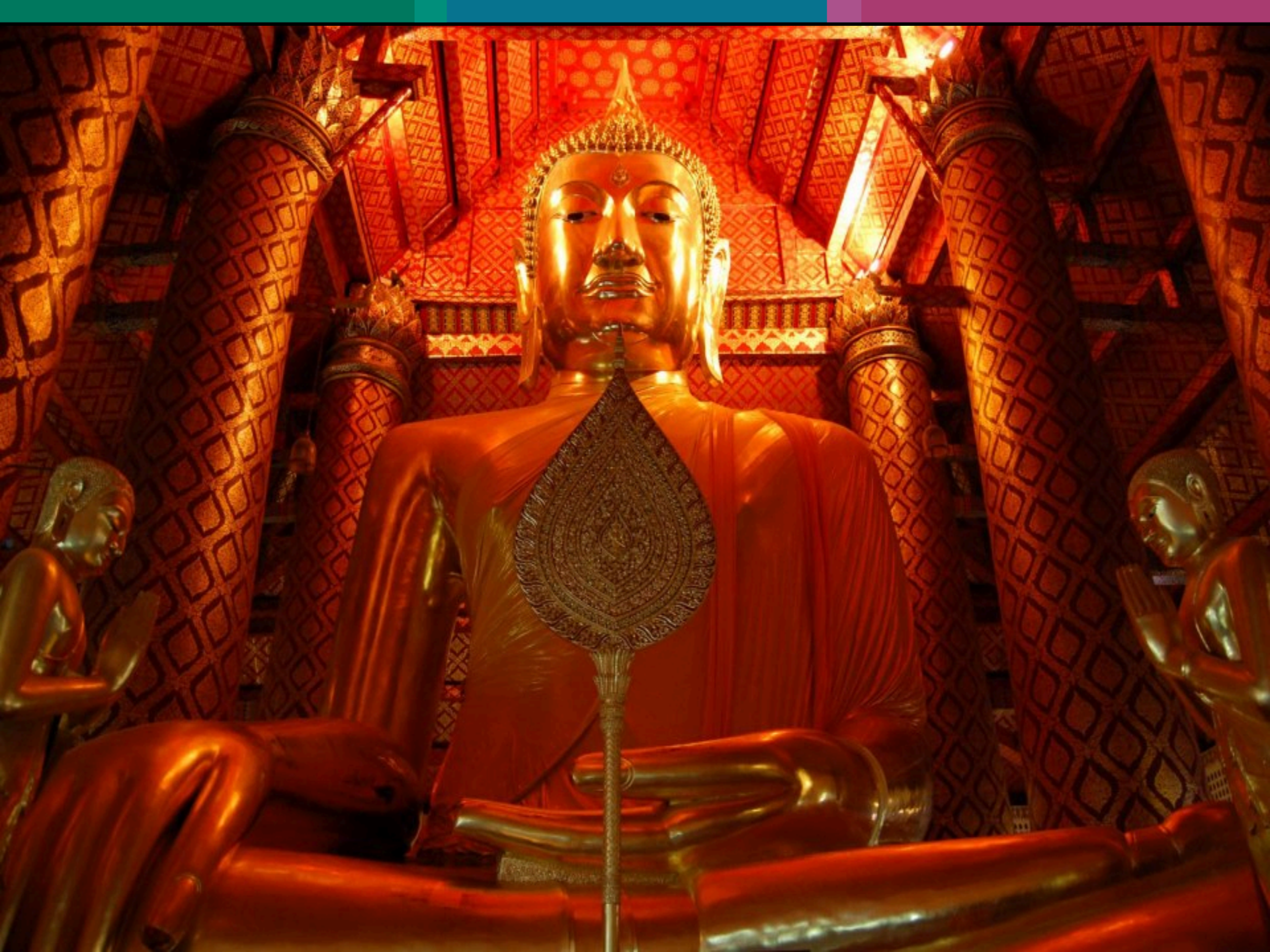


OPEN COMMUNITY APPROACHES TO DIGITAL PRESERVATION

ED FAY
OPEN PRESERVATION FOUNDATION






“ Everything is impermanent.

Source: The Buddha



“ Everything is impermanent.
Attachment causes suffering.

Source: The Buddha



“ Everything is impermanent.
Attachment causes suffering.

Source: The Buddha

Change is also the cause of
suffering in the digital world.



Hard Drives
SCSI / IDE



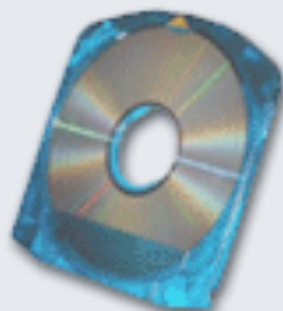
Raid Servers



Desktop PCs,
Laptops and Macs



CD / DVD



Digital Devices



Memory Sticks



Floppy Diskette



Smart Media



Tape Media





Carrier

Lifetime (years)

Stone tablet

10,000

Paper

1,000

Film

100

Disk

10

Source: Richard Wright (<http://dx.doi.org/10.7207/twr12-01>)

Carrier

Stone tablet

Paper

Film

Disk

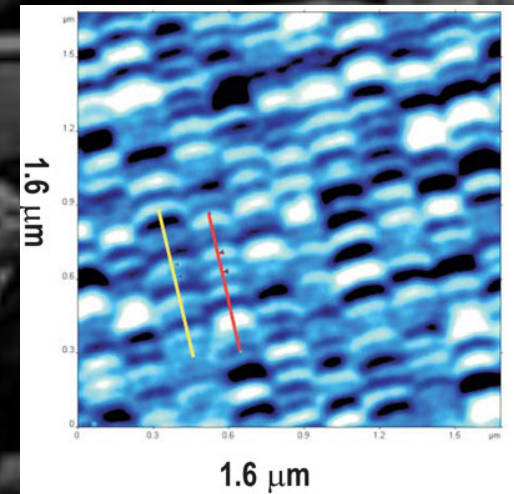
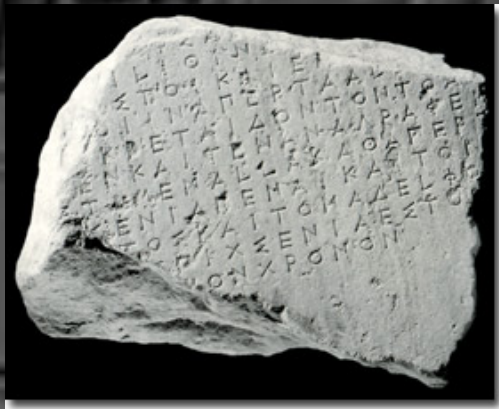
Lifetime (years)

10,000

1,000

100

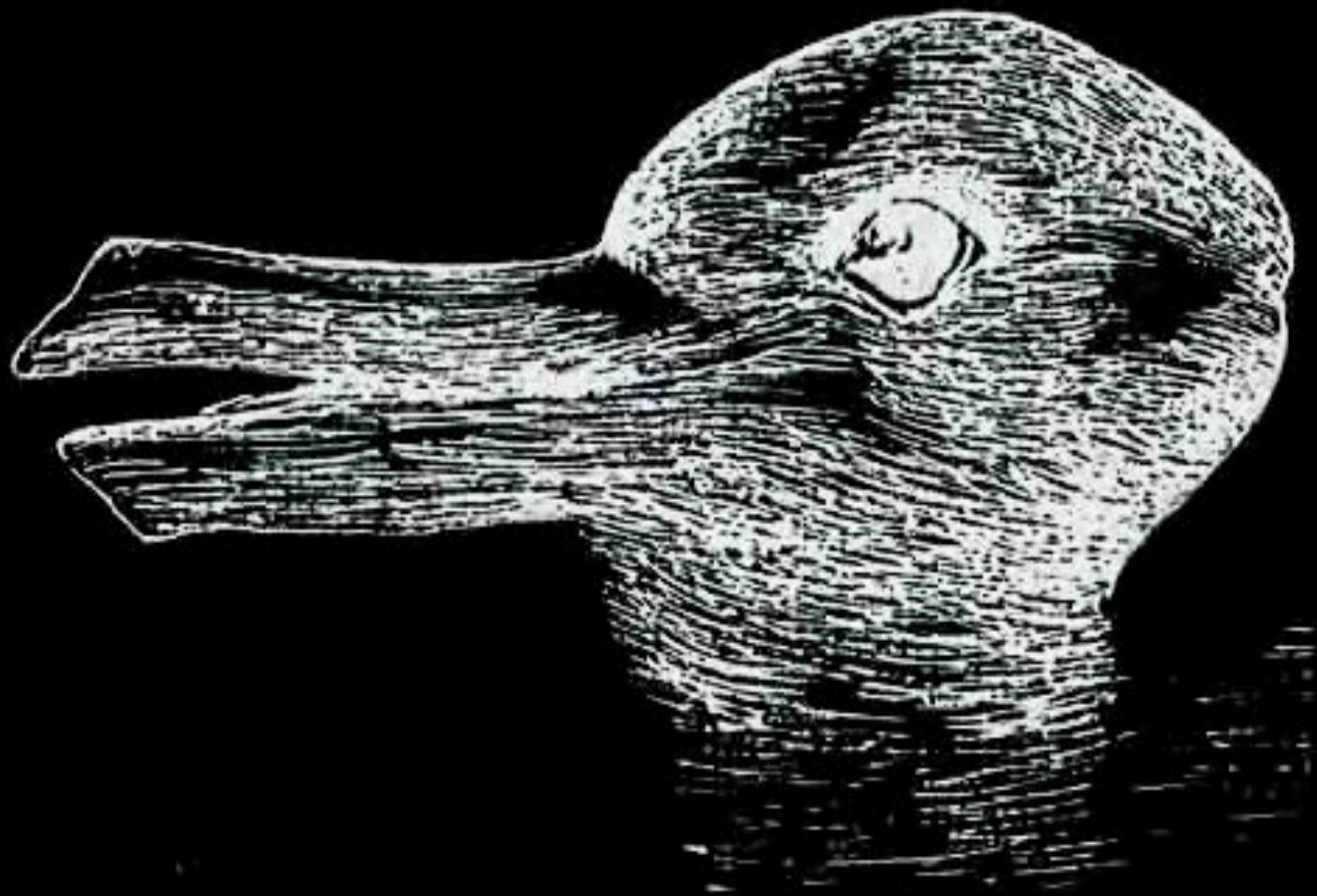
10

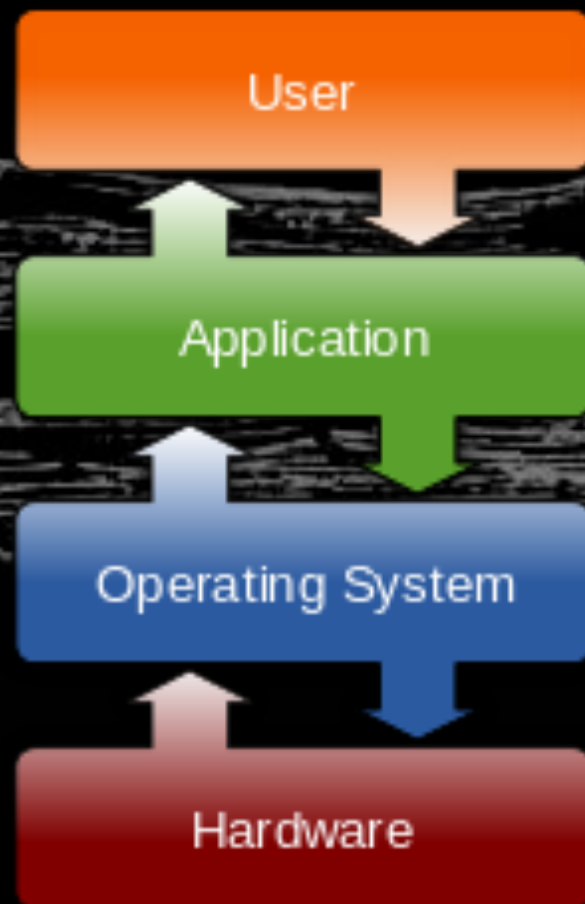


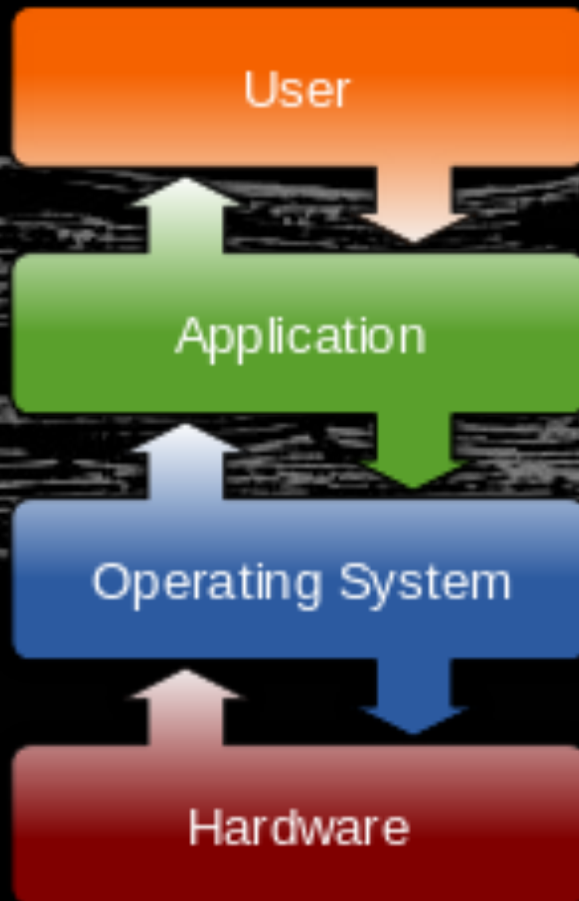
Source: Richard Wright (<http://dx.doi.org/10.7207/twr12-01>)







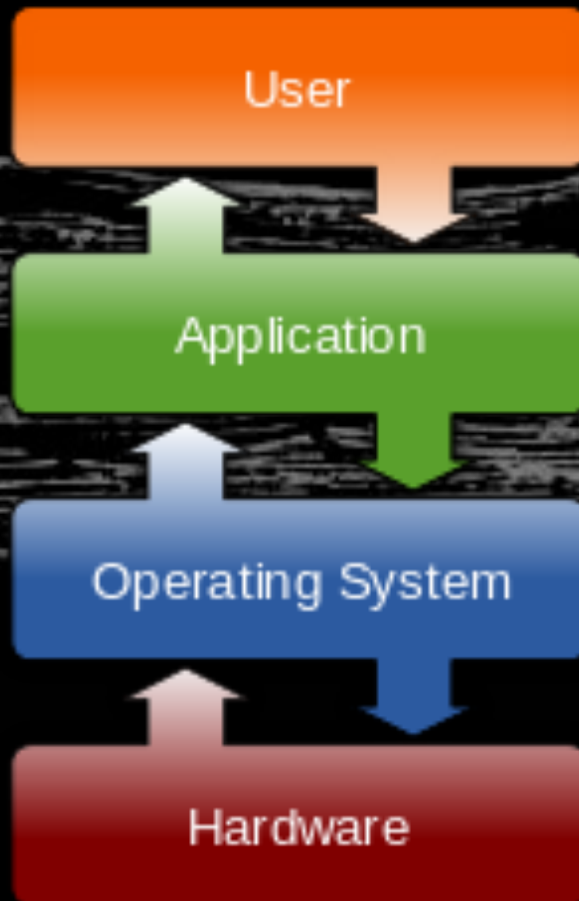




File system

Structures data for retrieval
Provides data as files

Hardware interaction

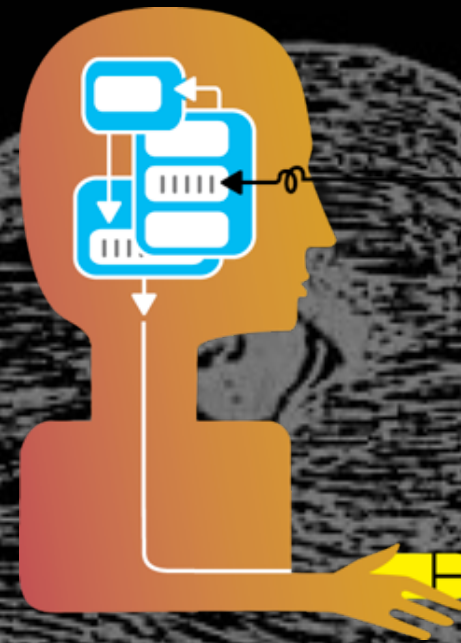
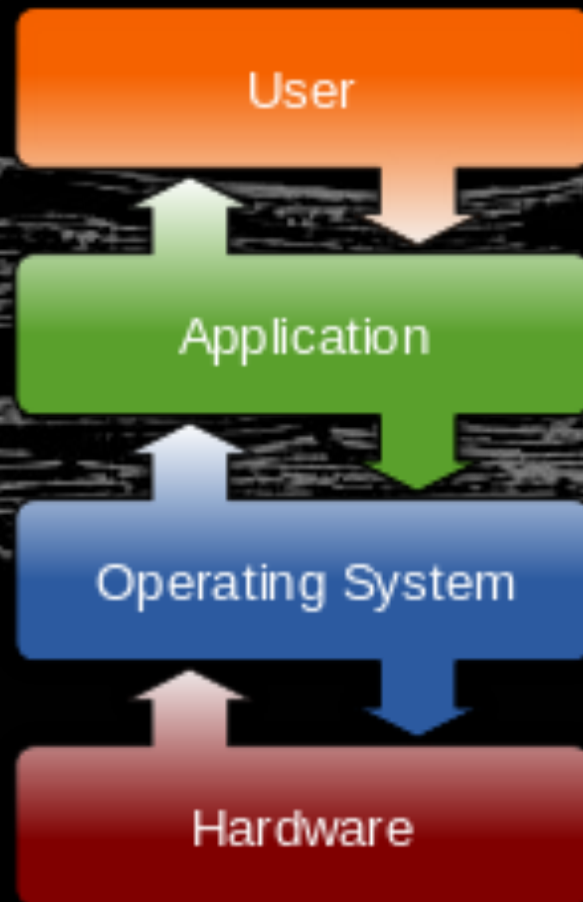


Software

Provides interactions with files
Off-the-shelf or bespoke

File formats

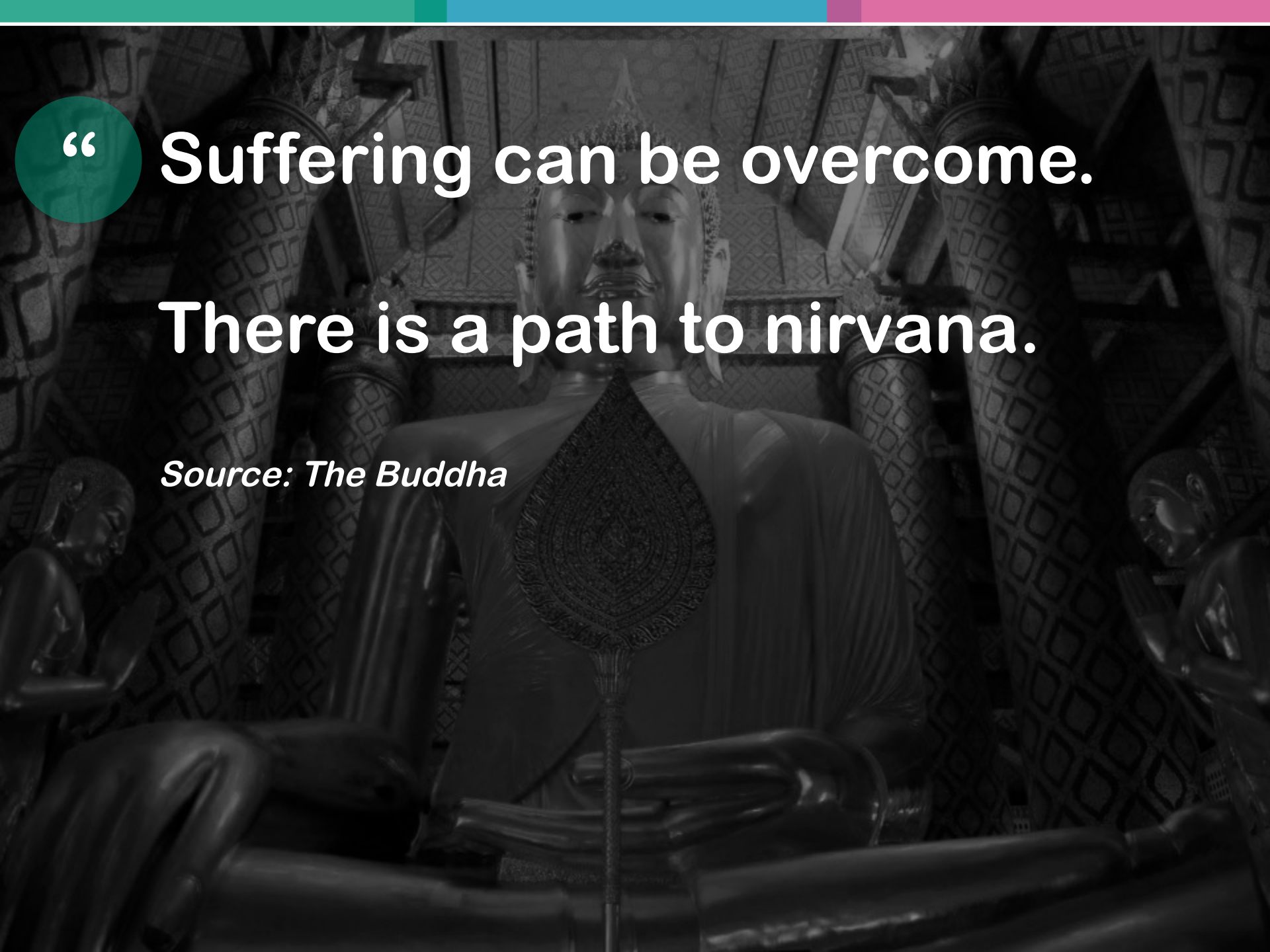
Encode data in meaningful ways
Proprietary or open standards






“ Suffering can be overcome.

Source: The Buddha

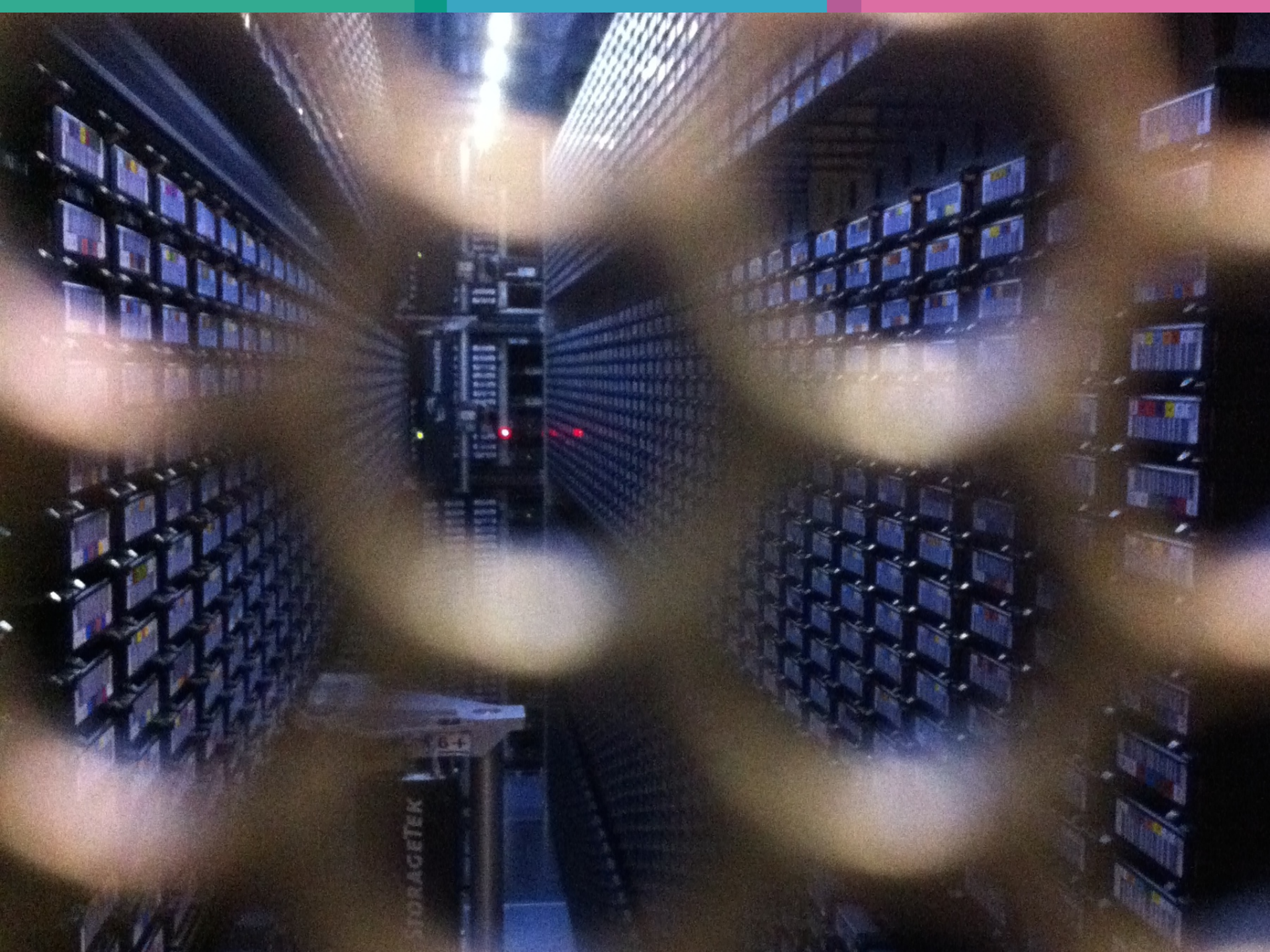


“ Suffering can be overcome.
There is a path to nirvana.

Source: The Buddha



Digital preservation is best understood as the active management of digital information to ensure its continued accessibility over time.





Physical preservation

Is keeping the bits safe



Physical preservation

Is keeping the bits safe

Replication

Multiple copies

Multiple storage media

Multiple locations

Fixity

Ensure identity over time (using 'checksums')

Prevent natural, accidental, or malicious alteration



Logical preservation Is understanding dependencies



Logical preservation

Is understanding dependencies

Characterise collections

Identify formats
Identify features

Characterise environments

Technical environments
Social environments

Logical preservation

Is understanding dependencies

“ Metadata is a love note
to the future

Source: Jason Scott (@textfiles)



Logical preservation

Is mitigating risk

Logical preservation

Is mitigating risk

Open standards and software

Conformance to standards and control of features

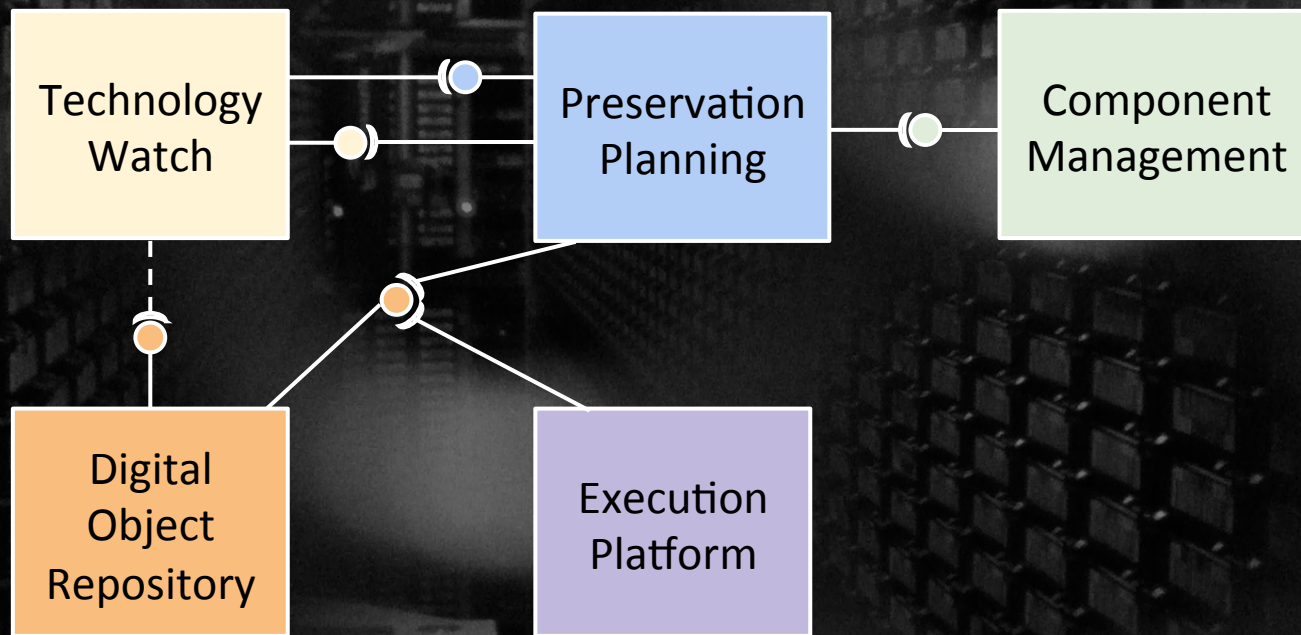
Transparent processing and quality assurance

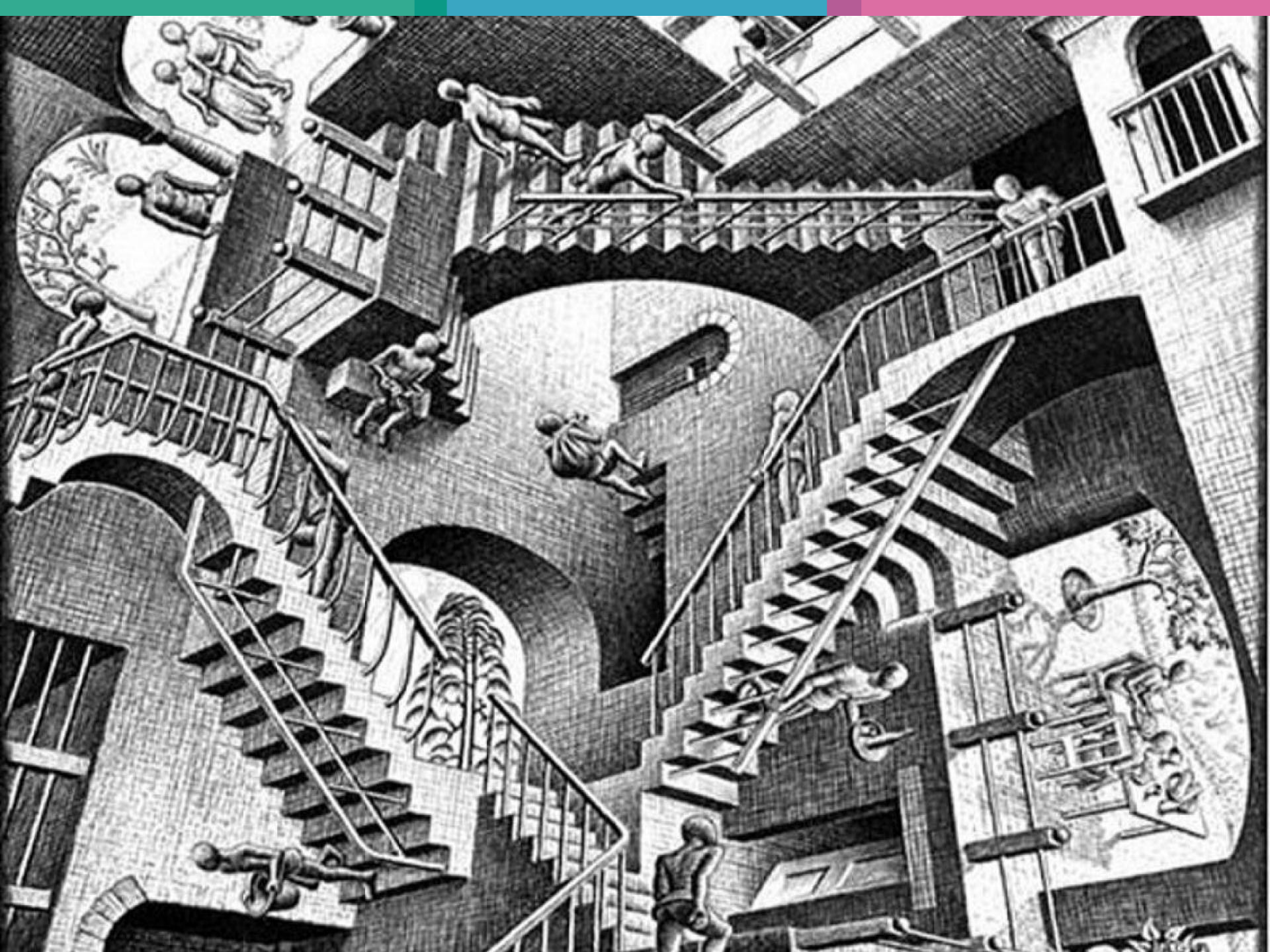
Intervention (as a last resort)

Migration of file formats

Emulation of legacy environments

Digital preservation

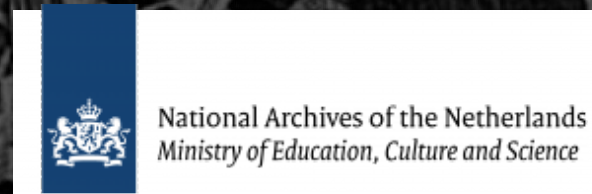
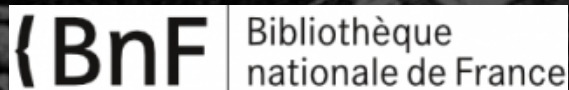




Open
Preservation
Foundation



LIBRARY
HSILIRB



Open
Preservation
Foundation

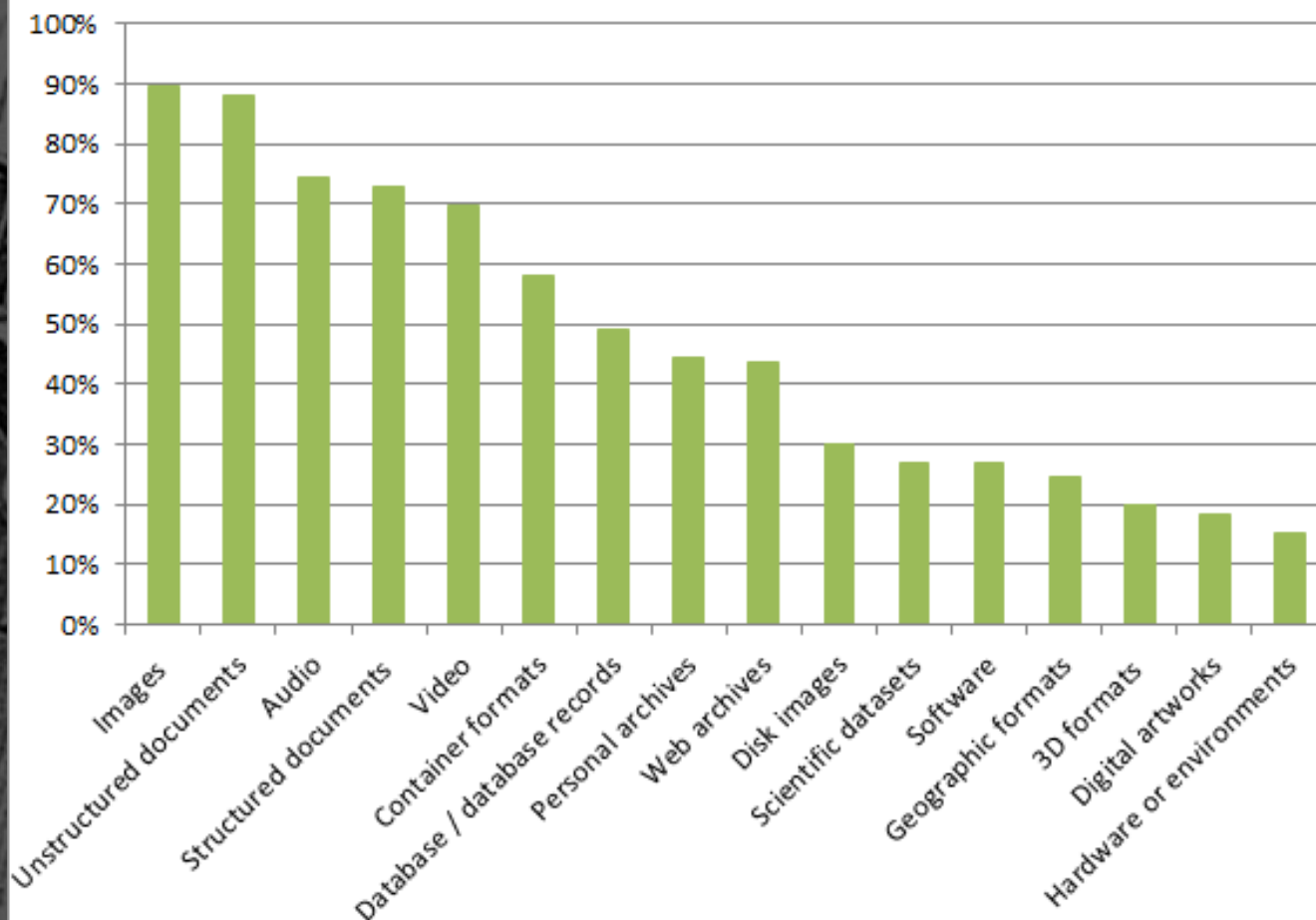


Technology

Knowledge

Advocacy and Alliances





Source: <http://openpreservation.org/knowledge/surveys/>

Technology

Open
Preservation
Foundation



Open source software

Maturity model (based on Apache, ISO25000)

Products (Production), Incubator (Labs), End-of-life (Attic)

Long-term sustainability

Consistency (hosting and packaging) = easy to use

Consolidation (evaluation and consensus) = easy to discover

Effectiveness (testing and roadmapping) = easy to adopt

Efficiency (collaborative development) = easy to maintain

<http://openpreservation.org/technology/>

Technology

OPF-Production

FIDO (format identification)

JHOVE (format validation)

JHOVE2 (format validation)

Jpylyzer (JP2K validation)

Matchbox (image analysis)

xCorrSound (audio analysis)

[hosted as a service]

PLATO (preservation planning)

SCOUT (preservation watch)

OPF-Labs

Bitwiser (bit analysis)

Flint (format policy)

Hardware Extractors (TIMBUS)

HaWarp (web archive processing)

Pagelyzer (web archive QA)

Simulator (scalability)

Strender (rendering analysis)

[and dozens more...]

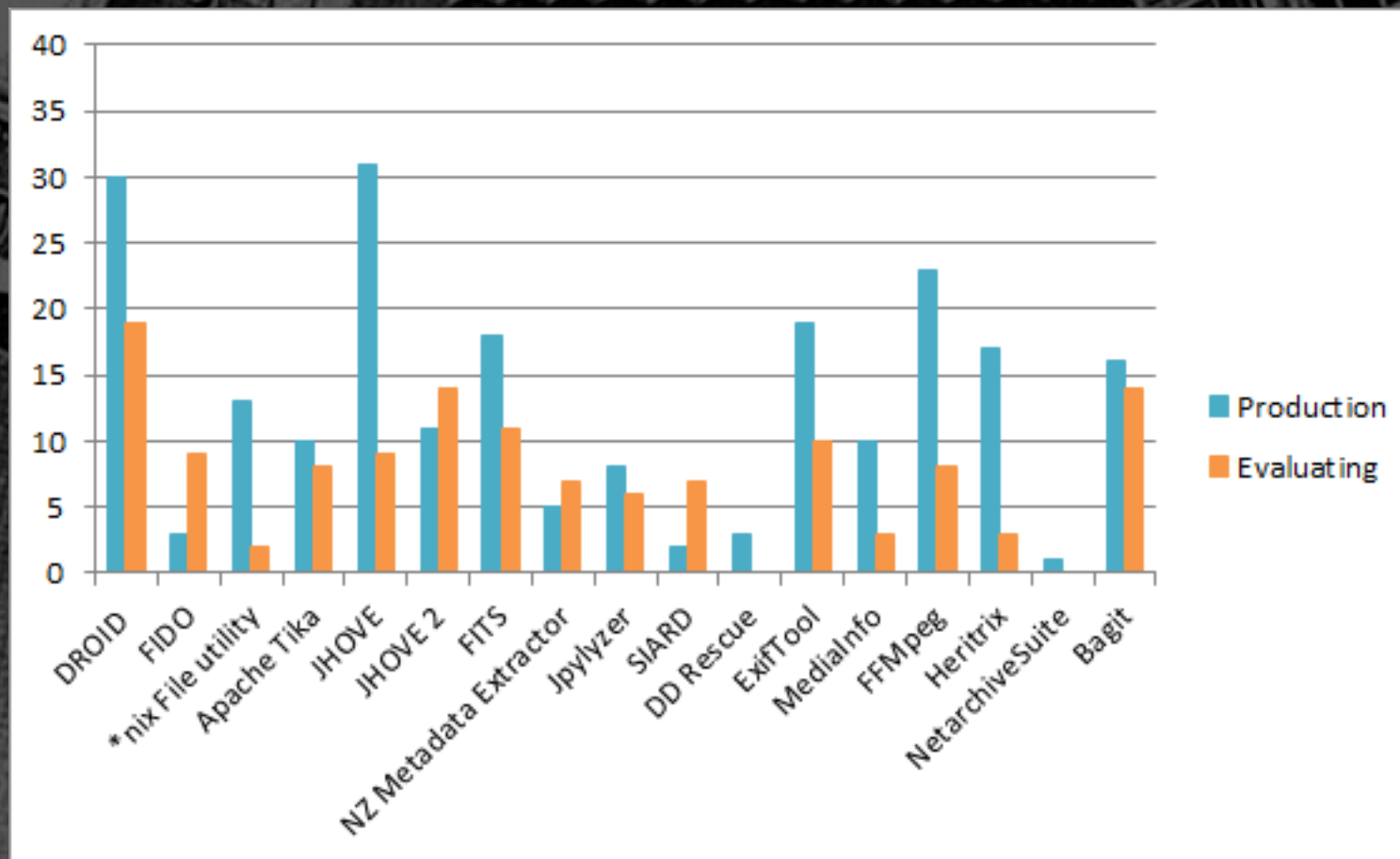
[contributed to]

BitCurator (digital forensics)

FITS (format characterisation)

c3po (collection profiling)

<http://openpreservation.org/technology/>



Source: <http://openpreservation.org/knowledge/surveys/>

Technology

Atlas of Digital Damages

- examples of preservation failures, e.g. corrupt files

Catalogue of Policy Elements

- guidance, procedures, control

Datasets, Issues, Solutions

- collection context, user requirements, solutions

File Format Risk Registry

- format-specific risks, recommendations and tools

Tools Registry (COPTR)

- reference for >350 digital preservation tools

Q&A site

- question and answer for digital preservation issues

<http://openpreservation.org/technology/>

Knowledge

Interest groups

Sharing experience (case studies, best practice)

Events

Expert workshops (webinars), collaboration (hackathons)

Training

Skills development

Surveys

<http://openpreservation.org/knowledge/>

Advocacy and Alliances

Projects

Open
Preservation
Foundation



<http://openpreservation.org/about/>



Open
Preservation
Foundation

Advocacy and Alliances

Collaborations

veraPDF: PDF/A validation

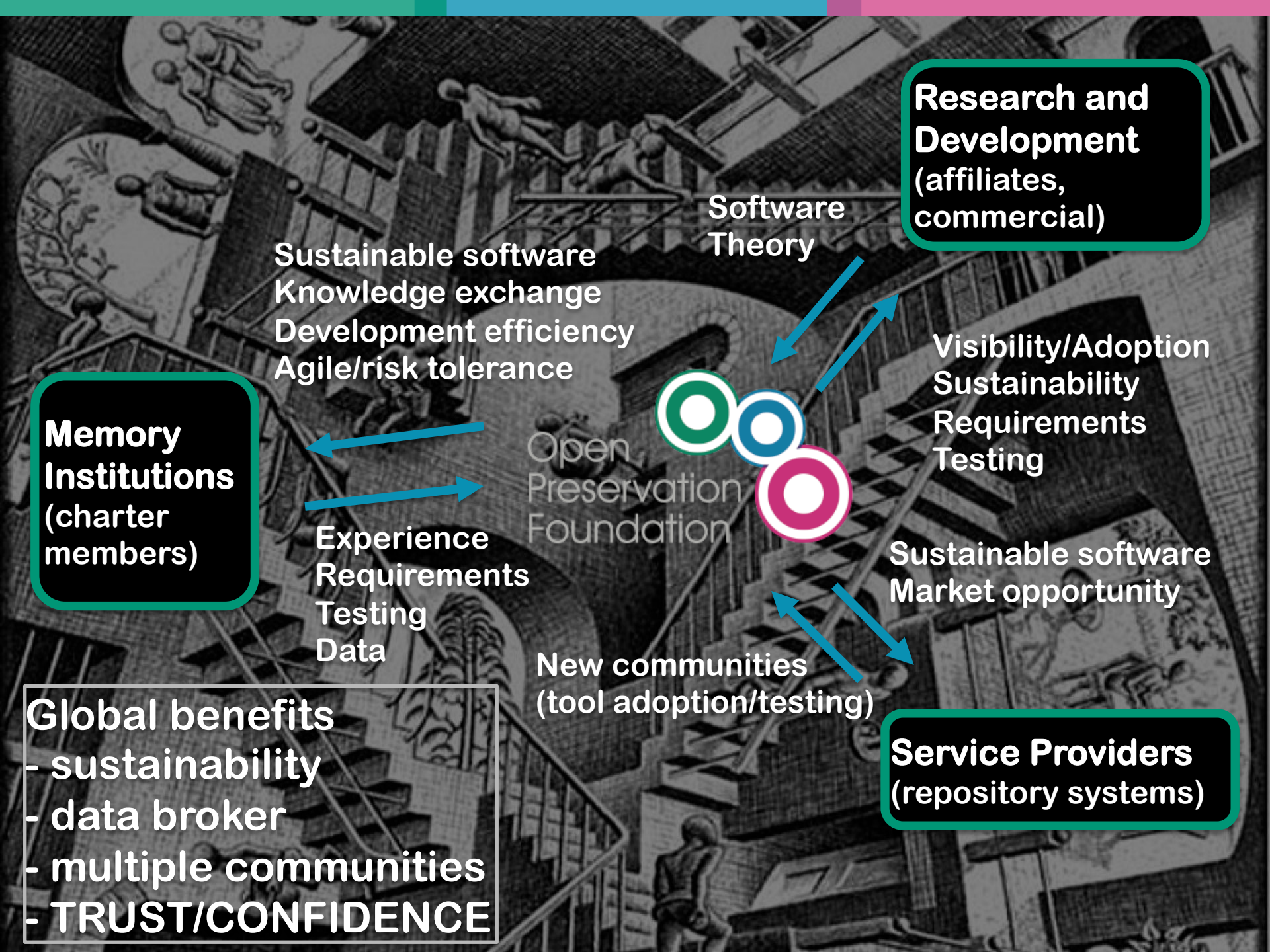
- objective frame of reference
- ISO committee liaison
- Policy Profile Registry



Digitisation MOOC



<http://openpreservation.org/about/>



Research and Development
(affiliates, commercial)

Software Theory

Sustainable software
Knowledge exchange
Development efficiency
Agile/risk tolerance

Memory Institutions
(charter members)

Experience
Requirements
Testing
Data

Open Preservation Foundation

Visibility/Adoption
Sustainability
Requirements
Testing

Sustainable software
Market opportunity

New communities
(tool adoption/testing)

Service Providers
(repository systems)

Global benefits

- sustainability
- data broker
- multiple communities
- TRUST/CONFIDENCE





! *Final thoughts*

The system is not the solution.

IT must understand
preservation as much as
librarians must understand IT.

Diversify by collaboration.



! *Final thoughts*

He who controls the present,
controls the past.

He who controls the past,
controls the future.

Source: George Orwell, 1984

ed@openpreservation.org

<http://openpreservation.org/>