arkivum

Bringing archived data to life

A Career in Digital Preservation

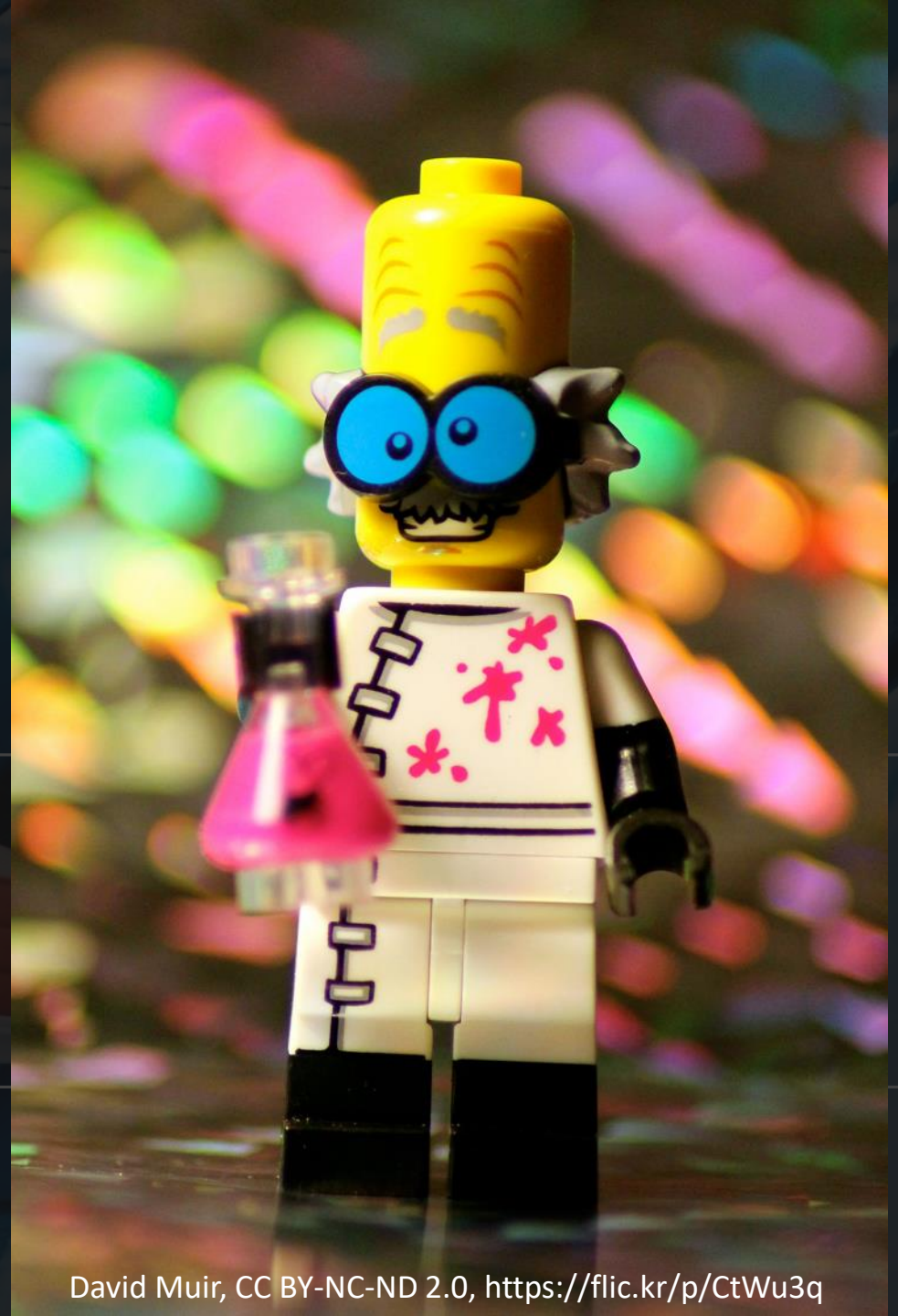
Matthew Addis

DPC Student Conference, 23 January 2019

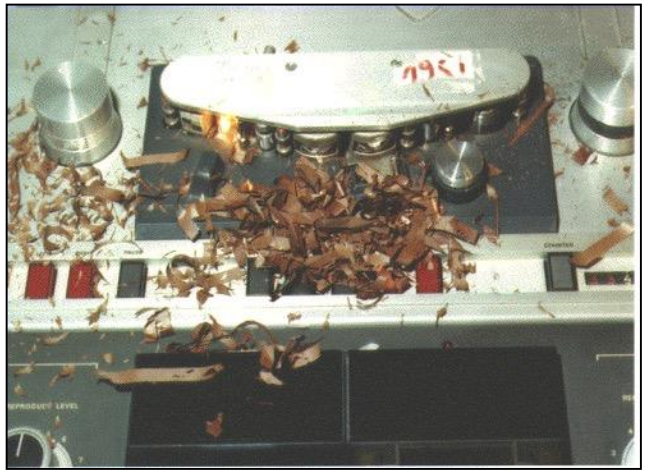
Career
/kəˈrɪə/

verb

move swiftly and in an uncontrolled way.
"the coach careered across the road and went through a hedge"

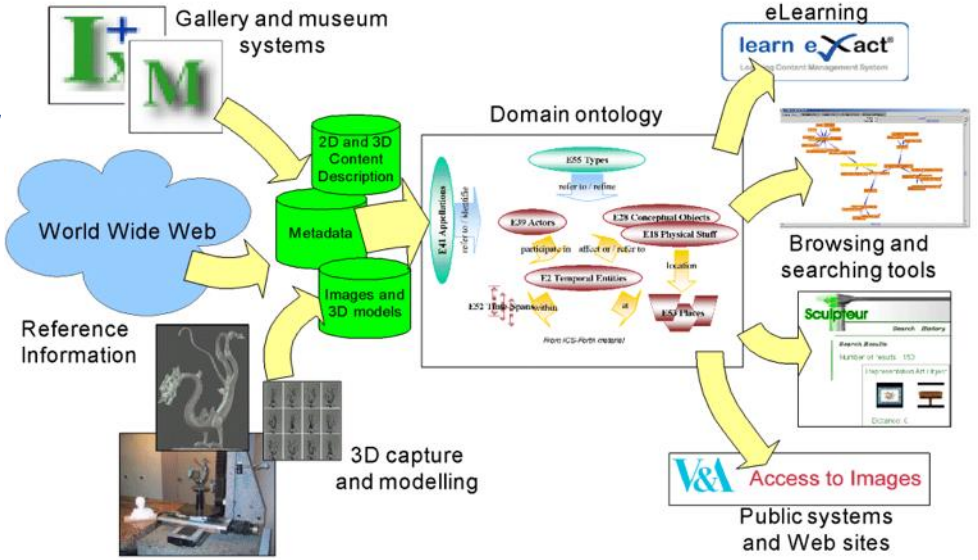


David Muir, CC BY-NC-ND 2.0, https://flic.kr/p/CtWu3q



https://flic.kr/p/ej5kzL

https://flic.kr/p/8UoMUe


Gallery and museum systems
eLearning
learn eXact
Domain ontology
2D and 3D Content Description
Metadata
Images and 3D models
World Wide Web
Reference Information
3D capture and modelling
Browsing and searching tools
Sculpteur
V&A
Access to Images
Public systems and Web sites

Warner Bros, Fox, Disney, Miramax, Evolutions, AMS pictures, Shaw cable, Smoke and Mirrors, BBC, B&G, United Space Alliance, BSkyB, Turner, Kingdom of Bahrain, The Pixel Farm, Film Museum NL, Pegasus Pictures, CineSite, FrameStore

Arkivum Today

>95% Customer renewal rate	100+ Customers	>80% CAGR	2011 Company founded
9 Go to Market Partners	>11 Petabytes under mgmt	3 Supporting investors	40+ Employees in UK, US & India

Neilson Financial Services

Bristol Genetics Laboratory

THE FRANCIS CRICK INSTITUTE

JAGUAR HERITAGE TRUST

LSE

itma Irish Traditional Music Archive Taisce Cheol Dúchais Éireann

Royal Botanic Gardens Kew

TATE

MoMA

OXFORD Molecular Diagnostics Centre

NHS North Bristol NHS Trust

SW SOUTH WEST HERITAGE TRUST

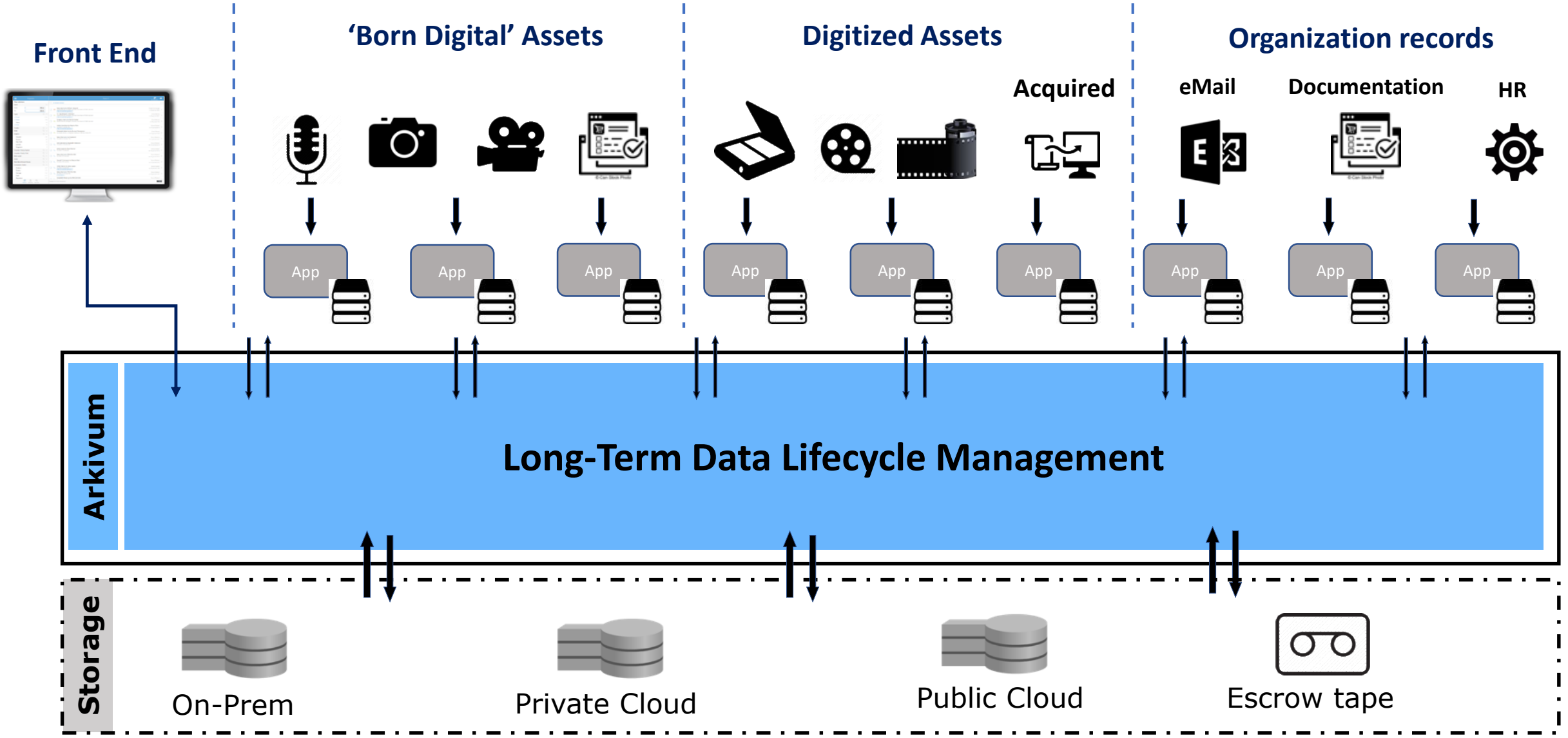
idorsia

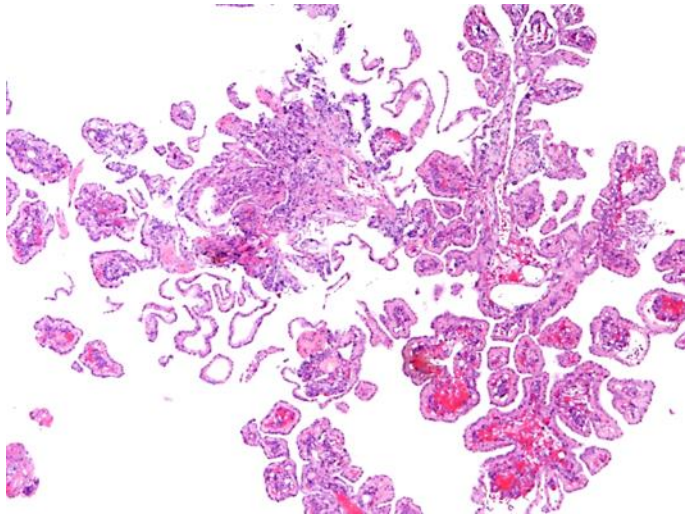
The Frick Collection

UNIVERSITY OF WESTMINSTER

The University of Nottingham

The Open University

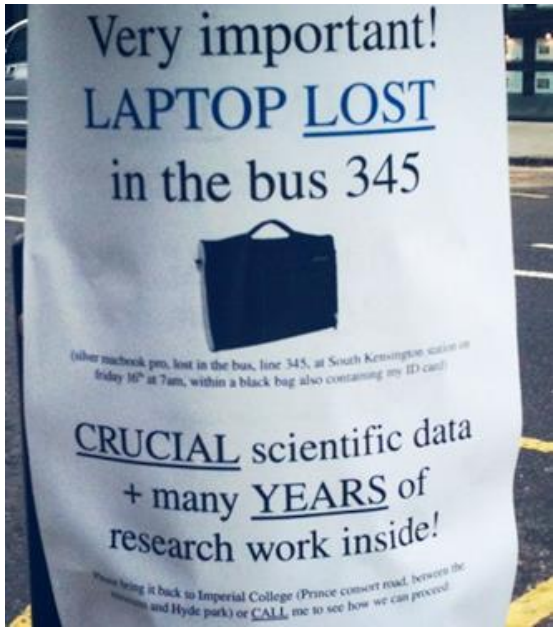

Front End
'Born Digital' Assets
Digitized Assets
Organization records
Acquired
eMail
Documentation
HR
App
Arkivum
Long-Term Data Lifecycle Management
Storage
On-Prem
Private Cloud
Public Cloud
Escrow tape



Ed Uthman, CC-BY-SA-2.0, https://flic.kr/p/muA53

Scott Beale, CC-BY-NC-ND-2.0, https://flic.kr/p/nNMX9d

Master phillip, CC BY-NC-ND 2.0, https://flic.kr/p/8Au7Zi



Romuald Le Peru, CC-BY-NC-2.0, https://flic.kr/p/3a8m7p

A typical day (there isn’t one)

Sharing Knowledge (Preservation Action Registries)

Sharing preservation good practice in communities and between preservation systems

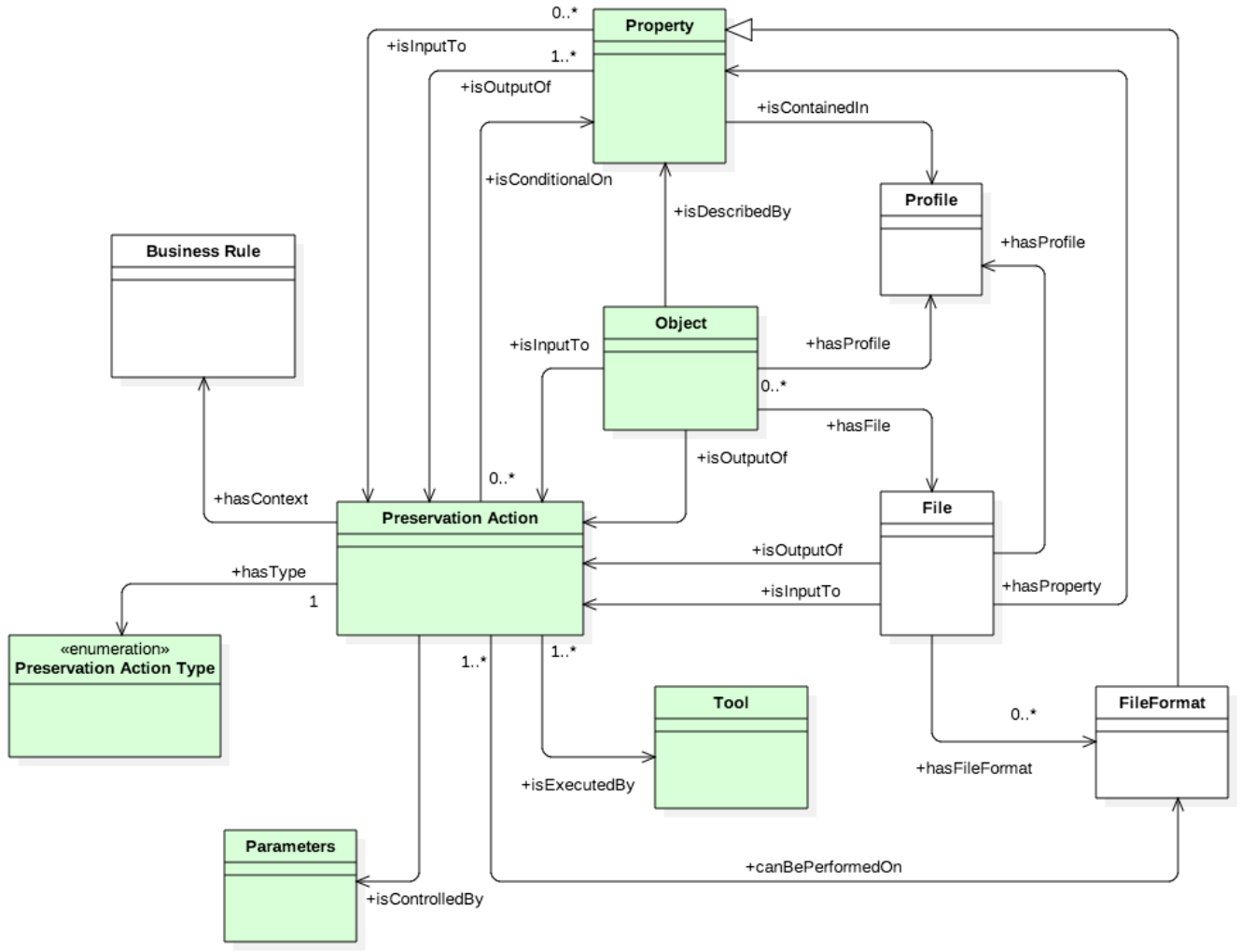
artefactual

Open Preservation Foundation

Jisc

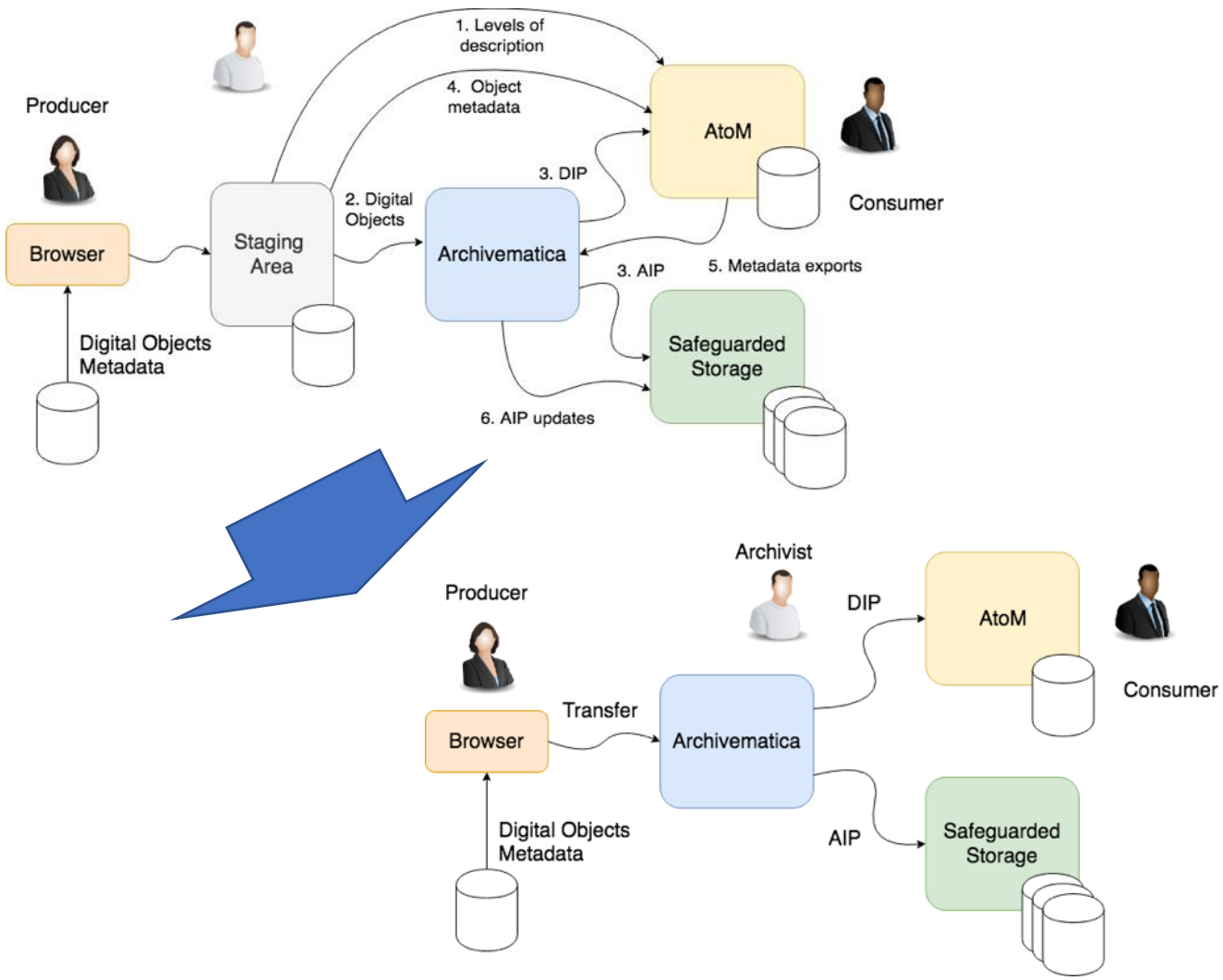
arkivum
Bringing archived data to life

Preservica
Active digital preservation



https://doi.org/10.6084/m9.figshare.6628418

Making Life Easier (Preservation Workflows)



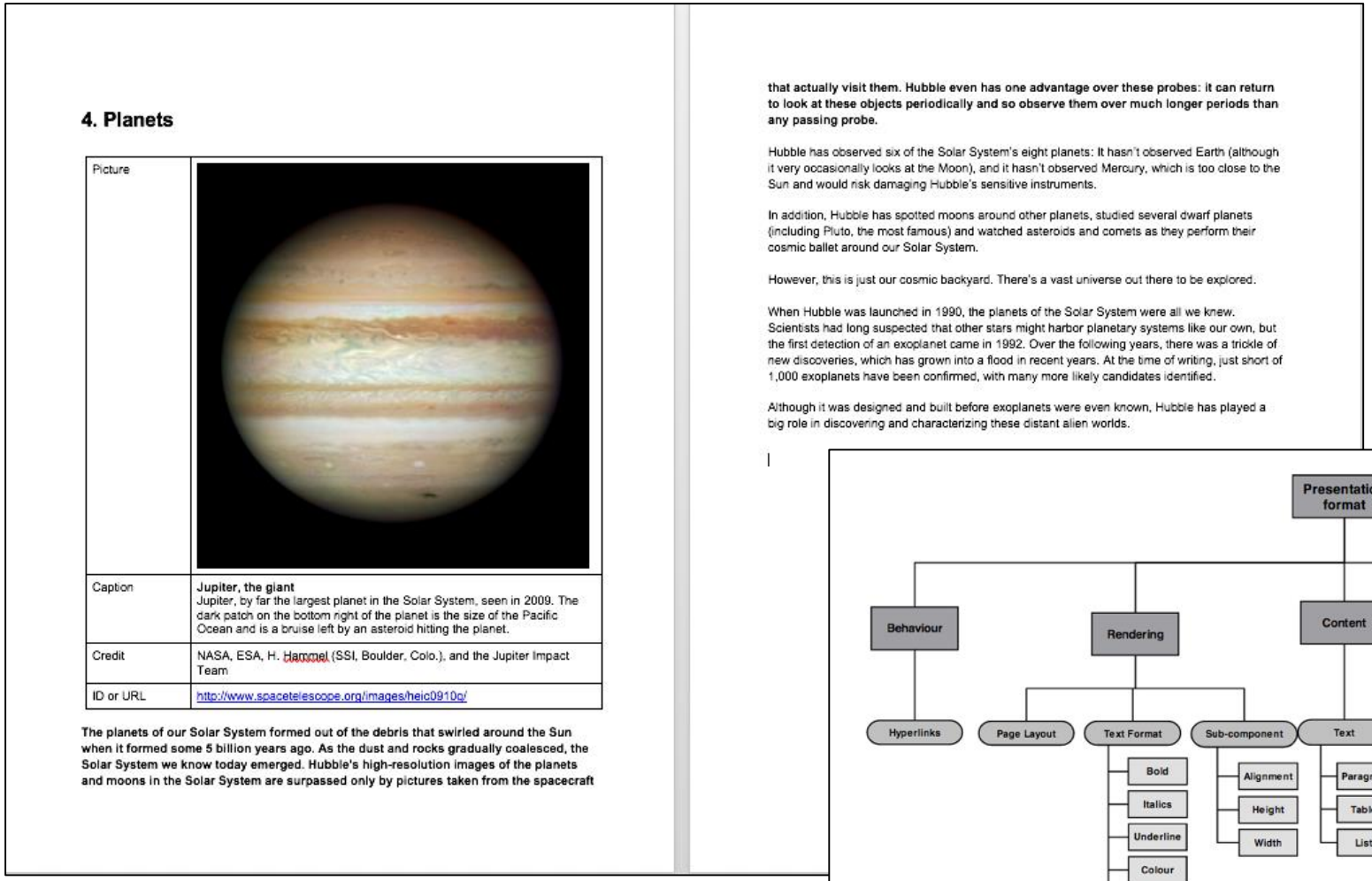
UNIVERSITY OF WESTMINSTER

LSE

THE ARCHIVES OF THE EPISCOPAL CHURCH

BIRMINGHAM CITY University

Preservation Tools and Techniques (Preserving Office Formats)



4. Planets

Picture	
Caption	**Jupiter, the giant** Jupiter, by far the largest planet in the Solar System, seen in 2009. The dark patch on the bottom right of the planet is the size of the Pacific Ocean and is a bruise left by an asteroid hitting the planet.
Credit	NASA, ESA, H. Hammel (SSI, Boulder, Colo.), and the Jupiter Impact Team
ID or URL	http://www.spacetelescope.org/images/heic0910g/

The planets of our Solar System formed out of the debris that swirled around the Sun when it formed some 5 billion years ago. As the dust and rocks gradually coalesced, the Solar System we know today emerged. Hubble's high-resolution images of the planets and moons in the Solar System are surpassed only by pictures taken from the spacecraft

that actually visit them. Hubble even has one advantage over these probes: it can return to look at these objects periodically and so observe them over much longer periods than any passing probe.

Hubble has observed six of the Solar System's eight planets: It hasn't observed Earth (although it very occasionally looks at the Moon), and it hasn't observed Mercury, which is too close to the Sun and would risk damaging Hubble's sensitive instruments.

In addition, Hubble has spotted moons around other planets, studied several dwarf planets (including Pluto, the most famous) and watched asteroids and comets as they perform their cosmic ballet around our Solar System.

However, this is just our cosmic backyard. There's a vast universe out there to be explored.

When Hubble was launched in 1990, the planets of the Solar System were all we knew. Scientists had long suspected that other stars might harbor planetary systems like our own, but the first detection of an exoplanet came in 1992. Over the following years, there was a trickle of new discoveries, which has grown into a flood in recent years. At the time of writing, just short of 1,000 exoplanets have been confirmed, with many more likely candidates identified.

Although it was designed and built before exoplanets were even known, Hubble has played a big role in discovering and characterizing these distant alien worlds.

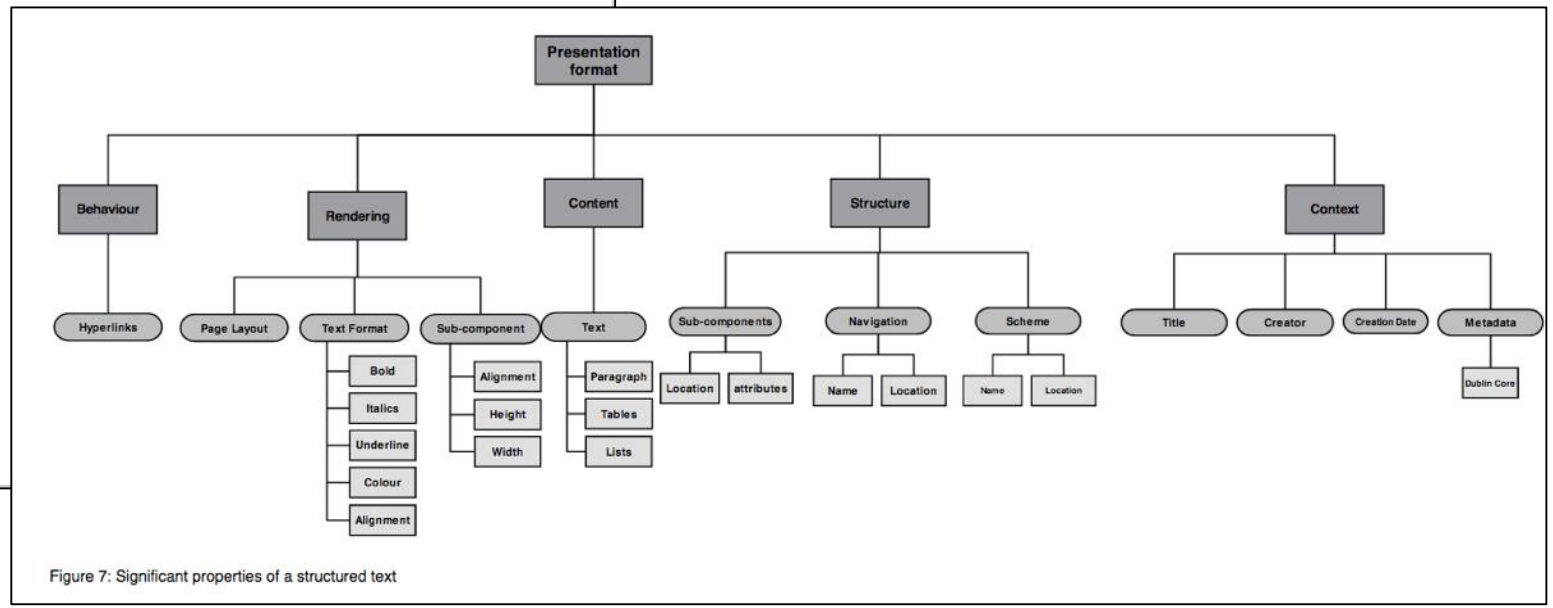
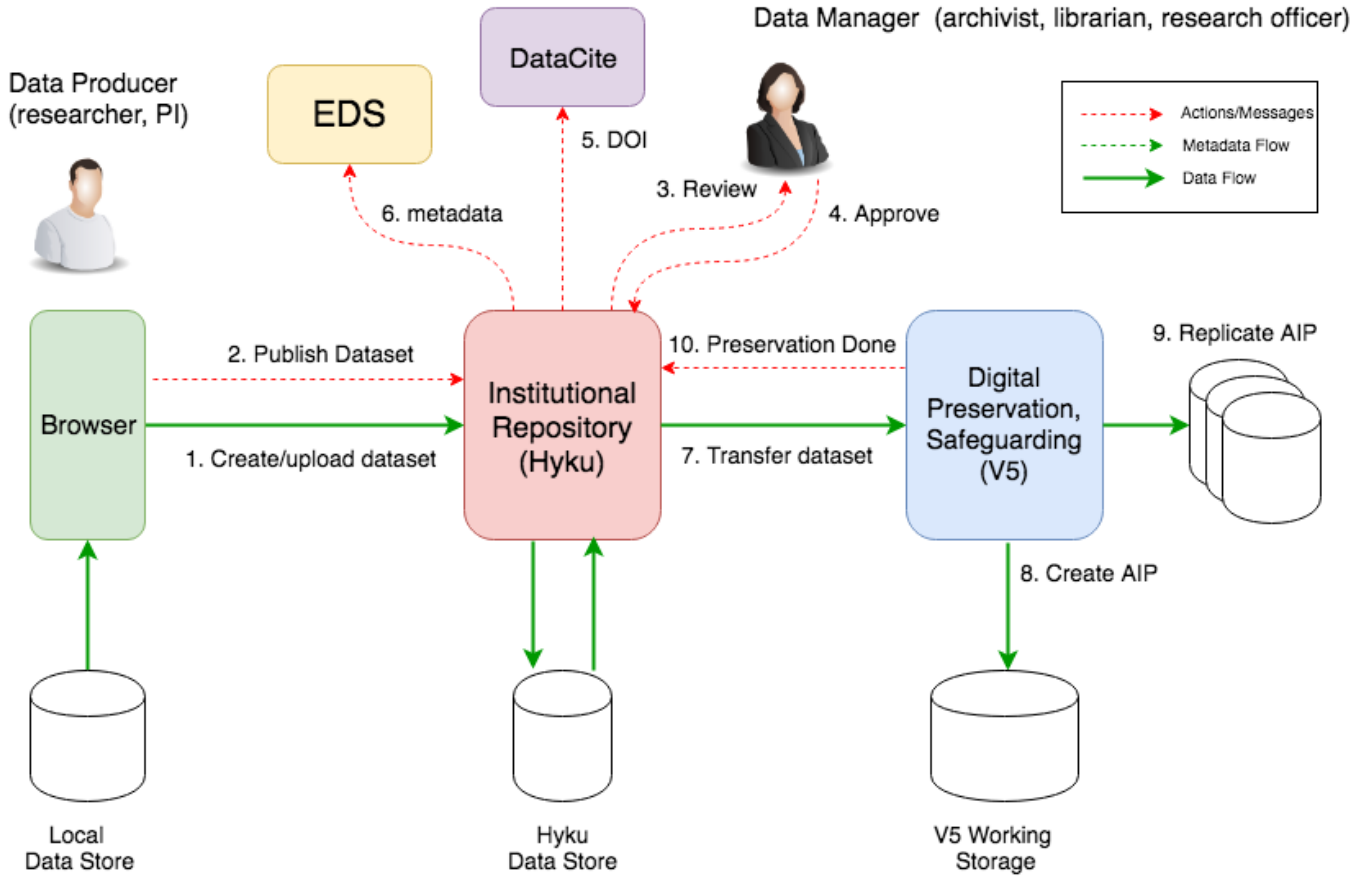


Figure 7: Significant properties of a structured text

Significant properties, lossy conversions, preservation v.s. access, open standards, open specifications, open source

https://www.kdl.kcl.ac.uk/fileadmin/documents/digifutures/materials/preservation/DF09_prsrv_knight-definingSigProperties.pdf

Preservation in New Domains (Research Data)



Joined up systems

Disappearing preservation

F.A.I.R

Long-term use and re-use

CODE OCEAN BETA

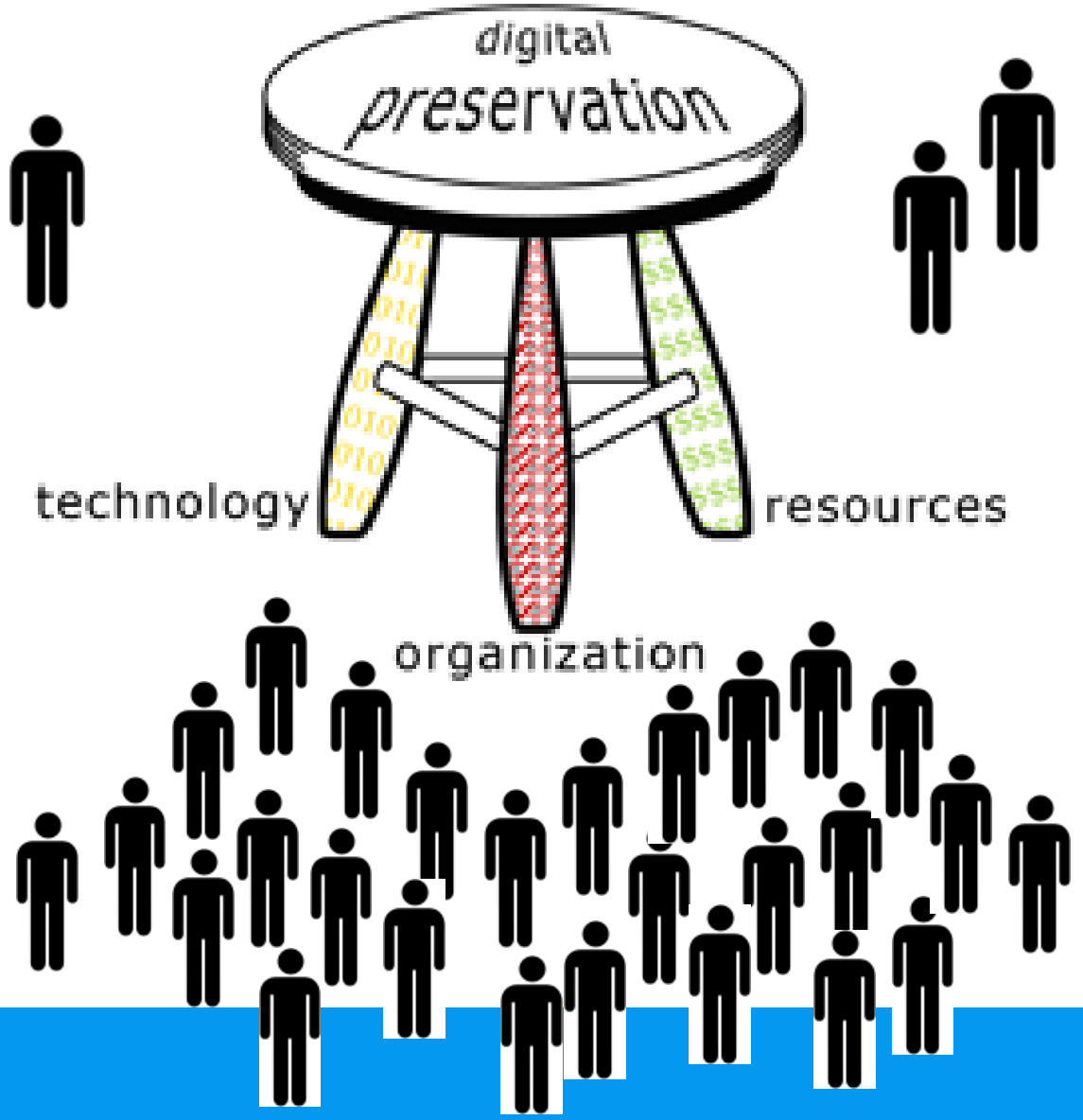
IwFLA
Legacy Environments at Your Fingertips

ReproZip

Software Preservation Network

(Just some of) What I wish I knew before I started

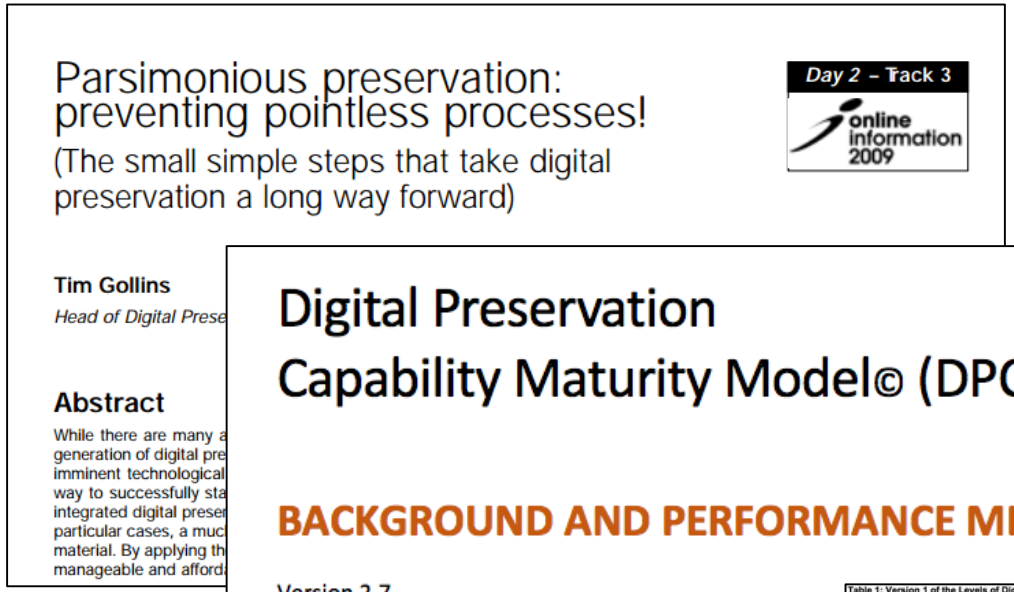
1. Not all legs of the preservation stool are equal

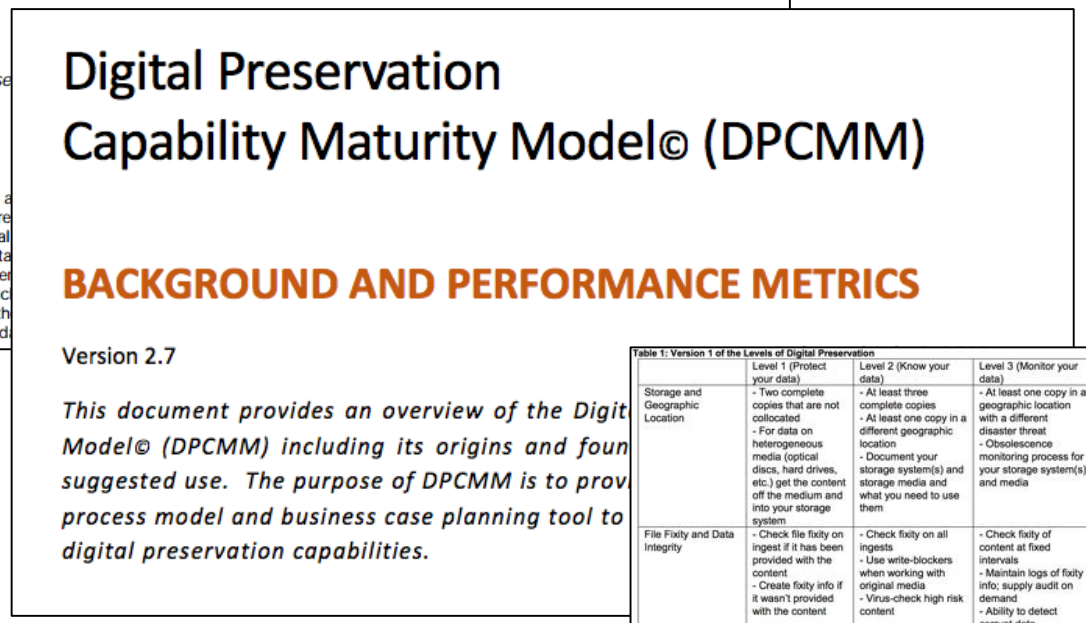


2. Starting simple is OK

S. Farik, CC BY, https://flic.kr/p/bvTM3k

Maturity models and assessment frameworks





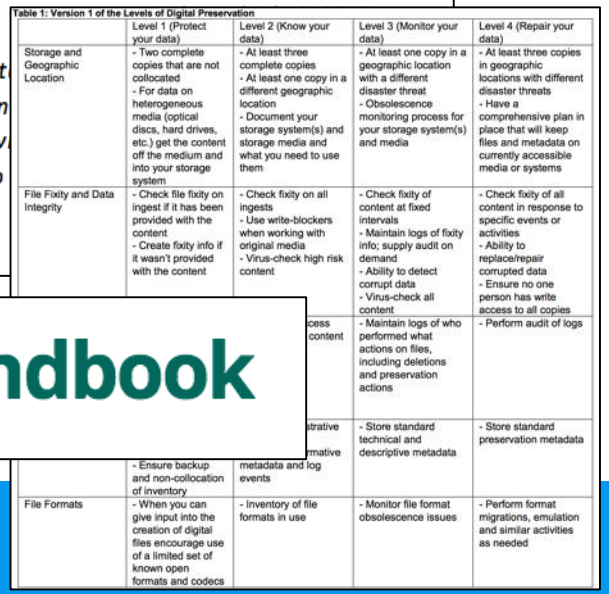
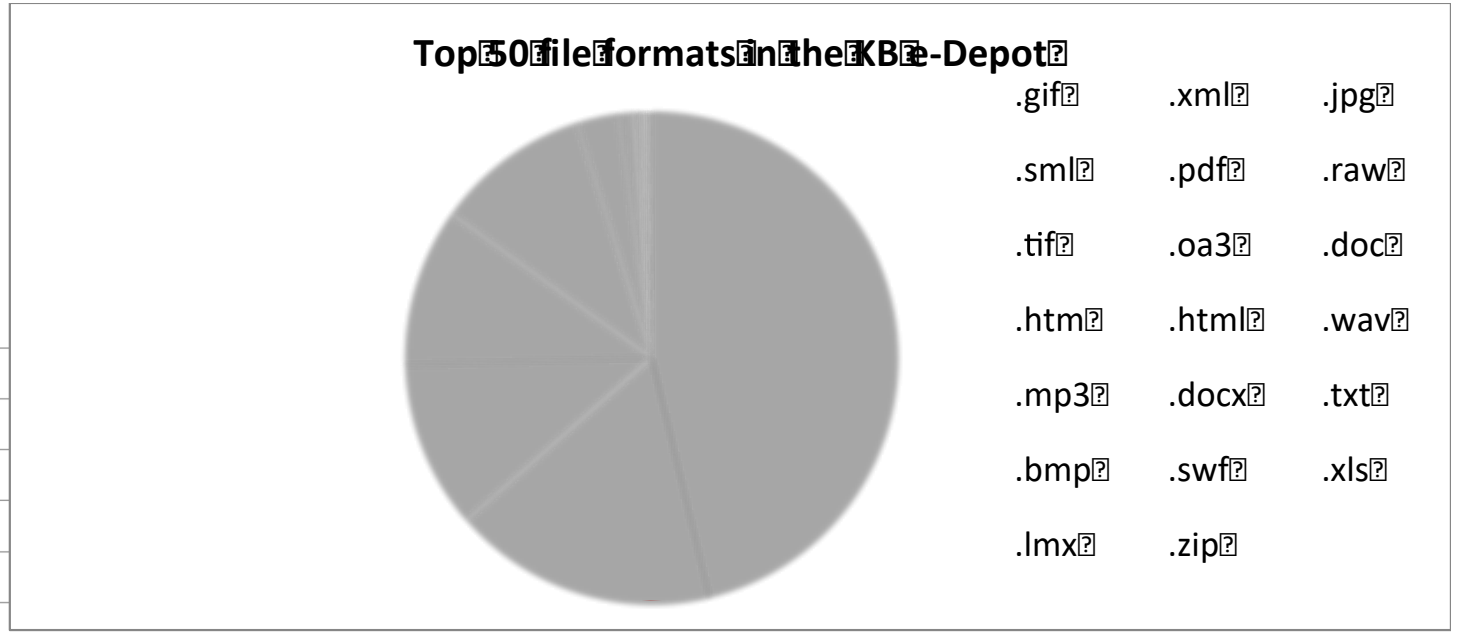


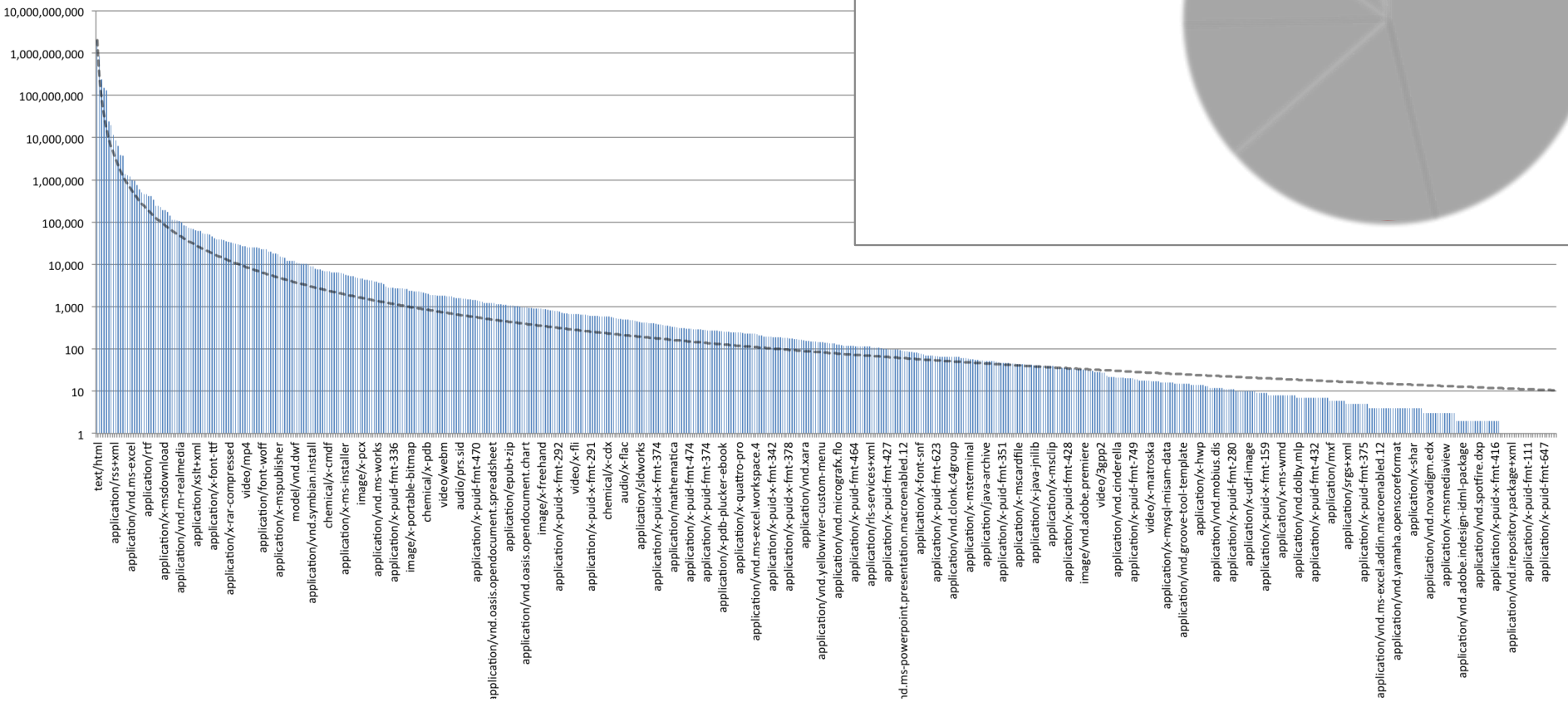
Table 1: Version 1 of the Levels of Digital Preservation

	Level 1 (Protect your data)	Level 2 (Know your data)	Level 3 (Monitor your data)	Level 4 (Repair your data)
Storage and Geographic Location	- Two complete copies that are not collocated - For data on heterogeneous media (optical discs, hard drives, etc.) get the content off the medium and into your storage system	- At least three complete copies - At least one copy in a different geographic location - Document your storage system(s) and storage media and what you need to use them	- At least one copy in a geographic location with a different disaster threat - Obsolescence monitoring process for your storage system(s) and media	- At least three copies in geographic locations with different disaster threats - Have a comprehensive plan in place that will keep files and metadata on currently accessible media or systems
File Fixity and Data Integrity	- Check file fixity on ingest if it has been provided with the content - Create fixity info if it wasn't provided with the content	- Check fixity on all ingests - Use write-blockers when working with original media - Virus-check high risk content	- Check fixity of content at fixed intervals - Maintain logs of fixity info; supply audit on demand - Ability to detect corrupt data - Virus-check all content	- Check fixity of all content in response to specific events or activities - Ability to replace/repair corrupted data - Ensure no one person has write access to all copies
		cess content	- Maintain logs of who performed what actions on files, including deletions and preservation actions	- Perform audit of logs
		strative rmative	- Store standard technical and descriptive metadata	- Store standard preservation metadata
	- Ensure backup and non-collocation of inventory	metadata and log events		
File Formats	- When you can give input into the creation of digital files encourage use of a limited set of known open formats and codecs	- Inventory of file formats in use	- Monitor file format obsolescence issues	- Perform format migrations, emulation and similar activities as needed

http://blog.kbresearch.nl/2015/04/29/top-50-file-formats-in-the-kb-e-depot/

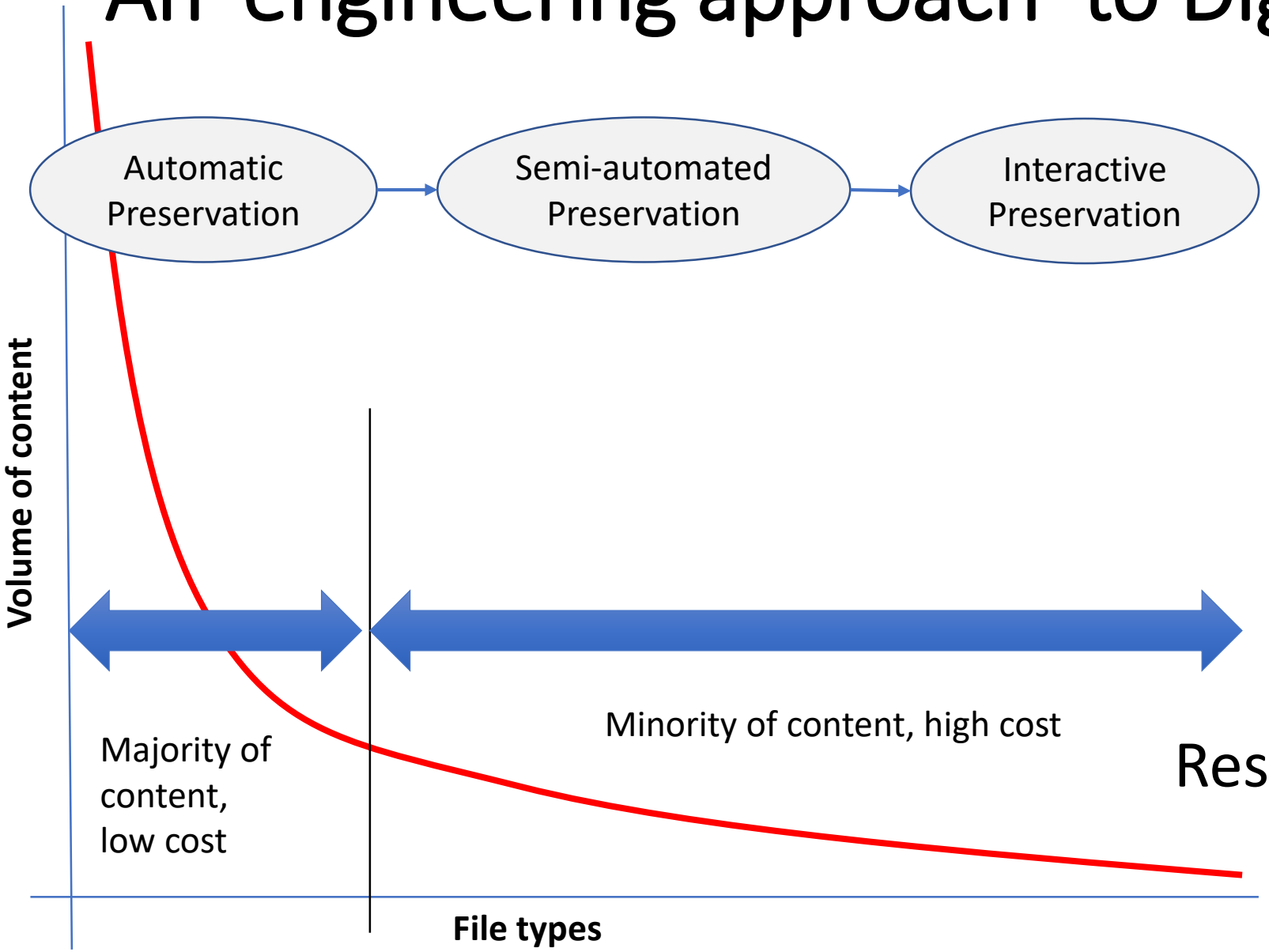
3. Prioritise

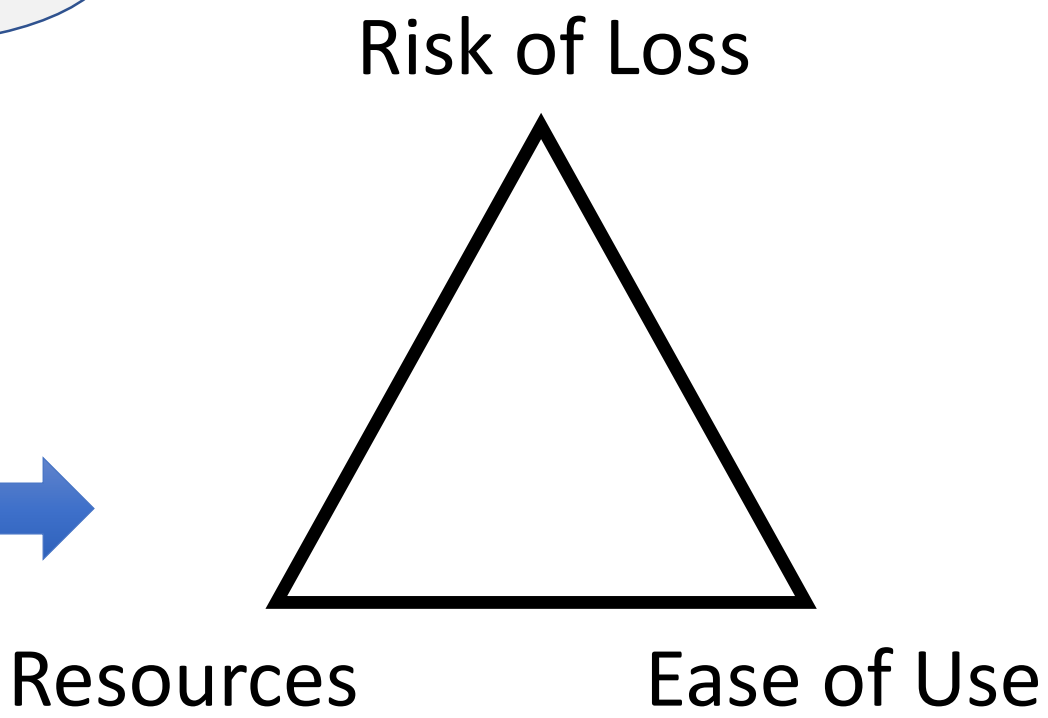




Andrew Jackson, British Library. https://speakerd.s3.amazonaws.com/presentations/a2fd5875bd0346a4be56f14d7900f015/dpc-searching-for-obsolescence.pdf

An ‘engineering approach’ to Digital Preservation





Digital Preservation is Totally Awesome!

matthew.addis@arkivum.com orcid.org/0000-0002-3837-2526 www.arkivum.com