# Digital Preservation Environmental Scan

Neil Jefferies

Bodleian Libraries

University of Oxford

@NeilSJefferies

# Session Structure

- Hardware Trends
- Software Trends
- Cloud
- OCFL (Oxford Common File Layout)

# Hardware

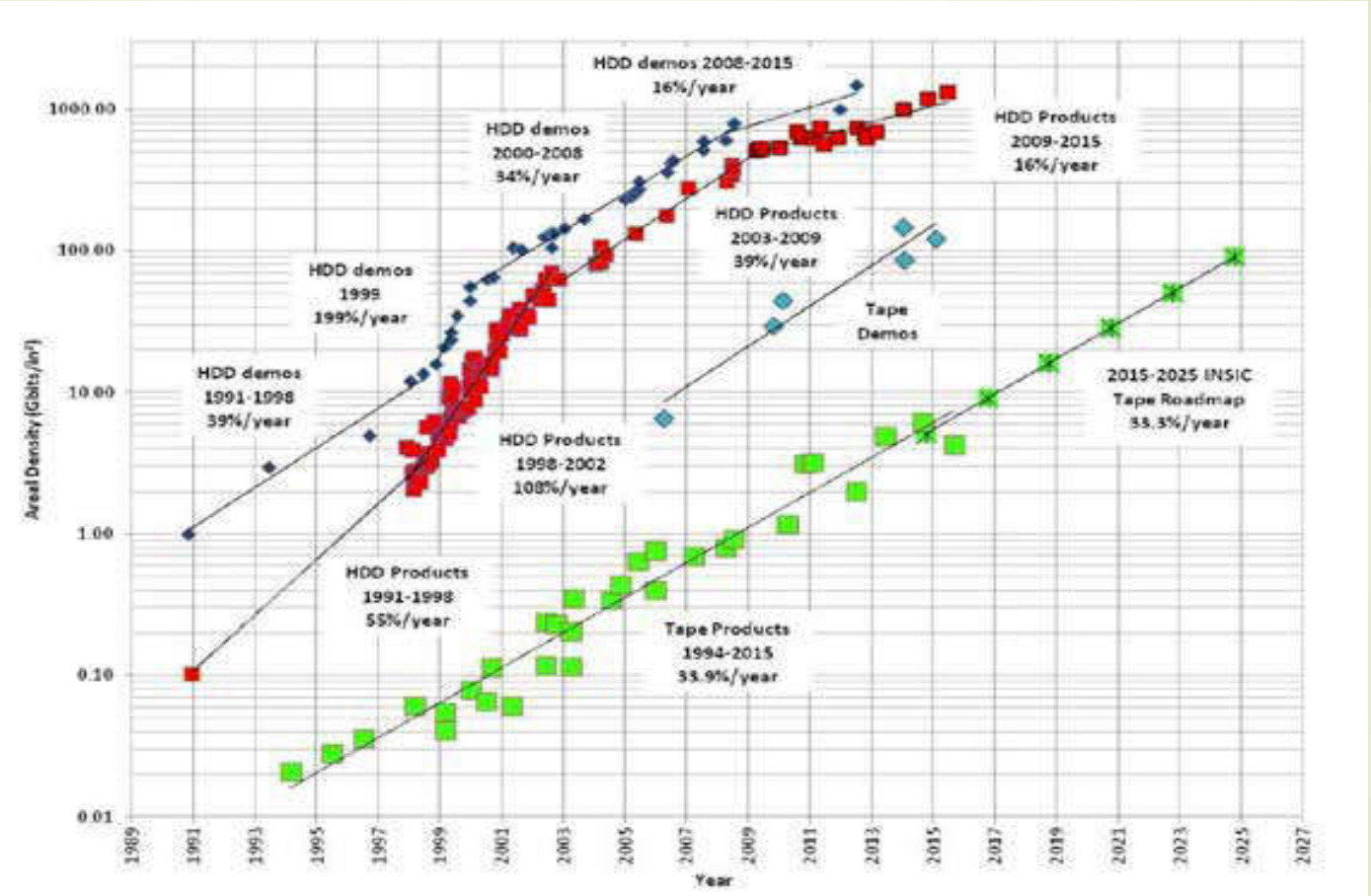
Keeping the bits safe

Bodleian Libraries
UNIVERSITY OF OXFORD

pasig
preserving and archiving special interest group

DigitalPreservationCoalition

# Trends: Disk vs Tape



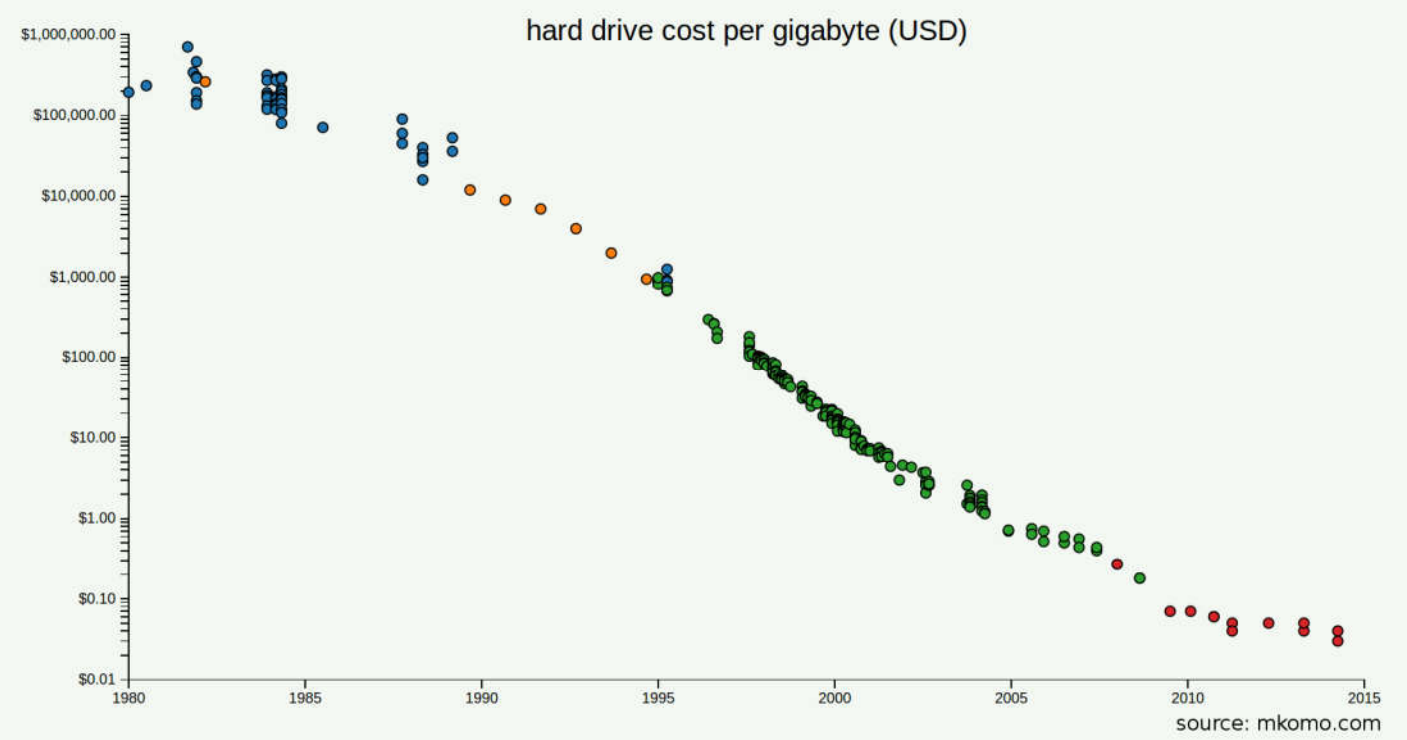



Bodleian Libraries
UNIVERSITY OF OXFORD

pasig
preserving and archiving special interest group

DigitalPreservationCoalition

# Some Independent Research




- Matt Komorowski http://www.mkomo.com/cost-per-gigabyte-update

Bodleian Libraries
UNIVERSITY OF OXFORD

pasig
preserving and archiving special interest group

DigitalPreservationCoalition

# Hard Disk

- Properties
  - Latency 5-15ms (this has not changed significantly for years)
  - Sustained data transfer rate 200MB/s
  - Capacity per unit (2018) 16TB
  - Cost per TB (2017) $50
  - Requires power (power cycling not recommended)
- Lifetime
  - 5 year warranties (MTBF figures are meaningless)
  - Interface longevity: SATA 2003, SAS 2004, FC (ANSI) 1994, Ethernet (802.3ab Gigabit) 1999
- Systemic Risk
  - 3 Manufacturers (Seagate, HGST and Toshiba)
  - Consumer market squeezed by PC substitutes (phones & tablets with flash)
  - Enterprise market squeezed by flash
  - Cloud enables higher utilisation by sharing -> lower unit shipments

# Hard Drive Technology

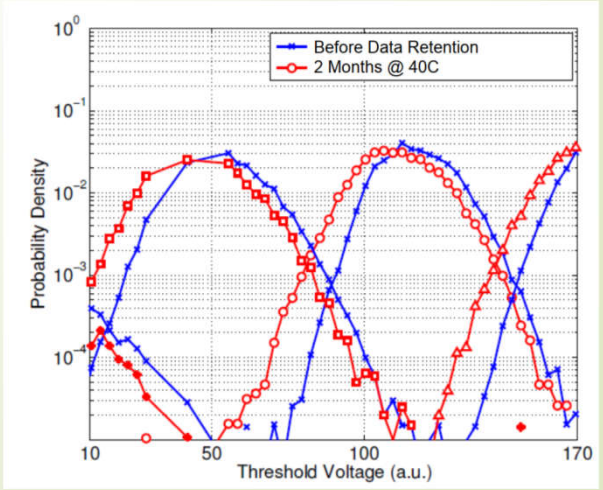
- Current technology limits being reached
- Short term fixes
  - He-filled drives allow more platters
  - Multiple banks of heads improve performance
  - Shingling
- Longer term face limitations of magnetic media
  - HAMR (Heat Assisted Magnetic Recording) Seagate
  - MAMR (Microwave Assisted…) WD, Toshiba
  - Patterned Media

# Tape


- Properties
  - Latency 100s (load from a robotic library), robot speeds gradually increasing
  - Sustained data transfer rate 360MB/s (faster than HDD once loaded!)
  - Capacity per unit (2017) 12TB
  - Cost per TB (2017) $21 (including library)
  - Media is unpowered, robot still needs power
- Lifetime
  - 2 Formats: IBM Magstar and LTO (Oracle T10K frozen in 2017)
  - 30 year media life (media warranty typically 1 year, though)
  - Drives typically can read back two generations (generations typically 2-3 years for LTO)
  - Drive warranties typically 5 years -> probably safe to keep media 10 years
    - IBM allows formatting older media at higher capacity (new with LTO-8, too)
- Systemic Risk
  - IBM: 1 drive manufacturer
  - LTO: 3 drive manufacturers (HPE, Quantum, IBM)

# Flash

- Properties
  - Latency 10us (decreasing rapidly)
  - Sustained data transfer rate 2000MB/s (generally limited by interface)
  - Capacity per unit (2017) 60TB, (2019) 100TB (higher density than disk)
  - Cost per TB (2017) $250 (decreasing rapidly)
  - Needs power! (but typically less than a hard drive)
- Lifetime
  - Enterprise SSD guaranteed retention 40 days (Consumer: 1 year, USB: indefinite)
    - In practice, retention is much longer
  - Warrantied according to total bytes written
    - Writing is primary degradation mechanism
  - Interface longevity similar to hard disks
- Low Systemic Risk
  - Many manufacturers (>10)

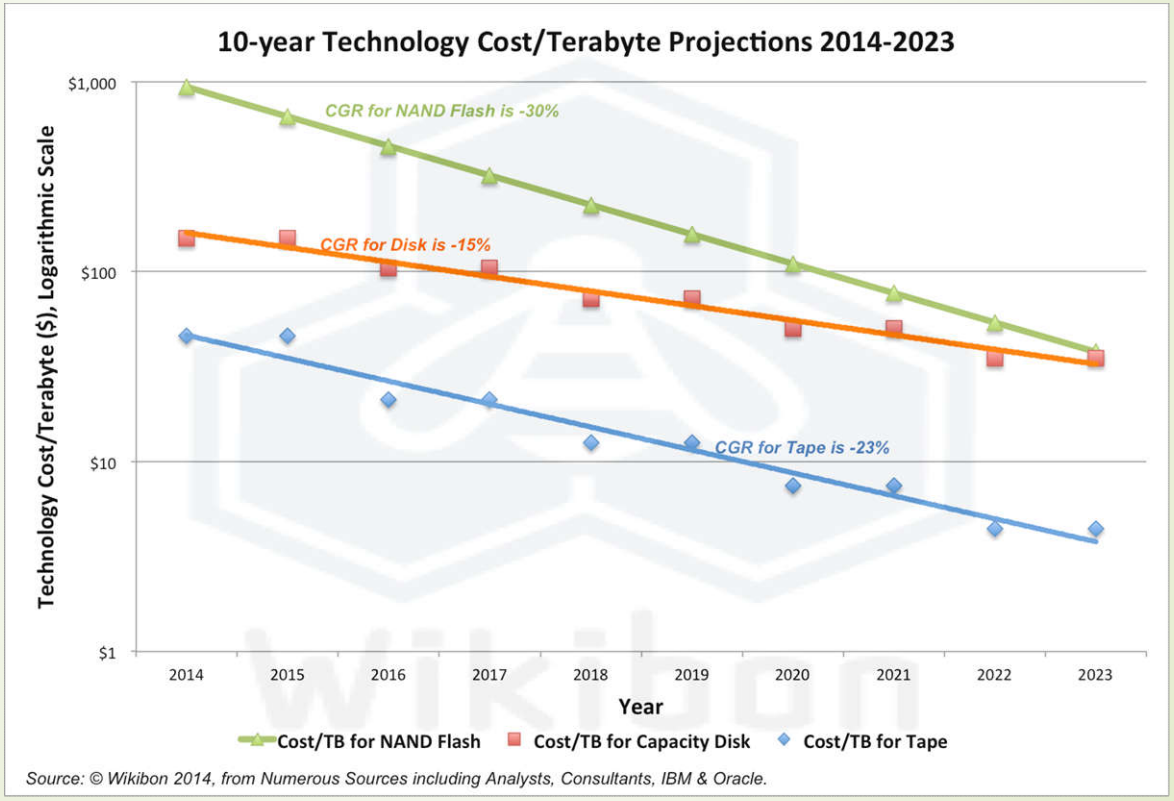



Bodleian Libraries
UNIVERSITY OF OXFORD

pasig
preserving and archiving special interest group

DigitalPreservationCoalition

# Long Term Total Cost of Ownership




Bodleian Libraries
UNIVERSITY OF OXFORD

pasig
preserving and archiving special interest group

DigitalPreservationCoalition

# Archive Optical


- Properties
  - Performance figures are scarce
  - Sustained data transfer rate 40MB/s (similar to Blu-Ray)
  - Capacity per unit (2017) 3.3TB (actually a cartridge of 9 disks)
  - Cost per TB (2017) $100
    - Enterprise SSD guaranteed retention 40 days (Consumer: 1 year, USB: indefinite)
    - In practice, retention is much longer
- Lifetime
  - Claimed 50 years+ for media
  - Drive promise backwards compatibility for all generations (only 2 exist so far)
- High Systemic Risk
  - Archive Optical: 2 Manufacturers (Panasonic, Sony)
    - Interoperability apparently not guaranteed
- M-Disc: 1 Manufacturer (Milleniata, has gone bankrupt once)
  - Variant of CD/DVD/Blue-Ray (LG, Lite-on, Asus produce compatible drives)
  - Has proved very robust in tests but low density (100GB)

Bodleian Libraries
UNIVERSITY OF OXFORD

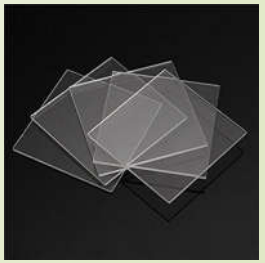
pasig
preserving and archiving special interest group

DigitalPreservationCoalition

# New technologies (briefly)

- Phase Change SSD's (e.g. X-Point)
  - More robust, potentially faster than flash
  - Costly, and less dense
  - Robustness not a selling point for many (it appears)
  - X-Point collaborators (Intel and Micron) have dissolved their JV

- DNA
  - Very robust (through replication) and high capacity
  - Read/write devices exist in many labs (though not for that purpose!) Quite large and cumbersome.
  - Bandwidth not that good at the moment



- Fused silica
  - Very robust (phase change material) and high capacity
  - Writing requires a specialised laser, reading much simpler

# Software

Making the bits useful/usable

Bodleian Libraries
UNIVERSITY OF OXFORD

pasig
preserving and archiving special interest group

DigitalPreservationCoalition

# Distributed orgnisations

- As data is distributed, so the organisations and processes follow
  - Geographic distribution is easier with partners
  - Technological distribution too
- Many preservation tools are open source
  - Operations are verifiable and repeatable
  - Need community
- Data can survive organisational failure
- Beware of lock in
  - Always have an exit strategy or rather "somewhere to go"
- Introduce additional complexities
  - Contractual
  - Governance
  - Rights/access and control

# Dissemination

- Beware the "dissemination copy"
  - It is the copy that people will reference, cite and care about
  - It will need to be regularly cross-checked with the archived material
  - It will need to be preserved
  - …so, ideally, generate it on-the-fly from an archival copy and cache it
- Emulation
  - Some formats just cannot be easily migrated or displayed
    - E.g. Macromedia Shockwave, FLASH, Multimedia titles
    - Security concerns with some formats too
  - Possible to emulate most hardware using modern software
    - Able to run older operating systems and software securely
    - "If it can play games then an emulator has almost certainly been written"
    - Most emulators are open source – easy to obtain
    - Long term support is harder – opportunity for DP community
- Discovery
  - Frequently neglected part of re-use
  - Depends on good metdata
  - Incremental curation – expect to add/update metadata over time

# Cloud-based Preservation

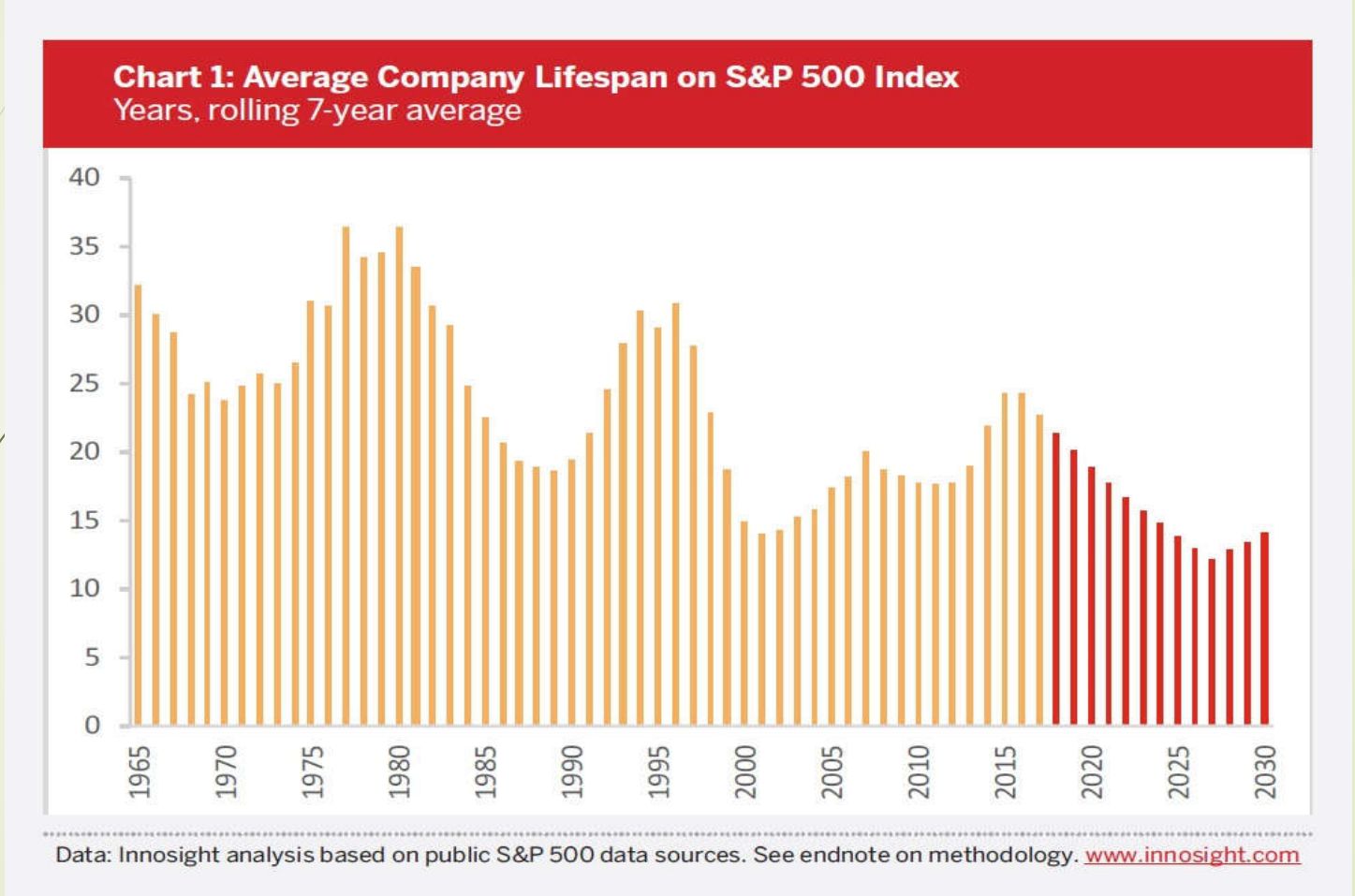
Is nothing new – it is the same hardware and software but with one important new risk factor…

Bodleian Libraries
UNIVERSITY OF OXFORD

pasig
preserving and archiving special interest group

DigitalPreservationCoalition

# Expect to Migrate!




Bodleian Libraries
UNIVERSITY OF OXFORD

pasig
preserving and archiving special interest group

DigitalPreservationCoalition

# OCFL

Oxford Common File Layout

https://ocfl.io

Bodleian Libraries
UNIVERSITY OF OXFORD

pasig
preserving and archiving special interest group

DigitalPreservationCoalition

# What is it

The Oxford Common File Layout (OCFL) specification describes an **application-independent** approach to the storage of digital information in a structured, transparent, and predictable manner. It is designed to promote long-term object management best practices within digital repositories.

**Observations**

- Archived objects change relatively slowly than archival software.
- Filesystems (and in particular POSIX filesystems) have been the most consistently implemented and widely tested API's for accessing storage in any form.
- Migration by export/ingest is slow and risky
- MOAB, BagIT, RDF DataBank as antecedents…

# Key attributes

- **Complete.** All the information about a digital object in an OCFL compliant repository should be serialised in the OCFL.
- **Application Independent.** Consequently, a repository should be rebuildable from just the data in an OCFL. Even if it is not the source repository.
- **Human parsable.** An OCFL should be understandable to a person (with a little effort). With basic filesystem tools they should be able to identify digital objects and their versions and contents.
- **Portable.** OCFL requires a minimal set of filesystem capabilities so it can be implemented on most filesystems, and be portable between them.
- **Provenance and Versioning**. OCFL allows the capture of a version history for objects and provide for the implementation of an audit trail.
- **Deduplication.** OCFL allows for deduplication of content between object versions so that unchanged parts
- **Fixity.** OCFL provides fixity as a by-product of its use of content-based addressing.
- **Burn-Line.** OCFL permits the loss of all systems except for a basic file-system storage node and still provide full recoverability.

# Who?

- Andrew Hankinson (Bodleian Libraries, University of Oxford)
- Neil Jefferies (Bodleian Libraries, University of Oxford)
- Rosalyn Metz (Emory University)
- Julian Morley (Stanford University)
- Simeon Warner (Cornell University)
- Andrew Woods (DuraSpace)

- https://groups.google.com/forum/#!forum/ocfl-community

- "Oxford" because of a Samvera meeting at Oxford…

Bodleian Libraries
UNIVERSITY OF OXFORD

pasig
preserving and archiving special interest group

DigitalPreservationCoalition

# What does it look like?

```
[storage_root]
├── 0=ocfl_1.0
├── ocfl_1.0.txt (optional OCFL spec)
├── d45
│   └── be6
│       └── 26e
│           ├──d45be626e024
│           │   ├── 0=ocfl_object_1.0
│           │   └── ...
│           └──d45be626e036
│               ├── 0=ocfl_object_1.0
│               └── ...
├── 310
│   └── 4ed
│       └── f03
│           └── 3104edf0363a
│               ├── 0=ocfl_object_1.0
│               └── ...
└── ...
```

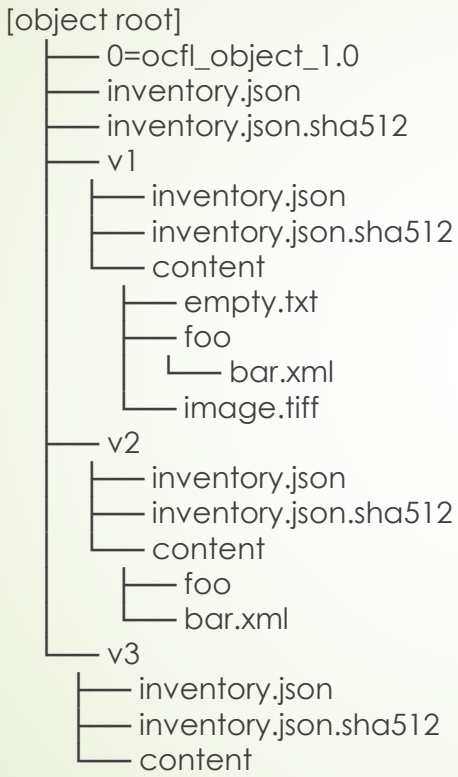
```
[object root]
├── 0=ocfl_object_1.0
├── inventory.json
├── inventory.json.sha512
├── v1
│   ├── inventory.json
│   ├── inventory.json.sha512
│   └── content
│       └── myfirstbag
│           ├── bagit.txt
│           ├── data
│           │   └── 27613-h
│           │       └── images
│           │           ├── q172.png
│           │           └── q172.txt
│           └── manifest-md5.txt
└── v2
    ├── inventory.json
    ├── inventory.json.sha512
    └── content
        └── myfirstbag
            ├── data
            │   └── 27614-h
            │       └── images
            │           ├── q173.png
            │           └── q173.txt
            └── manifest-md5.txt
```

Bodleian Libraries
UNIVERSITY OF OXFORD

pasig
preserving and archiving special interest group

DigitalPreservationCoalition

# Inventory





```
{
 "digestAlgorithm": "sha512",
 "head": "v3",
 "id": "ark:/12345/bcd987",
 "manifest": {
  "4d27c8...b53": [ "v2/content/foo/bar.xml" ],
  "7dcc35...c31": [ "v1/content/foo/bar.xml" ],
  "cf83e1...a3e": [ "v1/content/empty.txt" ],
  "ffccf6...62e": [ "v1/content/image.tiff" ]
 },
 "type": "Object",
 "versions": {
  "v1": {
   "created": "2018-01-01T01:01:01Z",
   "message": "Initial import",
   "state": {
    "7dcc35...c31": [ "foo/bar.xml" ],
    "cf83e1...a3e": [ "empty.txt" ],
    "ffccf6...62e": [ "image.tiff" ]
   },
   "type": "Version",
   "user": {
    "address": "alice@example.com",
    "name": "Alice"
   }
  },
  "v2": {
   "created": "2018-02-02T02:02:02Z",
   "message": "Fix bar.xml, remove image.tiff, add empty2.txt",
   "state": {
    "4d27c8...b53": [ "foo/bar.xml" ],
    "cf83e1...a3e": [ "empty.txt", "empty2.txt" ]
   },
   "type": "Version",
   "user": {
    "address": "bob@example.com",
    "name": "Bob"
   }
  },
  "v3": {
   "created": "2018-03-03T03:03:03Z",
   "message": "Reinstate image.tiff, delete empty.txt",
   "state": {
    "4d27c8...b53": [ "foo/bar.xml" ],
    "cf83e1...a3e": [ "empty2.txt" ],
    "ffccf6...62e": [ "image.tiff" ]
   },
   "type": "Version",
   "user": {
    "address": "cecilia@example.com",
    "name": "Cecilia" }}}}
```


Bodleian Libraries
UNIVERSITY OF OXFORD

pasig
preserving and archiving special interest group

DigitalPreservationCoalition

# Thank you

neil.jefferies@bodleian.ox.ac.uk

Bodleian Libraries
UNIVERSITY OF OXFORD

pasig
preserving and archiving special interest group

DigitalPreservationCoalition

# Sources

- Spectralogic Data Storage Outlook Reports
  - https://spectralogic.com/resources/white-papers/
  - Actually several good papers there!
- The Register
  - https://www.theregister.co.uk/
  - Good industry news, market figures (and somewhat British humour)
- Matt Komorowski
  - http://www.mkomo.com/cost-per-gigabyte-update
- ServetheHome
  - https://www.servethehome.com
  - In depth storage and server reviews
- Backblaze
  - https://www.backblaze.com/blog/
  - Cloud-based backup provider that publishes stats on their considerable hard drive estate

# Distributed Systems

- Storage controller hardware is getting simpler
  - Fixity and error correction now handled on-device
  - Redundancy, replication and caching handled by software
- Redundancy (rather than replication)
  - Geographic distribution (power failure etc.)
  - Technology distribution (disk **and** tape, different manufacturers etc.)
  - RAID – Redundant Array of Inexpensive disks
    - MAID – Massive Array of Idle Disks, powers disks down to sqve power
    - With large numbers of drives: time to recover > time to next disk failure
  - Erasure coding/clustering (Ceph, IPFS, ScoutFS, ZFS…)
    - Specify a number of fragments and how many needed to recover
    - Faster rebuild times
    - Tunable for fault tolerance/costs balance