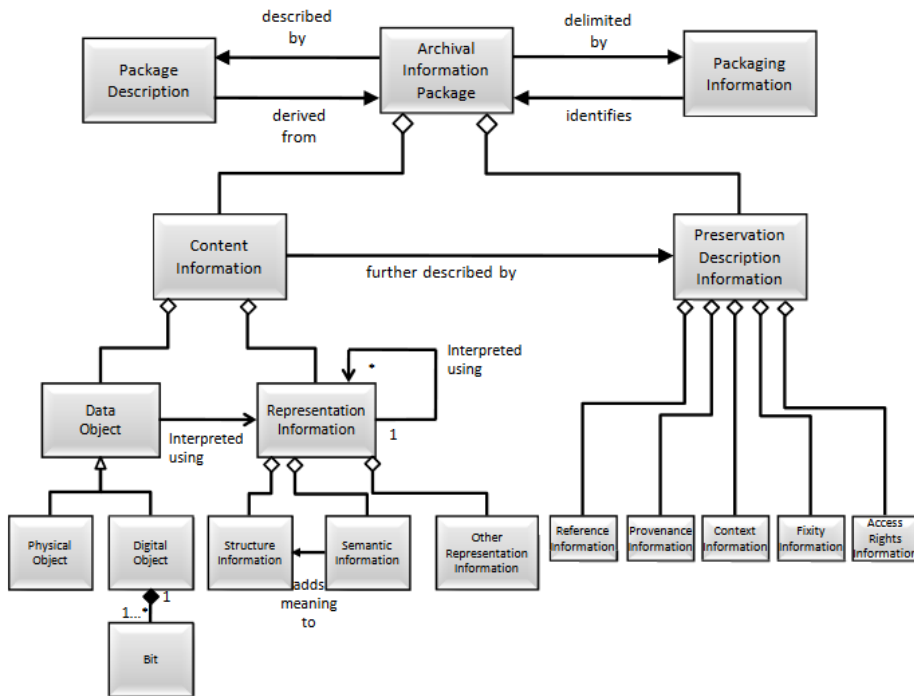organization

resources

How hard can it be?
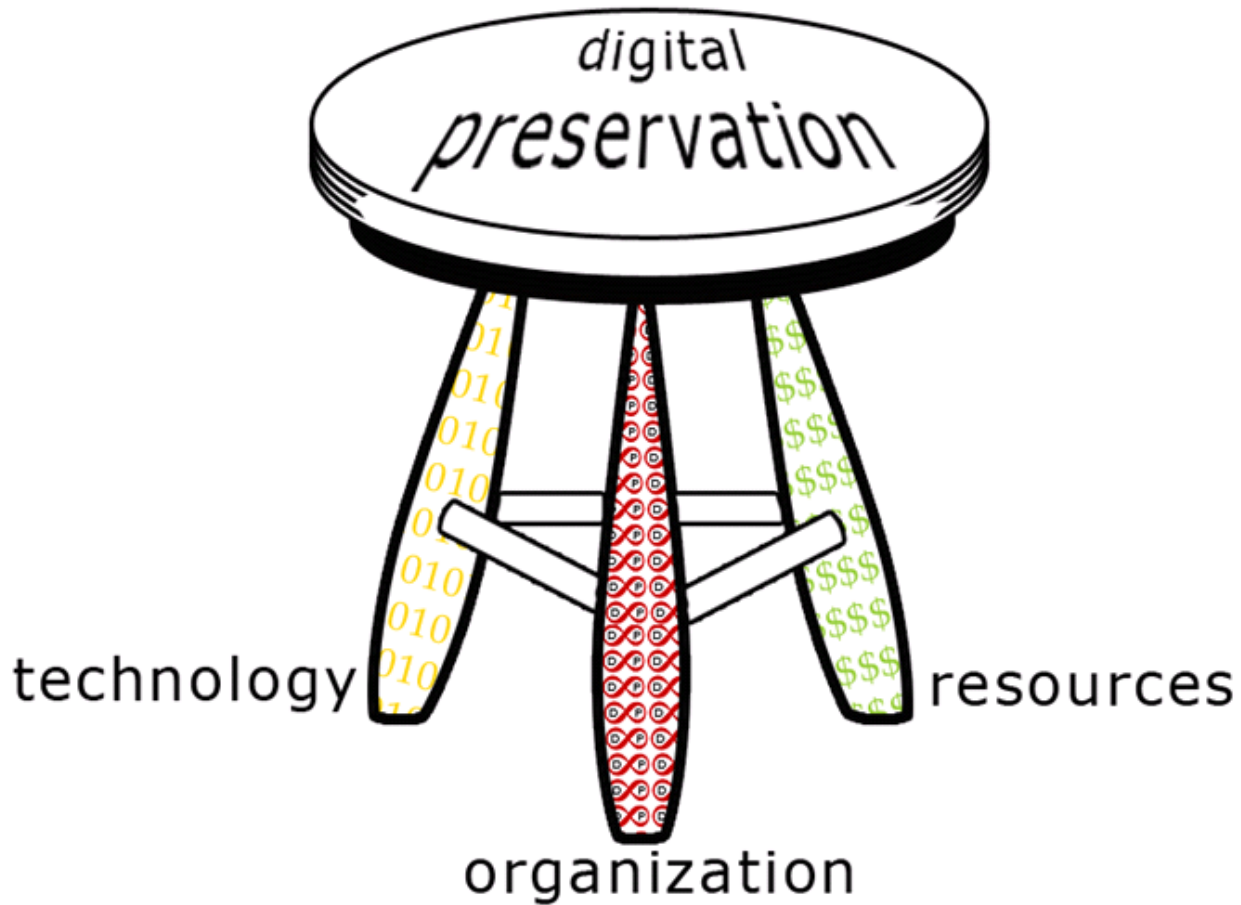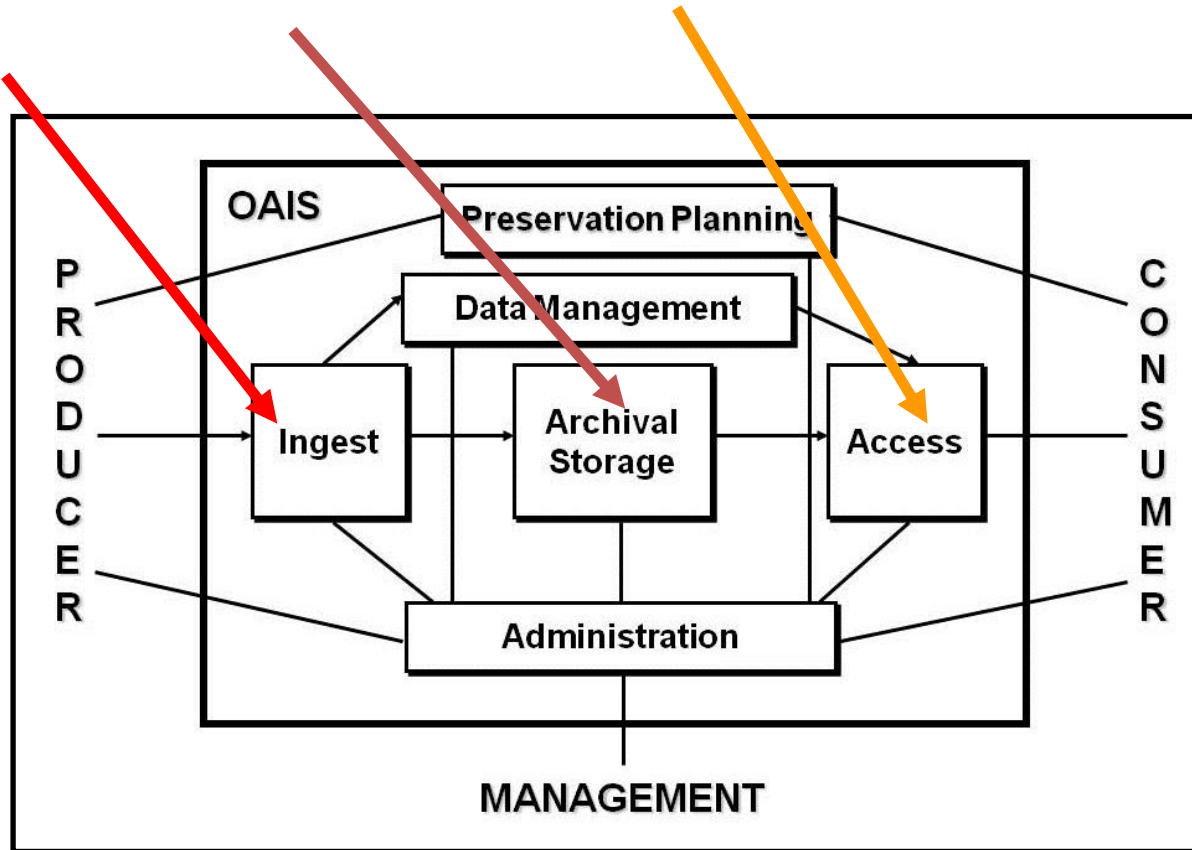

Lavoie 2015

Figure A-1: Composite of Functional Entities

OAIS:

- The Information Model is the more important of the three
- Designated community and representation information
- Inadvertently comparing yourself to NASA …

Lavoie, B 2014 *The Open Archival Information System Reference Model Introductory Guide* DPC Technology Watch Report http://dx.doi.org/10.7207/twr14-02
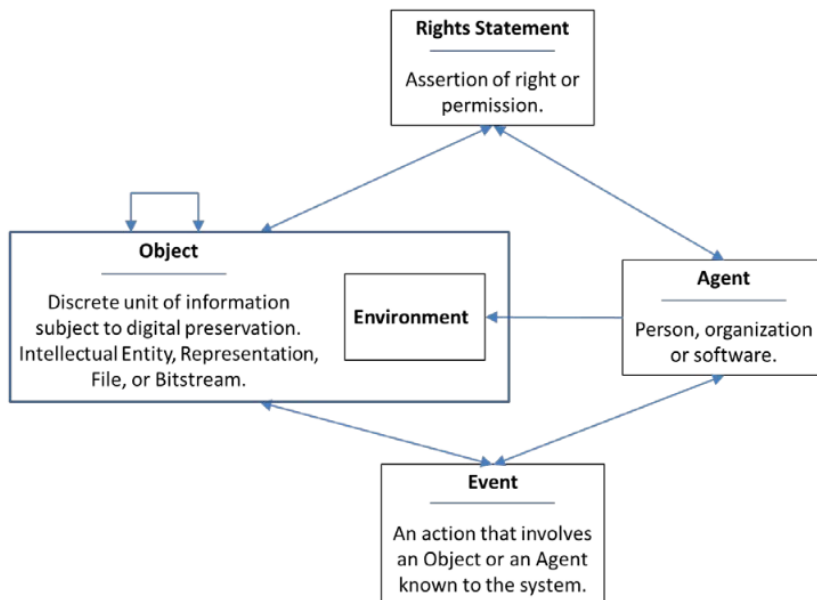
Documentation

PREMIS:
Preservation Metadata Implementation Strategy
METS:
Metadata Encoding and Transmission Standard

Lavoie, B and Gartner R 2013 *Preservation Metadata (2nd Edition)*
DPC Technology Watch Report http://dx.doi.org/10.7207/twr13-03

Tools

Standalone tools:
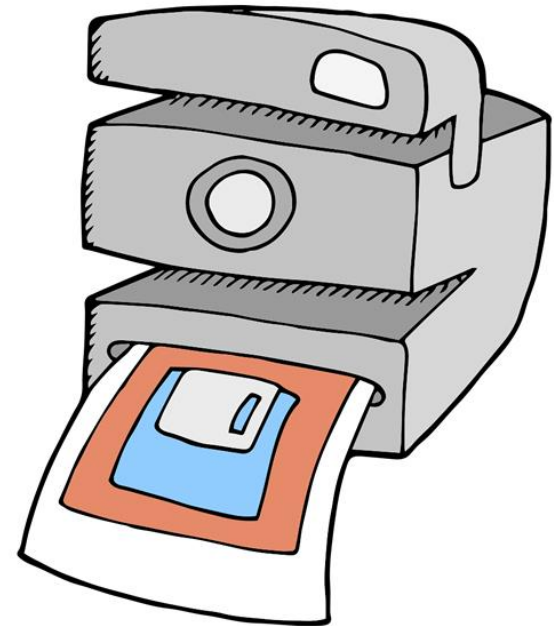
DROID, FITS, C3PO, BagIt, VeraPDF …

Repositories:

Fedora, Duraspace, GitHub …

Registries:

ProNOM, KEEP …

End to end:

Preservica, Rosetta, RODA, Archivematica, LibNova, Arkivum…

COPTR: The Community Owned Preservation Tools Repository
http://coptr.digipres.org/

DigitalPreservation Coalition

PRONOM and DROID on Characterization

… a short video

Checksums, fixity and replicated storage

… a short video

Finally: a trusted repository?

CoreTrustSeal

- *Lightweight self certification service for research archives*
- *Peer-review*
- *Based on Data Seal of Approval and World Data System*

'NESTOR Seal'

- *43 criteria based on DIN 31644*
- *External validation*
- *Service from nestor / DNB*

Repository Audit and Certification

- *ISO16363*
- *3 components, 100+ metrics*
- *Renewable external audit*

Theorize the Change Building Maturity

DPC Rapid Assessment Model

- *How are we doing?*
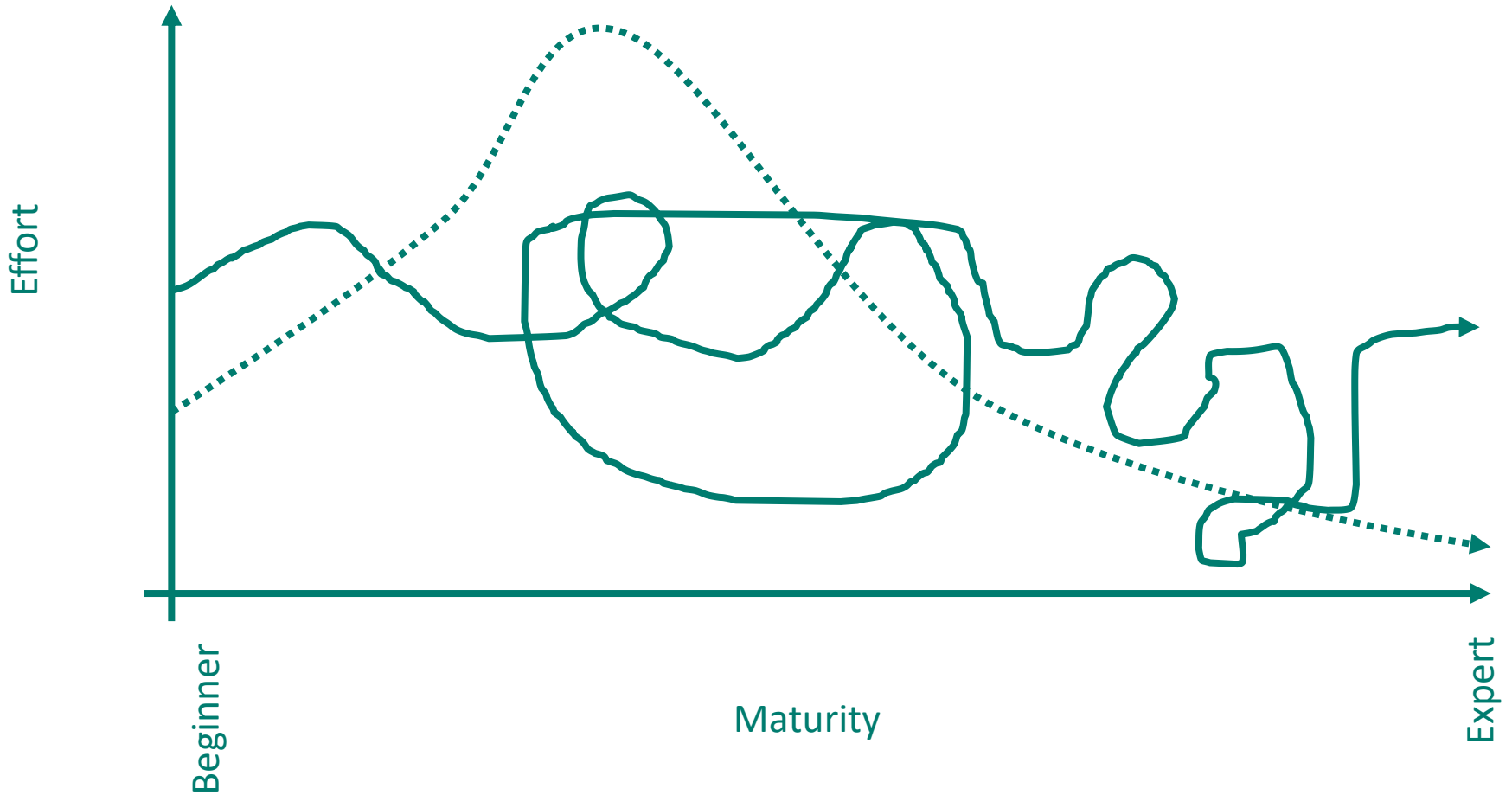- *How should we be doing?*
- *How can we move up?*

NDSA Levels

- *5 functional areas*
- *4 consistent steps*

5 Steps Model

- *Acknowledge*
- *Act*
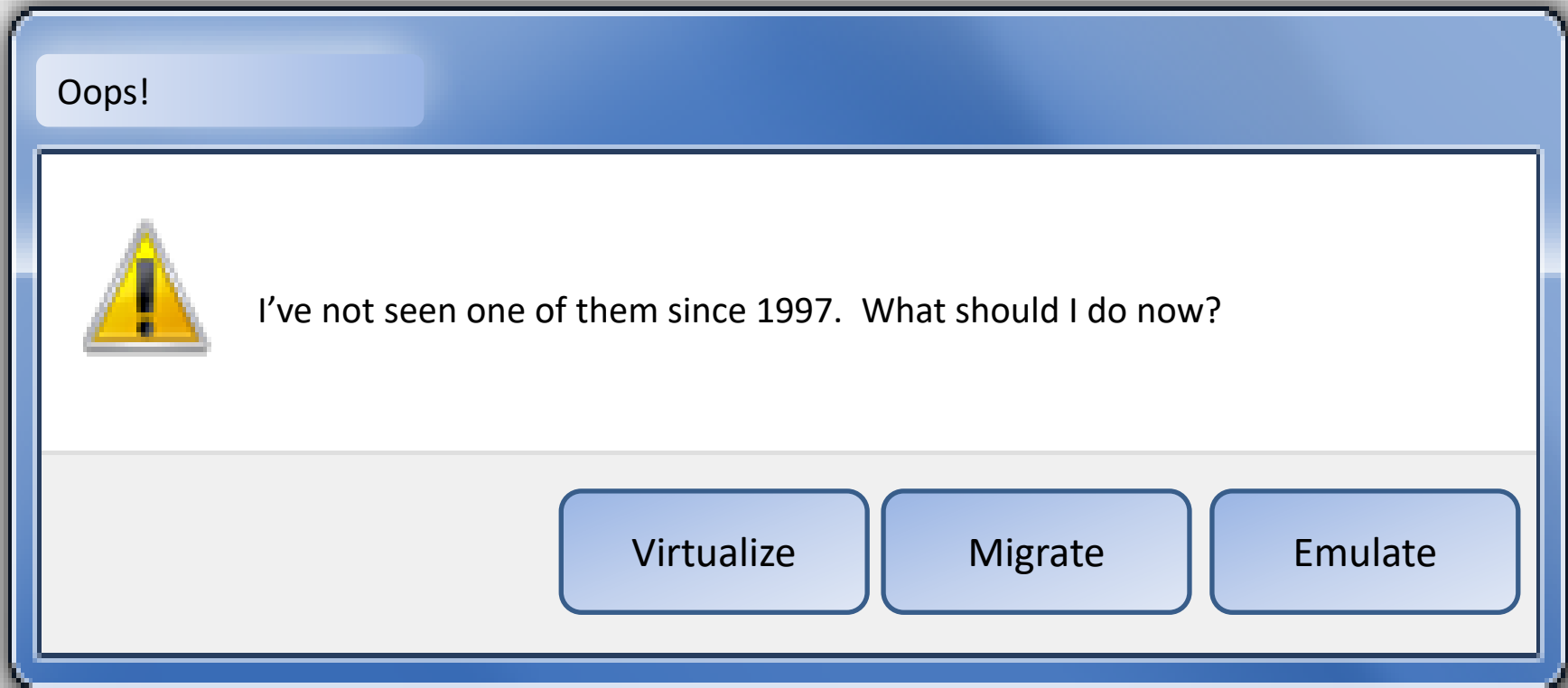- *Consolidate*
- *Institutionalize*
- *Externalize*

@williamkilbride

www.dpconline.org

Reality of change

@williamkilbride

This is what success will look like:
Simple; invisible; intensely clever; user-focussed

Oops!

⚠️ I've not seen one of them since 1997. What should I do now?

[Virtualize] [Migrate] [Emulate]

Pause for response, questions and answers.
Also sharing experience from the audience…

3. What I wish someone had to me 2: Lived Experience

Digital Preservation Not (Yet) Business as Usual

Was a daunting challenge
Decade of research and development
Now a daunting challenge with jargon
Configured like a research problem:
Hyper-specializing
Poor integration with other systems

The Bit List
ENDANGERED
Digital Species

CORPORATE ABANDONMENT IS AS BIG A THREAT TO THE DIGITAL ESTATE AS OBSOLESCENCE

Nook	Friends ReUnited	Yahoo 360
GoogleCode	Yahoo Mail Classic	del.icio.us
GeoCities	Blipfoto	MyBlogLog
GoogleWave	MySpaceBlogs	BeBo
knol	...	...

…money turns out to be the major problem facing the future of our digital heritage. (Rosenthal 2012)

Data creation:
Approx 60% per annum

Storage capacities:
Approx 40% per annum

IT Budgets:
Approx 2% per annum

3. What I wish someone had to me 3: Talking with your neighbours…

DigitalPreservationCoalition

William Kilbride @WilliamKilbride   4h
Paging everyone: thoughts welcome! I am
speaking next week to the National Archives of
Indonesia about #digitalpreservation. It will have a
short session entitled something like 'What I wish I
knew about Digital Preservation before I started'.
1/2

💬 2   🔁 4   ♡ 1   ooo

William Kilbride
@WilliamKilbride

I want to share the hard-won experience of
#digitalpreservation practitioners around the
world. What do you think? Replies below (or
DM if you prefer). 2/2 #digipres

9:14am · 19 May 2021 · TweetDeck

Jaana Pinnick @NottsFinn   4h
Replying to @WilliamKilbride
Don't be put off by the excellency of
technical experts, people skills are just as
important. Talk to your stakeholders - and
find out who they are! Encourage best
practice and remember that taking small
steps is better than doing nothing. Time
spent planning is time well spent.

💬 1   🔁   ❤ 2   ooo

David Underdown @DavidUnderdown9   4h
Replying to @NottsFinn @WilliamKilbride
And of course now you can use
#NLHF_diagram to help you with that
planning and work out which small steps
will have the biggest impact on your
preservation risks.

💬   🔁   ♡ 1   ooo

Neil Jefferies @NeilSJefferies   4h
Replying to @WilliamKilbride
Beware of dogma and "standards", they rot
faster than your media.

💬 1   🔁   ❤ 3   ooo

Neil Jefferies @NeilSJefferies   4h
Replying to @NeilSJefferies @WilliamKilbride
It's rarely just about the technology. People
are potentially your most important asset
and (hopefully different) people are likely to
be the biggest threat to your material.

💬   🔁 1   ♡ 3   ooo

@williamkilbride

Karyn Williamson @ArchivesnKaryn 2h
Replying to @WilliamKilbride
That there isn't a one size fits all solution. And noone knows completely what they are doing. The community is your best asset.

💬 🔁 1 ❤️ 6 ⋯

Jenny Mitcham @Jenny_Mitcham 🔒 5m
Replying to @WilliamKilbride
Remember there is no single right way of doing digital preservation and it is not always realistic to strive for 'best practice'. Think about what 'good enough' might look like for now and take steps towards achieving this. Don't let perfect be the enemy of 'good enough'!

💬 🔁 ♡ ⋯

Christopher Hilton @CHilton_BB 41m
Replying to @Jenny_Mitcham @WilliamKilbride
Yes – at its most extreme, I remember an organisation I dealt with when I was at Wellcome, who told me "we're very small, we don't have the technology, resources or knowledge to have a digital preservation policy – we're going to identify key documents, print them out and..." 1/2

💬 1 🔁 ♡ 1 ⋯

Christopher Hilton @CHilton_BB 40m
Replying to @CHilton_BB @WilliamKilbride @Jen...
"...add them to the paper filing system." They were startled when I pointed out that actually this *was* a digital preservation policy – they'd spotted the need, assessed their capacities and taken action that was appropriate to their ability. 2/2

💬 🔁 ♡ 1 ⋯

kevin wong @wzhkevin 1h
Replying to @WilliamKilbride
Digital preservation capability can be developed incrementally. You don't have to try to do everything at once, and you probably shouldn't. What can you do now? Maybe start by doing that.

💬 🔁 ♡ ⋯

Spellbound Blog @spellboundblog 20m
Replying to @WilliamKilbride
That #DigitalPreservation requires ongoing investment of time and resources. It isn't as simple as putting paper in a box with physical and environmental controls in place.

💬 🔁 ❤️ ⋯

Heidi Mcintosh @Mcintoshheidi 28m
Replying to @WilliamKilbride
Doing something, however small, is better than doing nothing at all. You don't need an all-singing, all-dancing system. And in many cases just getting the system without having policies and procedures in place, and without bringing people along with you could be catastrophic.

💬 🔁 ♡ ⋯

Simon P Wilson @simon_archivist 58s
Replying to @WilliamKilbride
At some point stop reading the text books and start playing with the tools and sample data. It helps make sense of things and helps make the learning fun. Make friends in ICT and other archives - you are not alone.

💬 🔁 ♡ ⋯

Sharon McMeekin 🏴󠁧󠁢󠁳󠁣󠁴󠁿🇪🇺 @SharonMcM... 15h
Replying to @WilliamKilbride
- Aiming for "best" practice is a white whale hunt, think "good" instead
- OAIS is not the answer, every org is different
- When planning think about risks you face and benefits you want to accrue more than standards you want to meet
💬 1 🔁 ❤ 4 ⋯

Sharon McMeekin 🏴󠁧󠁢󠁳󠁣󠁴󠁿🇪🇺 @SharonMcM... 15h
Replying to @SharonMcMeekin @WilliamKilbride
- Maturity Models offer the opportunity for flexibility, setting your own goals & continuous improvement
- Start small and build rather than doing it all at once. Check out the demos for some free/cheap tools in the #digipres Handbook
dpconline.org/handbook/tool-...
💬 1 🔁 ❤ 3 ⋯

Sharon McMeekin 🏴󠁧󠁢󠁳󠁣󠁴󠁿🇪🇺
@SharonMcMeekin

Replying to @SharonMcMeekin @WilliamKilbride
- Spending as much time (or more) on advocacy compared to developing processes is (unfortunately) the norm
And most importantly:
- People are the most important part of Digital Preservation. Value staff, build skills, share stories.

Euan Cochrane @euanc 14h
Replying to @WilliamKilbride
The Digital preservation systems are the part we know we will change over time. Don't get hung up on qualities of current systems and or open-source/proprietary nature, just have a good exit strategy.
💬 1 🔁 ❤ 6 ⋯

Remco van Veenendaal
@RvanVeenendaal

Replying to @euanc @WilliamKilbride
Related: that in the end, solving digital preservation challenges is more about people than about technology.

6:43pm · 19 May 2021 · Twitter for iPhone

Dragan Espenschied @despens 14h
Replying to @WilliamKilbride
Double your budget, and the double it again.
💬 1 🔁 ❤ 3 ⋯

Nick Krabbenho:ft 🆗
@NKrabben

Replying to @despens @WilliamKilbride
Then look at what your doing, and figure out ways to cut costs. (lossless compression, less fiddly workflows, etc) Managing more collections is often more valuable to an organization than being a bottleneck.

Crystal Sanchez @cristalyze 12h
Replying to @WilliamKilbride
Digital preservation is people doing the work, then policy, then technology. Or all of them, but people forget about the physical labor needed..
💬 🔁 ❤ ⋯

Jaye Weatherburn @jayechats 7h
Invest in advocacy & education skills for culture change and for building your people-force to tackle preservation challenges. Raising money know-how helps. Communities are everything for sharing and learning from others - #AusPreserves & @dpc_chat

Anne-Flore Laloë @ItStartsWithARC 13h
Replying to @WilliamKilbride
What helped me a lot was breaking down Digital Preservation into a bunch of smaller tasks & tiny steps. Eventually, the bigger picture started coming together and it was less intimidating.

Also, talking about my problems with people!🗣 @simon_archivist & @HJoyNam
💬 🔁 ❤ ⋯

www.dpconline.org

Euan Cochrane @euanc 12h
Replying to @WilliamKilbride
Advocacy and fundraising are the most important digital preservation activities, always.

💬 1   🔁   ❤️ 5   ⚬⚬⚬

Shira Peltzman @shirapeltzman 12h
Replying to @euanc @WilliamKilbride
+1 to @euanc but I would also add teaching/education to the mix

💬 1   🔁   ❤️ 4   ⚬⚬⚬

Euan Cochrane @euanc 12h
Replying to @shirapeltzman @WilliamKilbride
Yeah I should have been more nuanced but I liked the "soundbite" - It's more that, especially early on, the technology is not the top problem at all, so the first hire probably shouldn't be "technical" (developer/"digital archivist"), instead someone with more soft-skills

💬 1   🔁   ❤️ 4   ⚬⚬⚬

Euan Cochrane
@euanc

Replying to @euanc @WilliamKilbride @shirapeltzman

and on the "digital archivist" note - I mean, not a practitioner who expects to have the resources to get stuff done

7:20pm · 19 May 2021 · TweetDeck

Andrew Waugh @MAndrewWaugh 6h
Replying to @WilliamKilbride
It only seems complex and expensive because it's new. The cost and complexity of preserving physical objects has been normalised over hundreds of years. Think of the cost of building a physical library (or archive), or of the purple Darth Vader.

💬 1   🔁   ❤️   ⚬⚬⚬

Andrew Waugh @MAndrewWaugh 5h
Replying to @MAndrewWaugh @WilliamKilbride
Also normalised are the risks of physical preservation (e.g. deterioration due to access or storage, or destruction due to war or disaster). The risks of digital preservation have not yet been normalised, so seem massive.

💬 1   🔁   ❤️   ⚬⚬⚬

Andrew Waugh
@MAndrewWaugh

Replying to @MAndrewWaugh @WilliamKilbride

This is not to say that digital risks and costs aren't higher.

But we've only been doing this seriously for 20 years. We've been serious preserving physical objects for hundreds of years. For physical, we know what to do, accept what it costs, & overlook the residual risk.
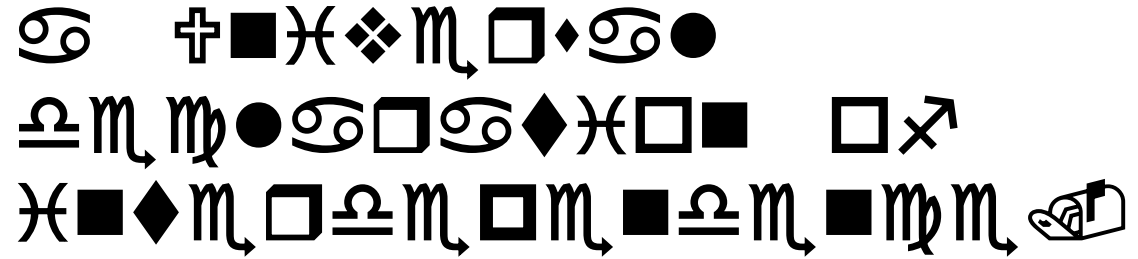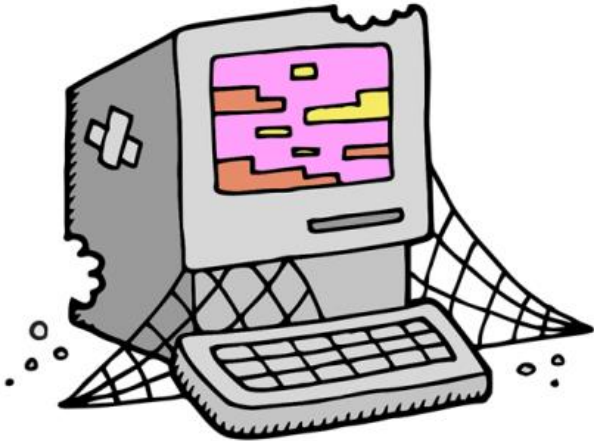
1:34am · 20 May 2021 · Twitter Web App

www.dpconline.org

Pause for response, questions and answers.
Also sharing experience from the audience…
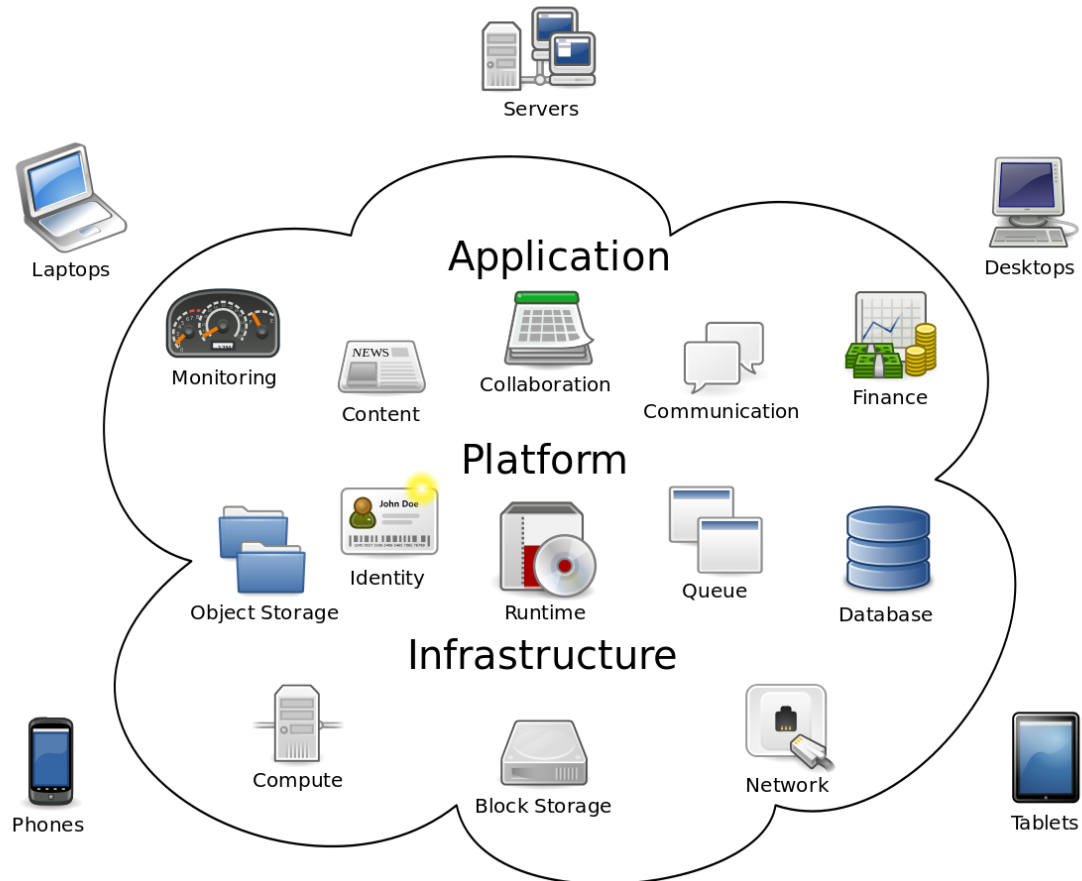
4. Emerging Trends:
planning for an uncertain future

DigitalPreservation**Coalition**

**Emerging trend one:
the boundary of data**
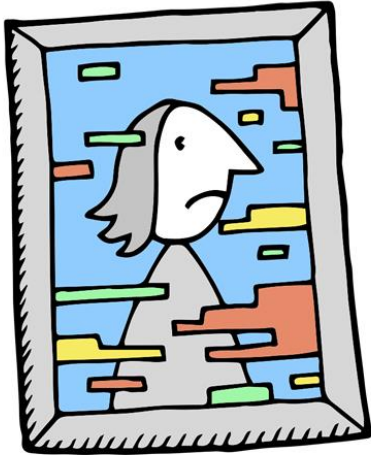

⚋ ☦■✥✤❡□◆⚋● ⚎❡❡●⚊□⚊◆✥□■ □↗ ✥■◆❡□⚎❡□❡■⚎❡■❡❡⍾

or

A universal declaration of interdependence.

Cloud Computing

Cloud computing is the delivery of computing as a service rather than a product

I can no longer distinguish between data, information, processes, software and applications

The file is not the atomic unit of preservation: it never has been

Preservation is an evolving problem: it's getting harder.

**Emerging trend two:
Digital Preservation and AI**

1. AI for better digital preservation:
   – ingest, automated metadata extraction, technology watch, risk management
2. AI as Producer
   – preserving the outputs of AI
3. AI as consumer:
   – machine to machine data processing, computational access, text mining, pattern recognition …
4. AI as object of preservation:
   – preserving algorithm, reproducibility, POV
5. AI as target of archival disruption
   – Make the ethical challenge and change for better AI

@williamkilbride

www.dpconline.org

DigitalPreservationCoalition

Emerging trend four:
The Digital Preservation Community is Growing
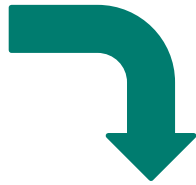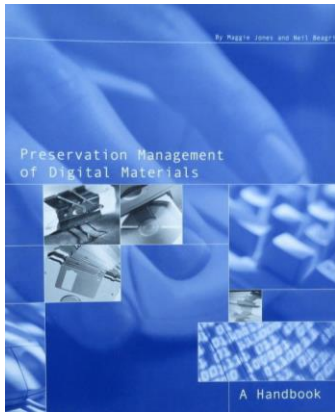
- DPC Members: 33 in 2009, 118 in 2021
- DP Awards: 1 in 2010, 7 in 2020
- DPC Staff: 2 in 2009, 10 in 2021

- More diverse
- More data
- More complicated
- More use cases
- More regulated
- More requirements
- More demanding?

How do we respond to change like this?

@williamkilbride

www.dpconline.org

Technology: the paradox of obsolescence

- Digital preservation systems are products of their own generation
- They are subject to the same processes of obsolescence as the objects they safeguard
- They are a contingent solution to an enduring problem
- The drivers of this change are outside of our control!!!

How do we respond to change like this?

@williamkilbride

www.dpconline.org

Where am I going with this?

The People

- We understand that we need to renew the technical infrastructure in response to emerging technology.
  We have been quite good at this.
- We also need to **renew constantly the social infrastructure** in which digital preservation occurs.
- Especially as the community grows, becomes more complicated

The Technology

- Digital preservation is getting harder through time, not easier and it's never really 'done'
- Digital preservation is not an app or a workflow or a department
- Digital preservation is a commitment through time

Pause for response, questions and answers.
Also sharing experience from the audience...

5. Digital Preservation as Community

The Digital Preservation Coalition

... *a secure digital legacy* ...

Digital**Preservation**Coalition

(your logo here)

@williamkilbride

www.dpconline.org

We do 6 things …

- **Advocacy**
- **Community engagement**
- **Workforce development**
- **Capacity Building**
- **Good Practice and Standards**
- **Management and governance**

@williamkilbride

www.dpconline.org

DPC as a partner and friend!

Join us: http://www.dpconline.org/join-us

@williamkilbride

www.dpconline.org

So ... a warm welcome and invitation:

Let's work together on this

because I want to learn from you.

@williamkilbride

www.dpconline.org