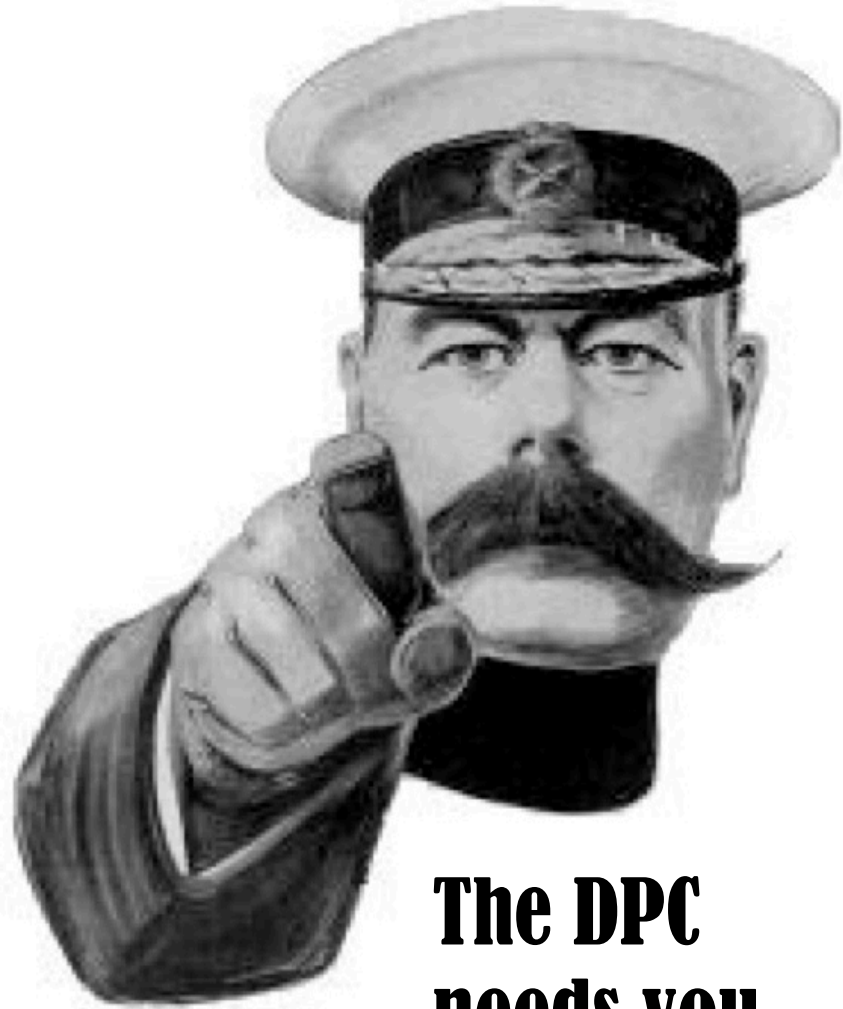


Bit preservation: opportunities and burdens

/Arkivum


Every bit archived

What I wish I knew before I started



**The DPC
needs you**

What I also wish I knew before I started

 LIBRARY OF CONGRESS


ASK A LIBRARIAN

DIGITAL COLLECTIONS

LIBRARY CATALOGS

The Library of Congress > Blogs > Digital Preservation > We Want You Just the Way You Are: The What, Why and When of Fixity

Search Search Loc.gov



[Print](#) [Subscribe](#) [Share/Save](#)

About This Blog

Categories

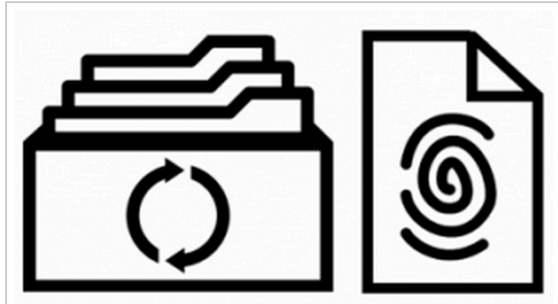
- [At the Museum](#)
- [Content Matters Interview](#)
- [Digital Content](#)
- [Education and Training](#)
- [FADGI](#)
- [Inside the Library](#)
- [Insights Interview](#)
- [NDSA](#)
- [Outreach and Events](#)
- [Partners and Collaboration](#)
- [Personal Archiving](#)
- [Publications and Resources](#)
- [Tools and Infrastructure](#)

We Want You Just the Way You Are: The What, Why and When of Fixity

October 6, 2014 by [Trevor Owens](#)

Fixity, the property of a digital file or object being fixed or unchanged, is a cornerstone of digital preservation. Fixity information, from simple file counts or file size values to more precise [checksums and cryptographic hashes](#), is data used to verify whether an object has been altered or degraded.

Many in the preservation community know they should be establishing and checking the fixity of their content, but less understood is *how*, *when* and *how often*? The National Digital Stewardship Alliance Standards and Practices and Infrastructure working groups have published [Checking Your Digital Content: What is Fixity and When Should I Be Checking It?](#) (PDF) to help stewards of digital objects answer these questions in a way that makes sense for their organization based on their needs and resources.



Icons for Archive and Checksum, Designed by Iconathon Los Angeles, California, US 2013. Public Domain

Bit preservation

Keeping *data* safe and accessible for the long-term

Both a corner and cornerstone

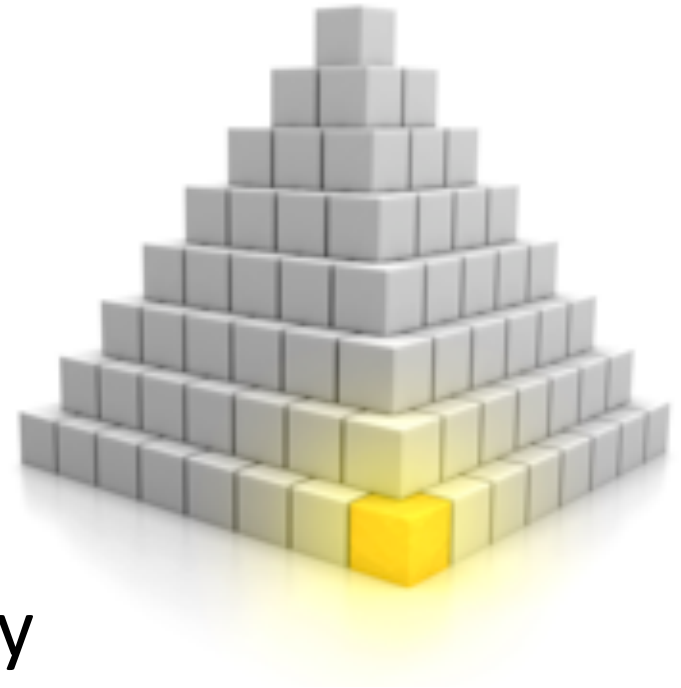
Costs, benefits, sustainability

Rights, licensing, access

Trusted Digital Repositories

File formats, metadata, usability

Bit preservation



Bit preservation

Opportunity:

- Fast and easy access
- Trusted content
- Unlock the value
- Helps with the business case



Burden:

- It never ends, no 'file and forget'
- You can't take your eye off the ball
- Time, money, skills, infrastructure



Toy Story



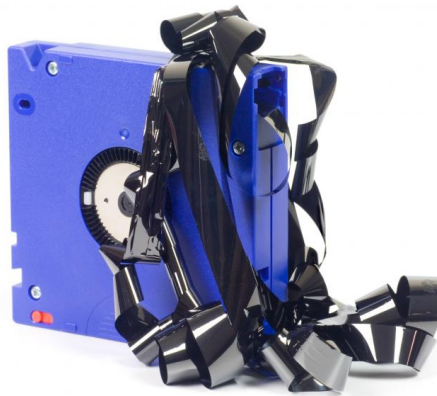
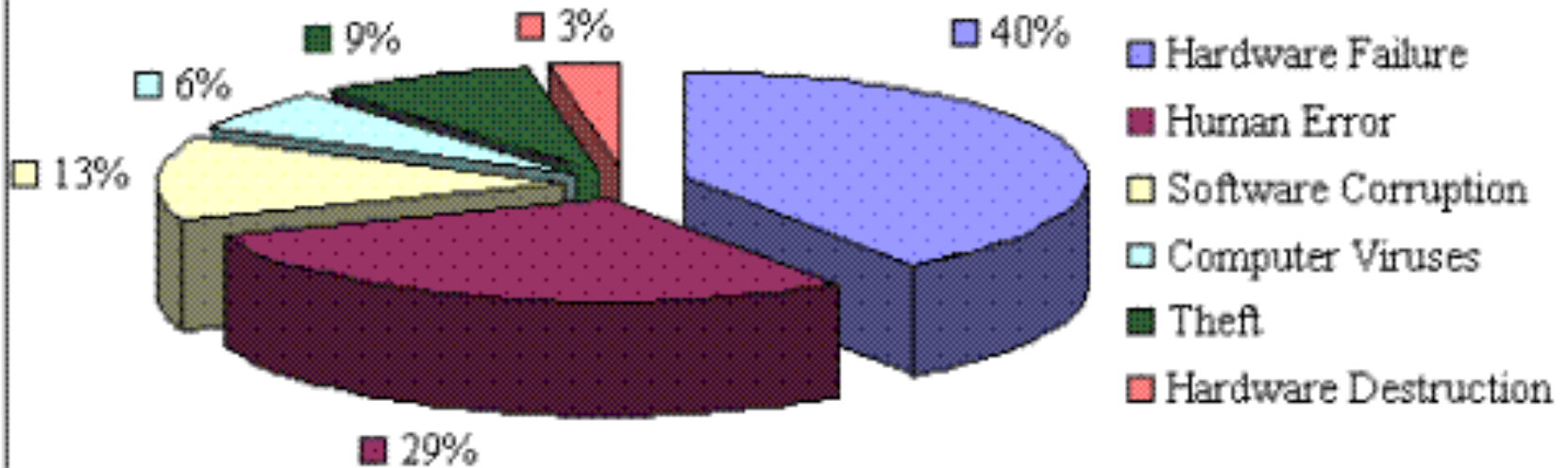
https://www.youtube.com/watch?v=8dhp_20j0Ys

Bit preservation challenges

- People do stupid things
- Failure to follow processes and procedures
- Technology is unreliable
- Things happen very fast in digital systems
- Live systems are not safe places for data
- Backups are not preservation



Causes of Data Loss



Examples of damaged discs.

Further reading: bit preservation and storage technologies

- Bit Preservation: A Solved Problem?
- Mean time to meaningless: MTTDL, Markov models, and storage system reliability
- Hard-Disk Drives: The Good, the Bad, and the Ugly



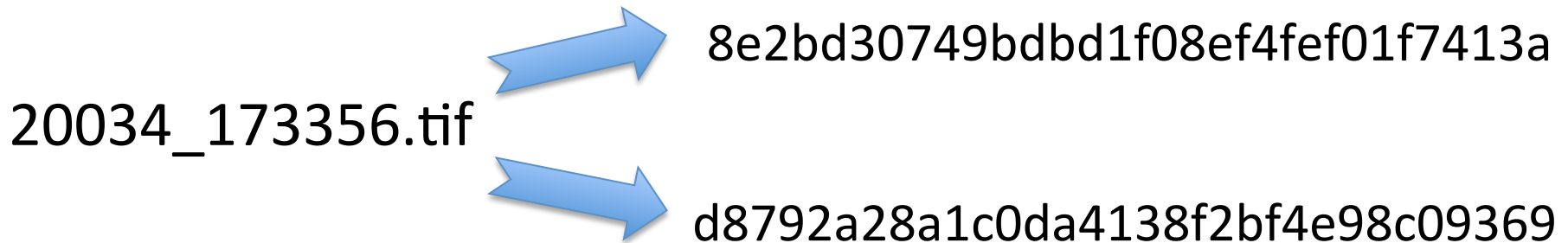
<http://cacm.acm.org/magazines/2009/6/28493-hard-disk-drives-the-good-the-bad-and-the-ugly/fulltext>

https://www.usenix.org/legacy/event/hotstorage10/tech/full_papers/Greenan.pdf

<http://www.ijdc.net/index.php/ijdc/article/view/151/224>

Fixity

- Fingerprints on data
- Detect even the smallest changes
- Know what you have
- Know if it's changed



How to approach the problem

99.99% chance of a car journey with no break down

600 trips per person per year in the UK

20,000 breakdowns per day

1,000,000 data objects

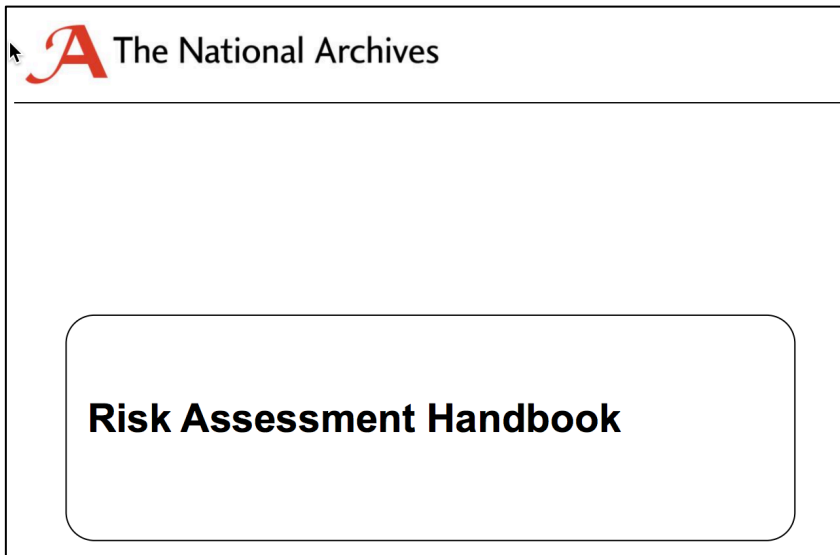
Each object makes 1,000 'data trips'

20 year total data journey



More information: risk assessment

DRAMBORA



<http://www.nationalarchives.gov.uk/documents/information-management/risk-assessment-handbook.pdf>
<http://www.repositoryaudit.eu/>
<http://public.ccsds.org/publications/archive/652x0m1.pdf>

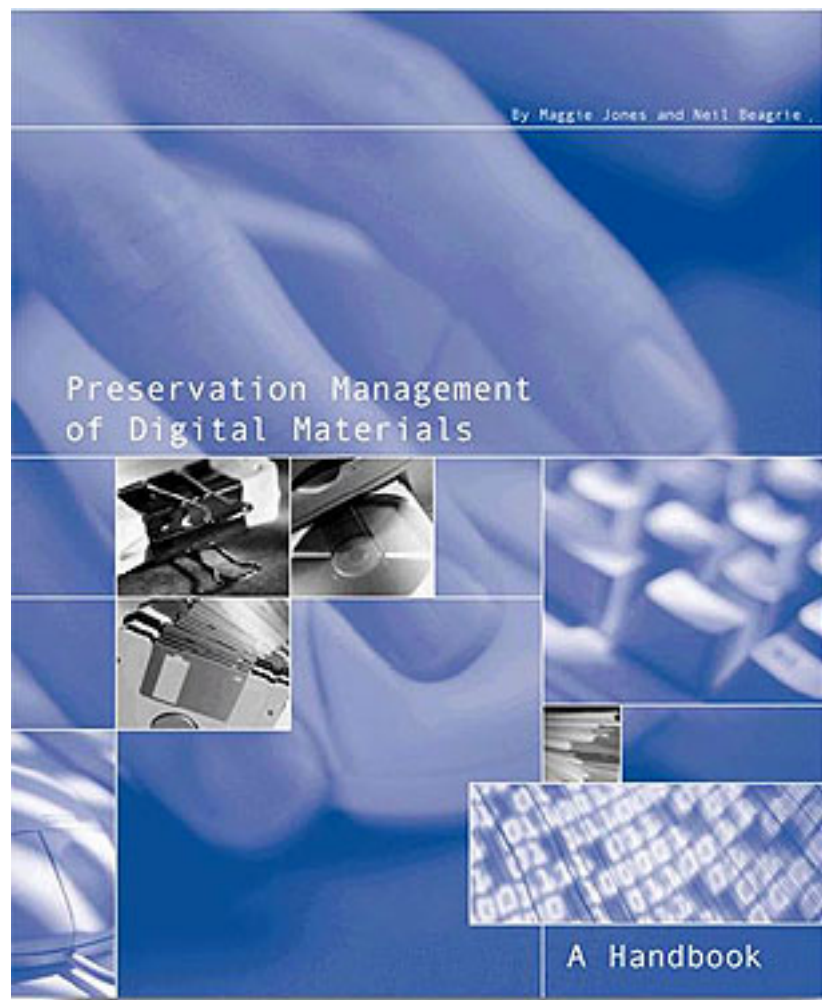
Recommended Practice (1)

- Redundancy and diversity (copies)
 - Lots of copies keep stuff safe
 - Online and offline
- Fixity (checksums)
 - Establish and monitor bit-level data integrity
 - Fix problems as soon as they occur
- Migration (obsolescence)
 - Monitor technologies/vendors
 - Migrate in a planned and managed way

Recommended Practice (2)

- Consolidate and document (KISS)
 - Trusted Digital Repository
 - Write everything down!
- Exit plan
 - Exports, escrow, standards
 - Avoid vendor or service provider lock-in

More Information: Digital Preservation Handbook



Emerging technical solutions

- Checksum tools (e.g. md5sum)
- Resilient file systems (e.g. ZFS)
- Archive storage (e.g. LTFS on LTO tape)
- Repository tools (e.g. DSpace, Fedora)
- AIP tools (e.g. Archivematica)
- Distributed preservation (e.g. LOCKSS, DPN)
- Digital Forensics tools (e.g. bitcurator)
- Cost and risk modelling (e.g. AVPreserve, 4C)
- Bit preservation services (e.g. Arkivum, DuraCloud)

Tool directory and comparison: COPTR and POWRR

<http://coptr.digipres.org/>

<http://www.digipres.org/tools/about/>

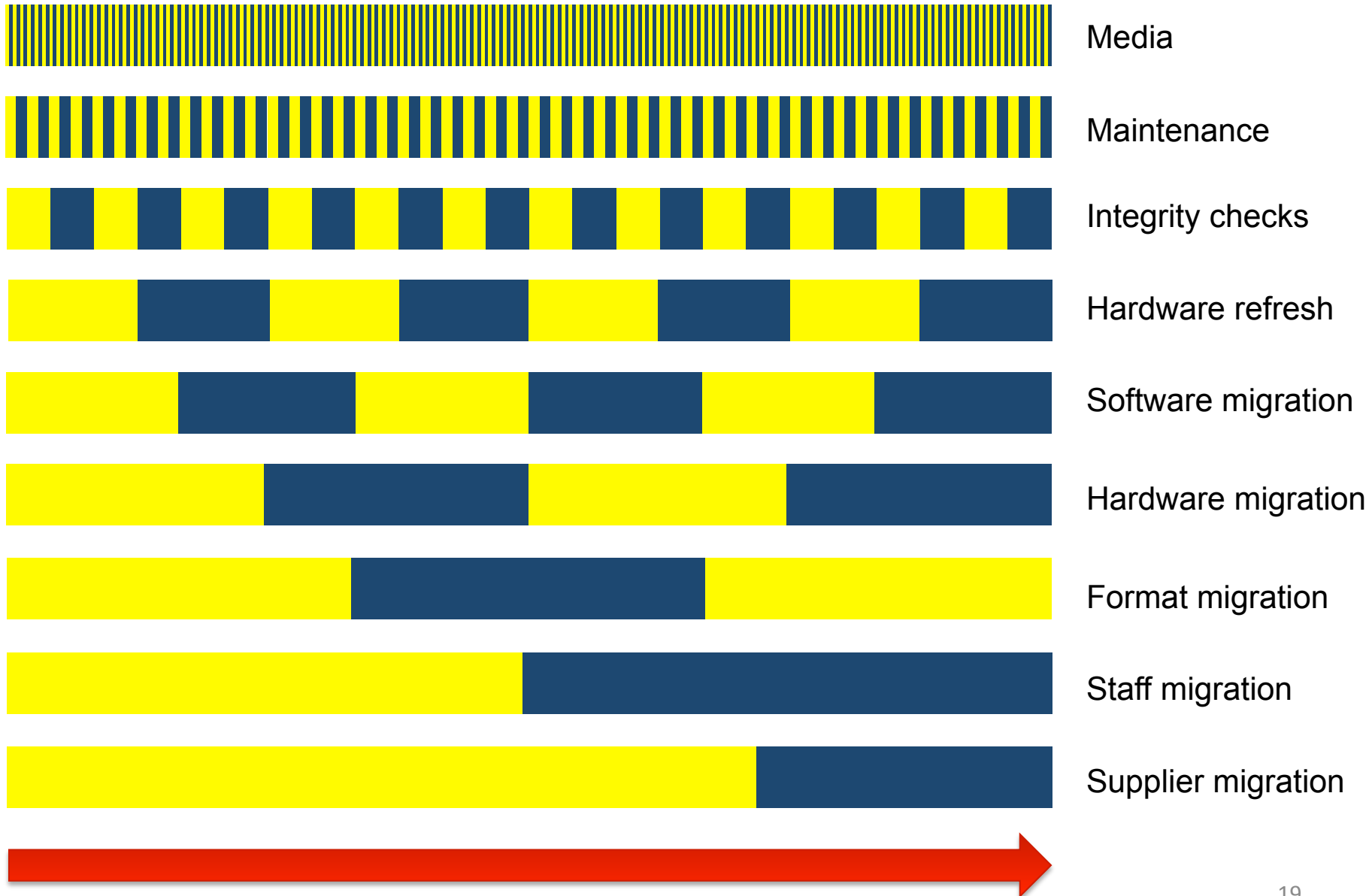
Lots of options, but no silver bullets

IT Storage	Cheap, quick access Short-lived, relatively unreliable Widely available, good adoption Needs monitoring and migrating
Long-lived media	Expensive, slow access Long-lived, relatively reliable Niche products, vendors go bust Players become obsolete
Cloud services	Moderate costs, online access Can be reliable, but need a solid contract Take advantage of skilled providers Don't put eggs all in one basket, need an exit plan

10 Megabyte Hard Disk \$3,495*



20 years of bit-preservation



Arkivum: bit-preservation as a service



Flagship Arkivum100 service with 100% data integrity guarantee



World-wide professional indemnity insurance – Arkivum100



Long term contracts for enterprise data archiving



Fully automated and managed solution

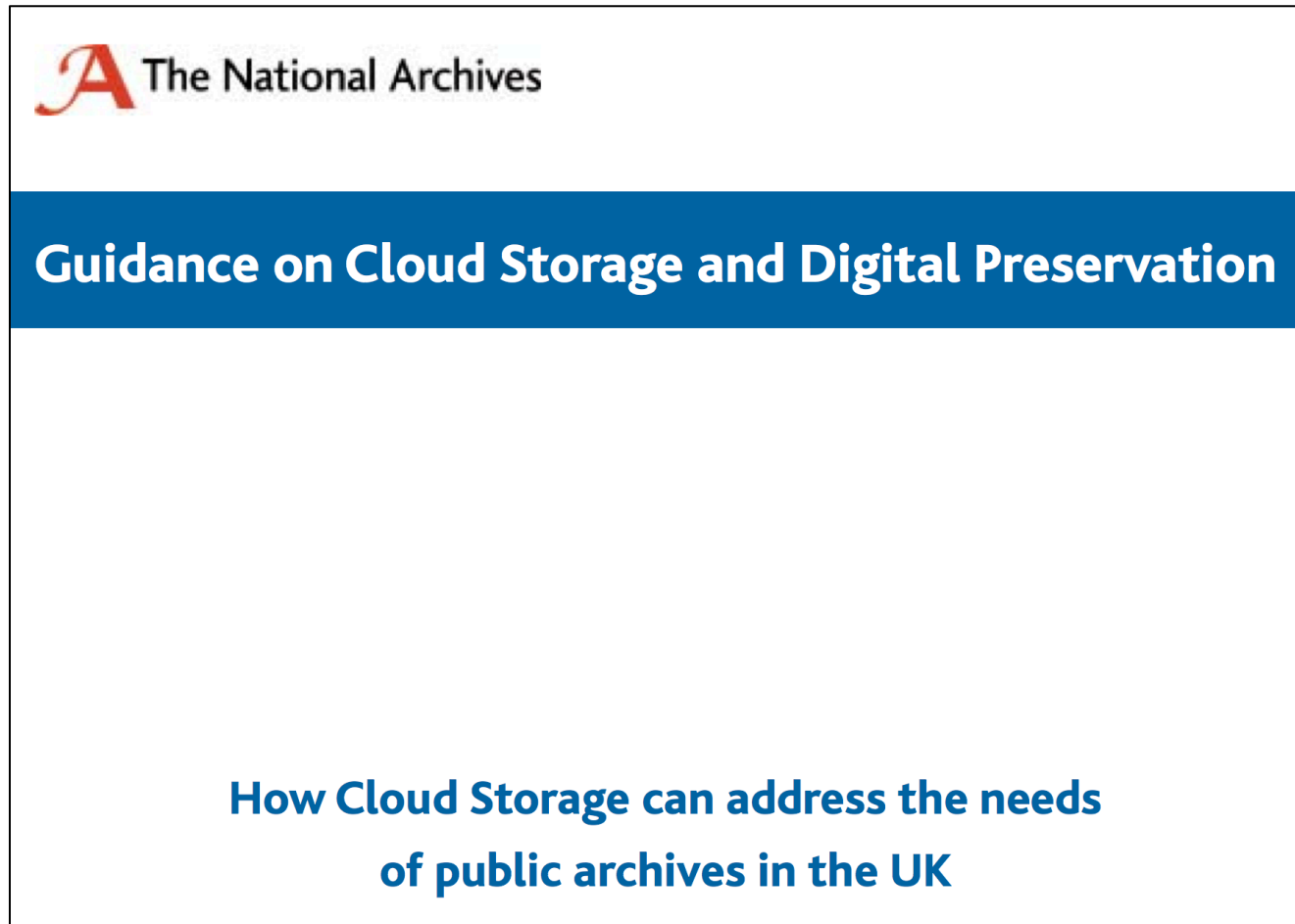


Audited and certified to ISO27001



Data escrow, exit plan, no lock-in

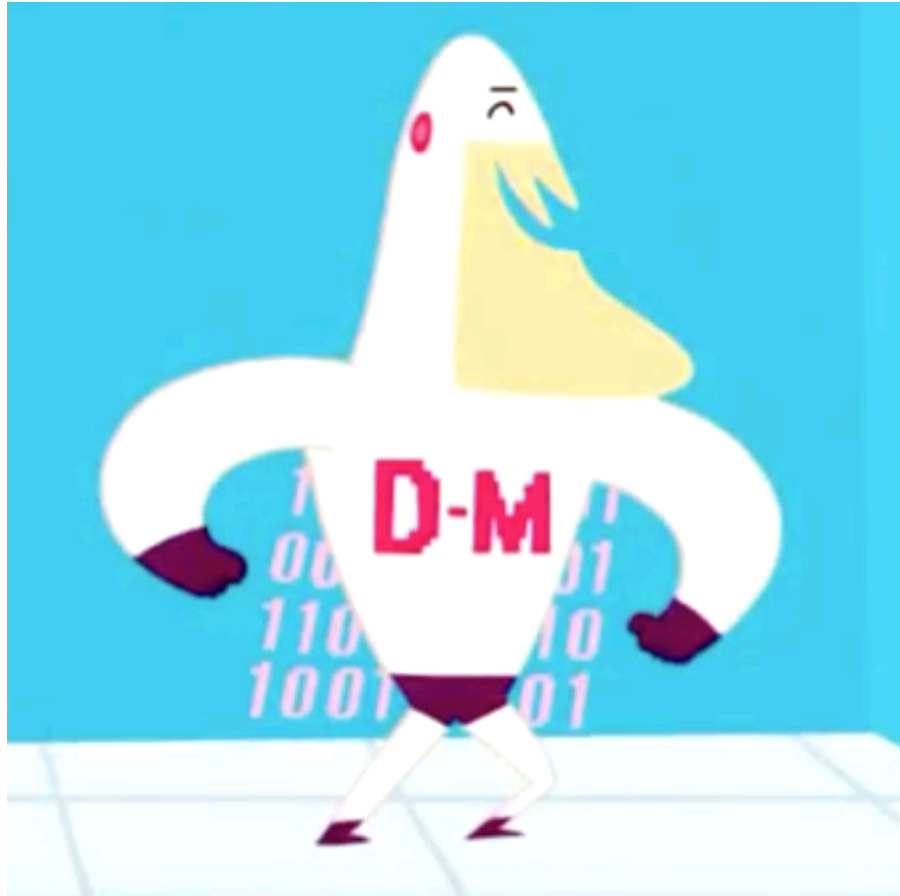
More Information: TNA guide on the cloud for preservation



Summary

- People, processes, skills, infrastructure
- Get data into a Trusted Digital Repository
- Plan for failures, corruptions, bugs, errors
- Establish and monitor fixity
- Redundancy and diversity
- Regular migrations
- Good metadata and documentation

Summary by DigiMan



<https://www.youtube.com/watch?v=pbBa6Oam7-w>

Questions?



matthew.addis@arkivum.com

www.arkivum.com