



# Digital Preservation Technology Futures

Matthew Addis - Arkivum

25 June 2017





# Digital Preservation Technology ... Where Next?



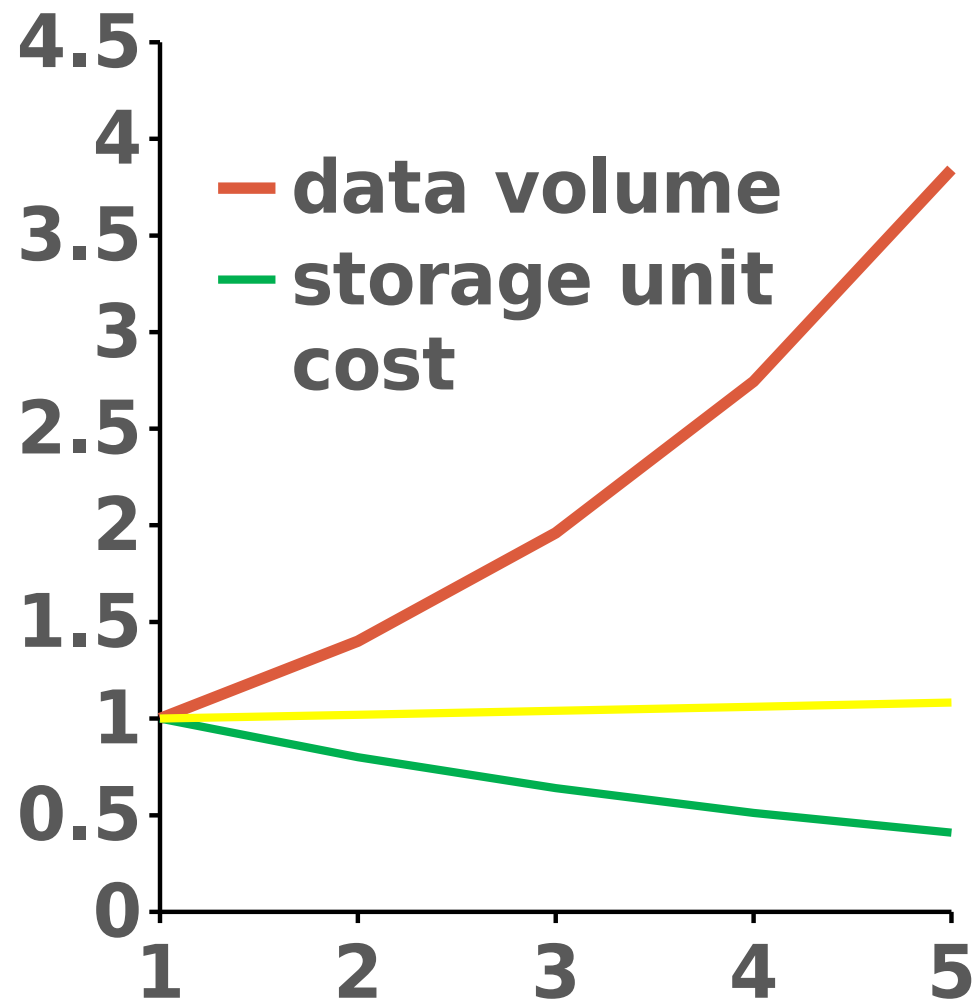


# 1. Volume

## 2018 *This Is What Happens In An Internet Minute*







- Keep everything is not an answer
- Falling storage costs is not enough



- Automation
- Appraisal
- Selection
- Disposition





# 2. Variety

## CONVERSATION PRISM 5.0

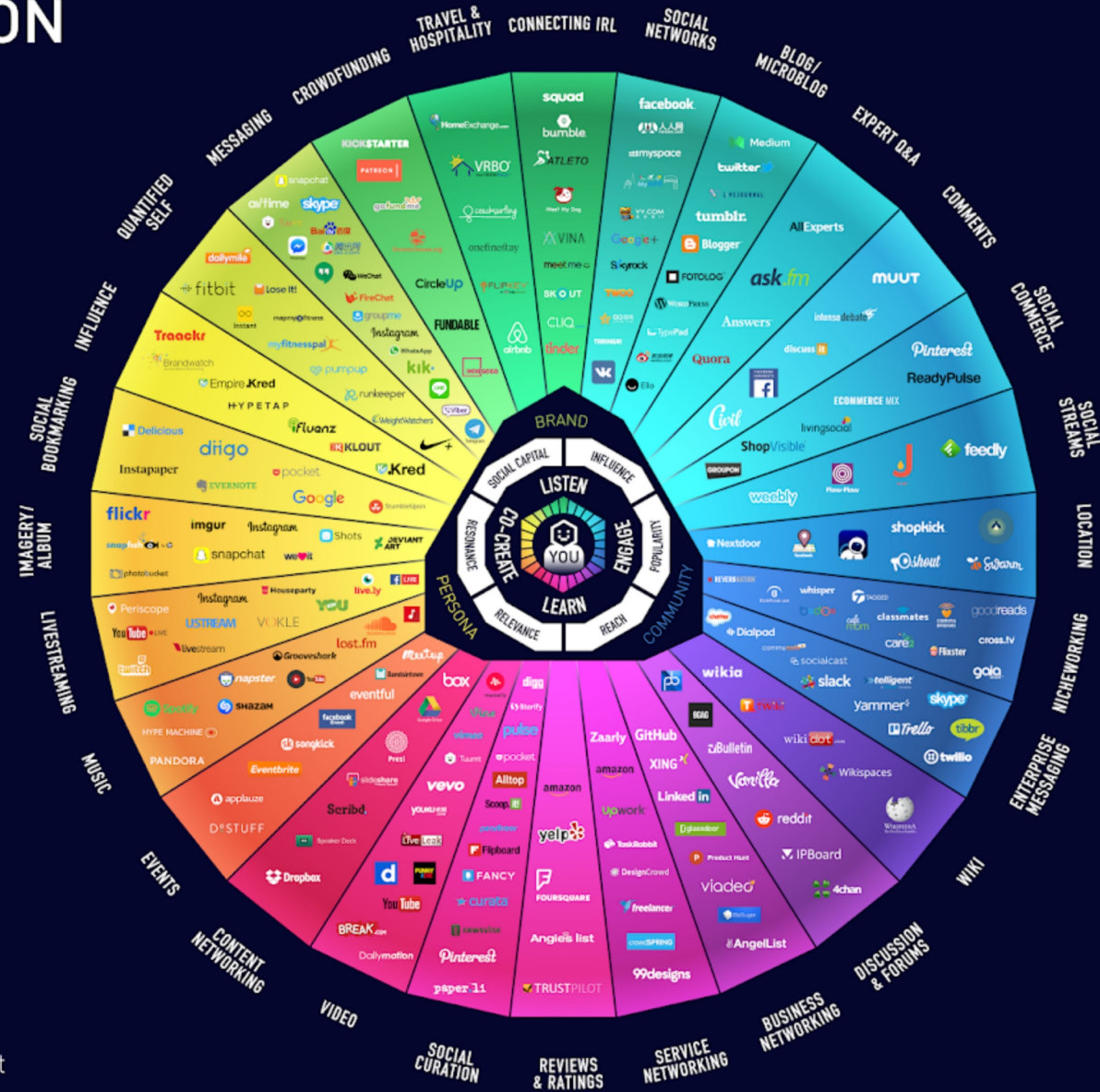
Brought to you by  
Brian Solis & JESS3

### Social Media Gave Everyone a Voice

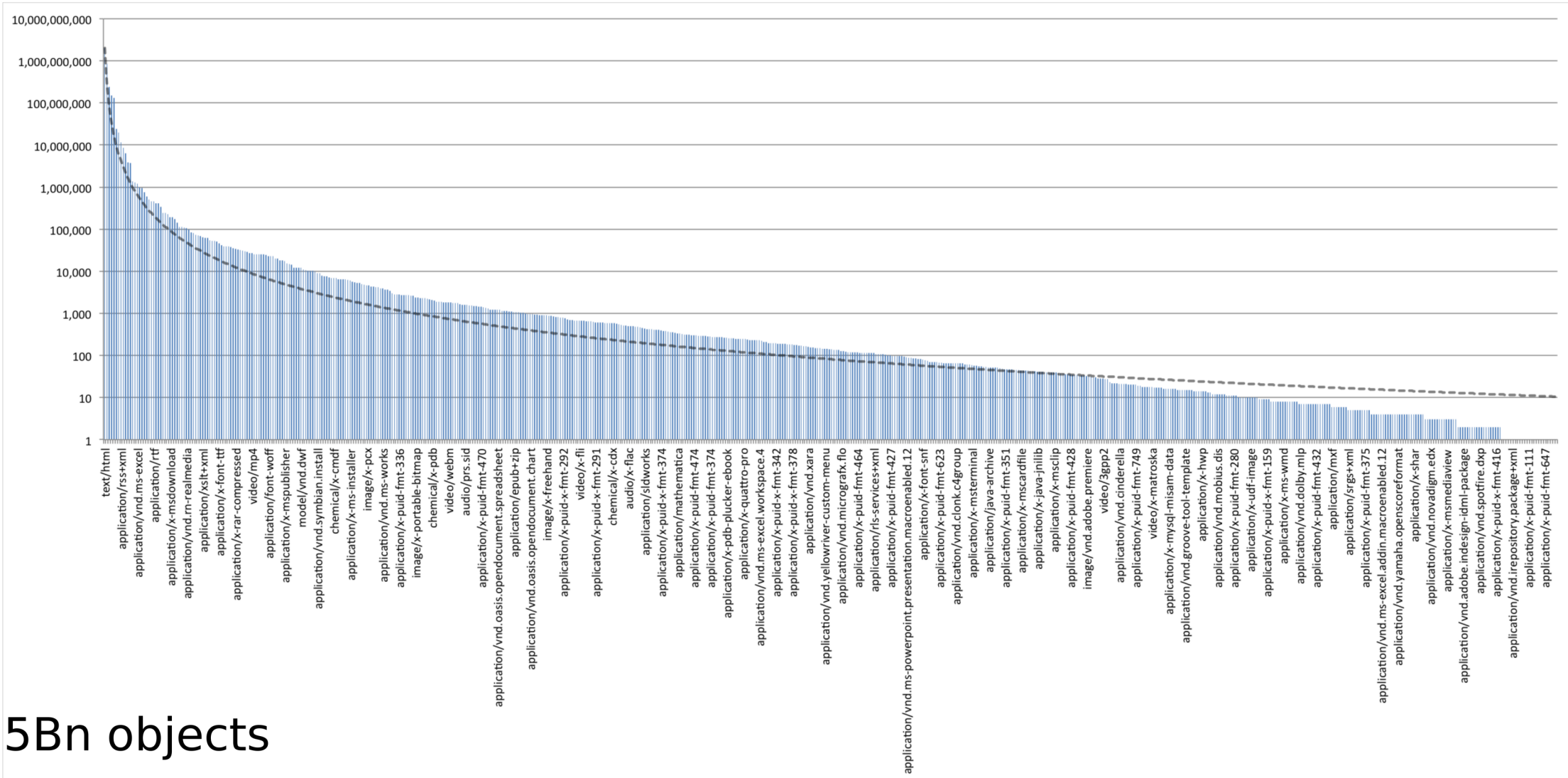
The Conversation Prism debuted in 2008 as social media was exploding online. Social media would change everything about how we communicate, learn and share. It forever democratized information and reset the balance for influence.

The Conversation Prism was designed as a visual map of the conversational networks that continue to reshape everything. Its purpose is to help you understand and appreciate the statusphere so that you can play a productive and defining role in the conversations shaping our future.

For more information check out  
[conversationprism.com](https://conversationprism.com)







3.5Bn objects

99.82% in just 20 formats (html, xml, text, image, video, MS office)

19<sup>th</sup> is 'unknown format' (5,015 extensions)



Secure <https://webrecorder.io/temp-CFSBAPTQ/temp/recording-session/record/http://www.dpconline.org/events>

Webrecorder / Temporary Collection / Recording Session

Recording 2.49 MB (native) Current http://www.dpconline.org/events

Preservation Day

DPC  
EVENTS  
RSS  
FEED

LSE Library

### Research Data Management for All?

18 January 2018 | 10:00 - 16:15 Glasgow, G2 5RJ | Royal College of Physicians and Surgeons Co-Sponsored Event (All Welcome, Registration Essential)

### What I Wish I Knew Before I Started - DPC Student Conference 2018

24 January 2018 | 10:00-16:00 London | Senate House

### Email Preservation: How Hard Can It Be? 2

## Task Force on Technical Approaches to Email Archives

About Working Groups Friends of the TF TF Documents Bibliography

## Managing Email for Preservation

Published March 1, 2018

People and organizations create and receive ever-increasing quantities of mail, which can easily accumulate into an undifferentiated morass of information. But good information management practices can mitigate this risk, making it more likely that valuable messages will remain findable, navigable, and useful. As a side benefit, good email management ensures that that digital files can be preserved and that future generations will be better able understand the activities of the people or organizations that left them behind

## Automatically pack your research to be run elsewhere!

ReproZip allows you to pack your research along with all necessary data files, libraries, environment variables and options.

Then anybody can reproduce the research on a different machine, without tracking down and installing the dependencies, or even having to run the same operating system!

<https://www.reprozip.org/>

Software Library

The Internet Archive Software Library is the ultimate software crate-digger's dream: Tens of thousands of playable software titles from multiple computer platforms, allowing instant access to decades of computer history in your browser through the

Share  
★ Favorite

About Collection

SORT BY VIEWS · TITLE · DATE ARCHIVED · CREATOR

36,859 RESULTS

Search this Collection

collections 28  
software 36,821  
texts 6  
images 3  
data 1

TOPIC

program 1,495  
file 1,264  
apple2 810  
files 697

The Software Library: MS-DOS

7,570 ITEMS

Software Library: MS-DOS Games

2,576 ITEMS

The Oregon Trail

Oregon Trail, The  
by MECG

1.7M 1,153 135

The Software Library: Apple Computer

3,852 ITEMS

Prince of Persia

by Jul 29, 2014

713,056 647 29

PAC-MAN

1. PLAYED GAME  
2. PAC-MAN  
3. PAC-MAN

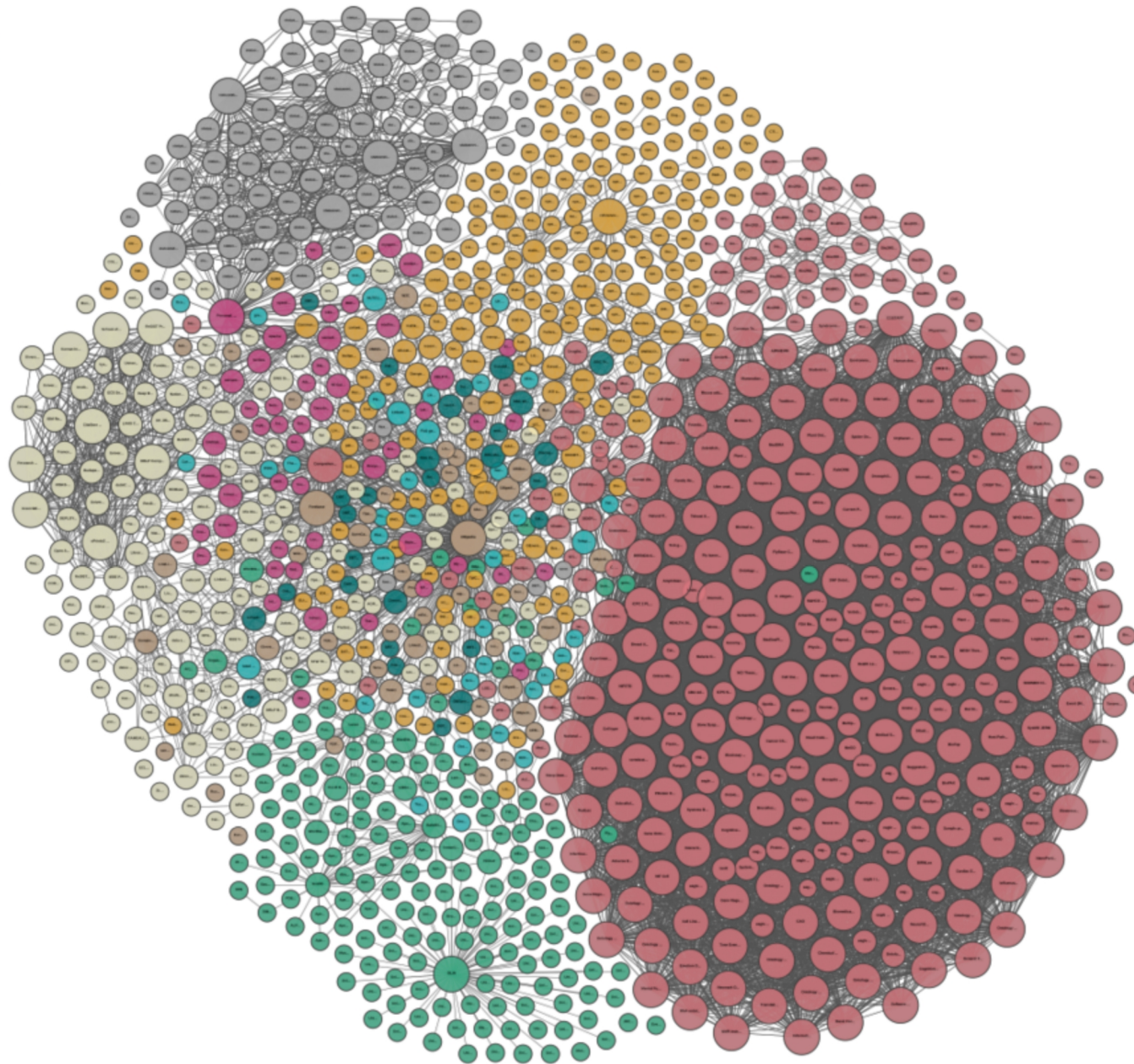


# 3.

## Connectedness







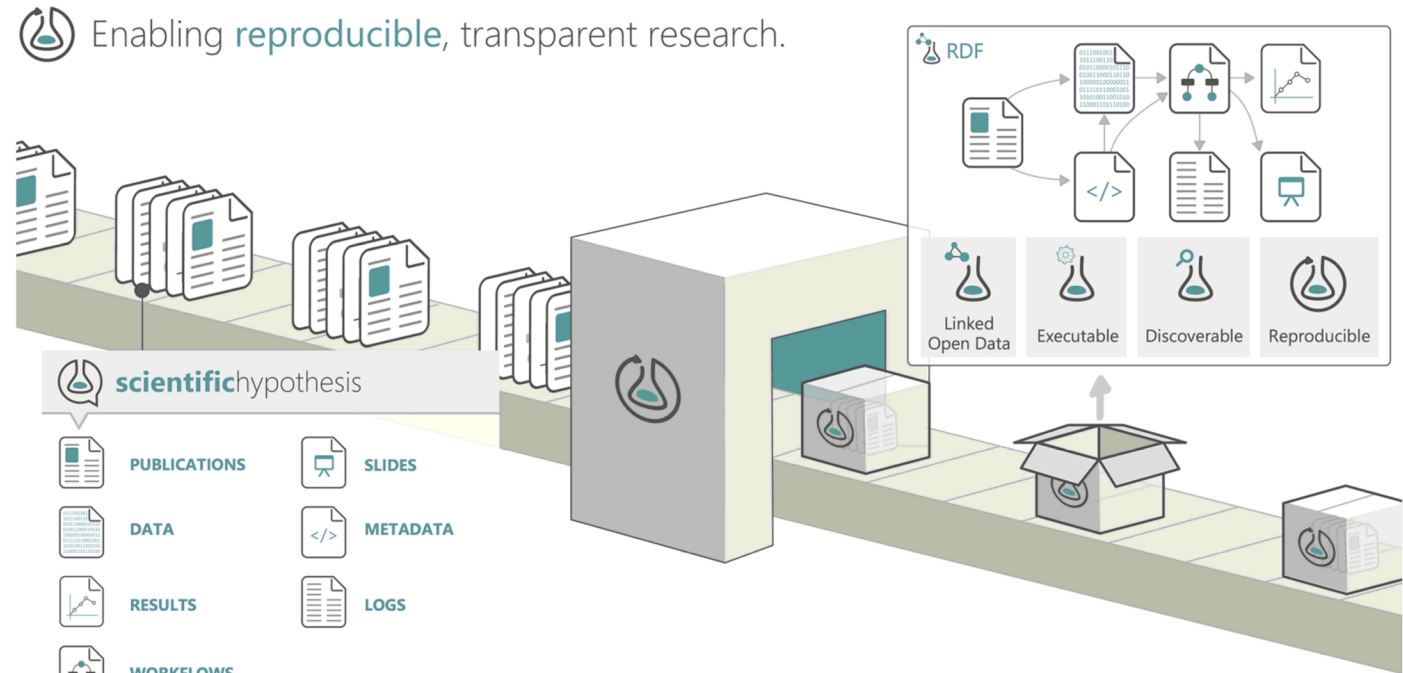


# 4. User experience



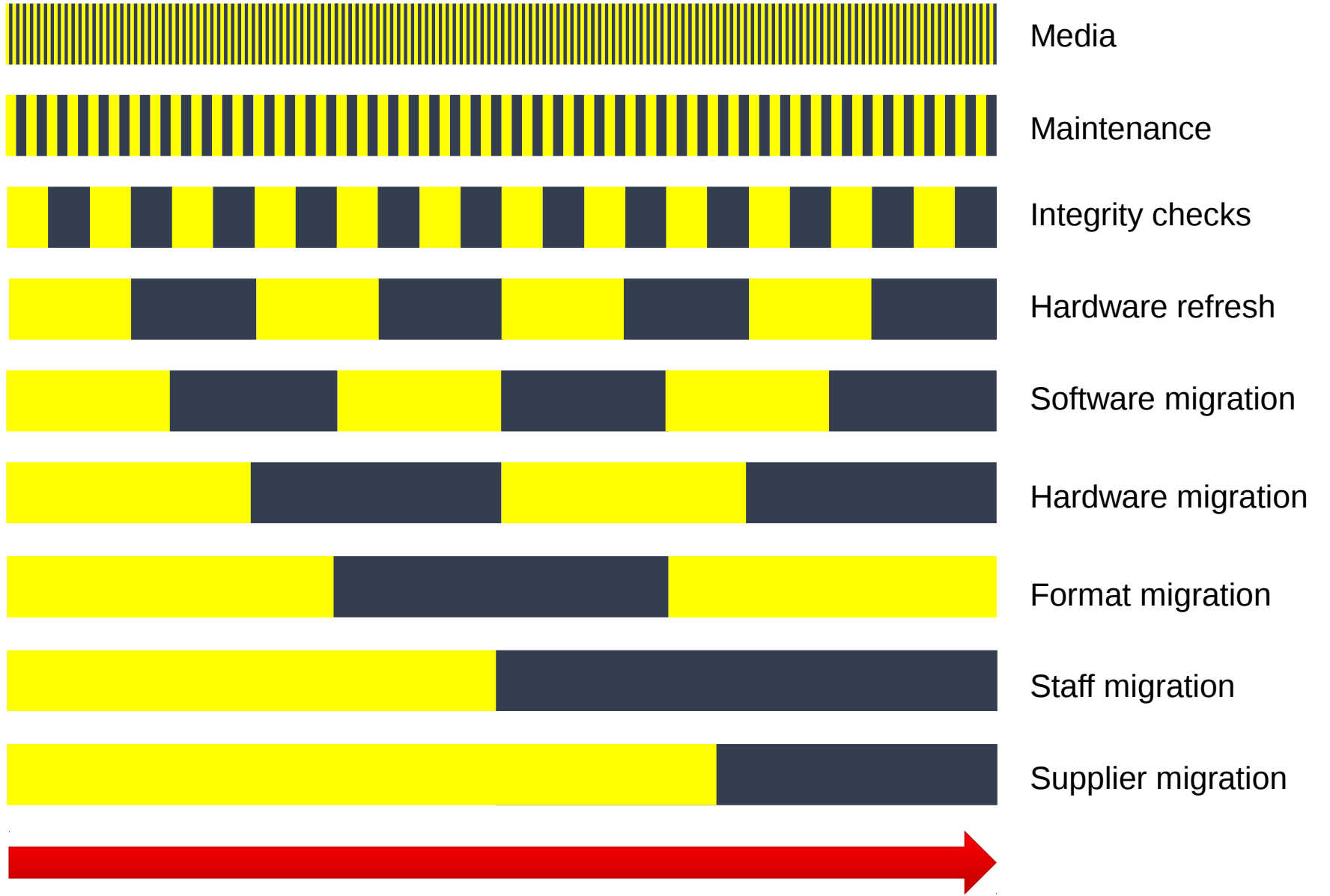


- Record ‘what it was like at the time’
- Ability to recreate an experience, event or outcome
- Evidential weight and authenticity
- Access environments that allow users to interact with and use digital objects
- Packaging and linking of data, metadata, software, workflows, provenance





# 5. Change





- Preservation systems need to be modular
- Choose, integrate, migrate components



- Transparency
- Open Standards
- Open implementations and open source
- Open APIs



# A digital preservation future

**Digital Preservation happens through integrated systems**

**Integration with content sources**

**Integrated tools for different content types**

**Integration with systems for content discovery and use**

**Digital Preservation becomes more automated**

**Automated processing**

**Shared Knowledge**

**An 'engineering' approach**

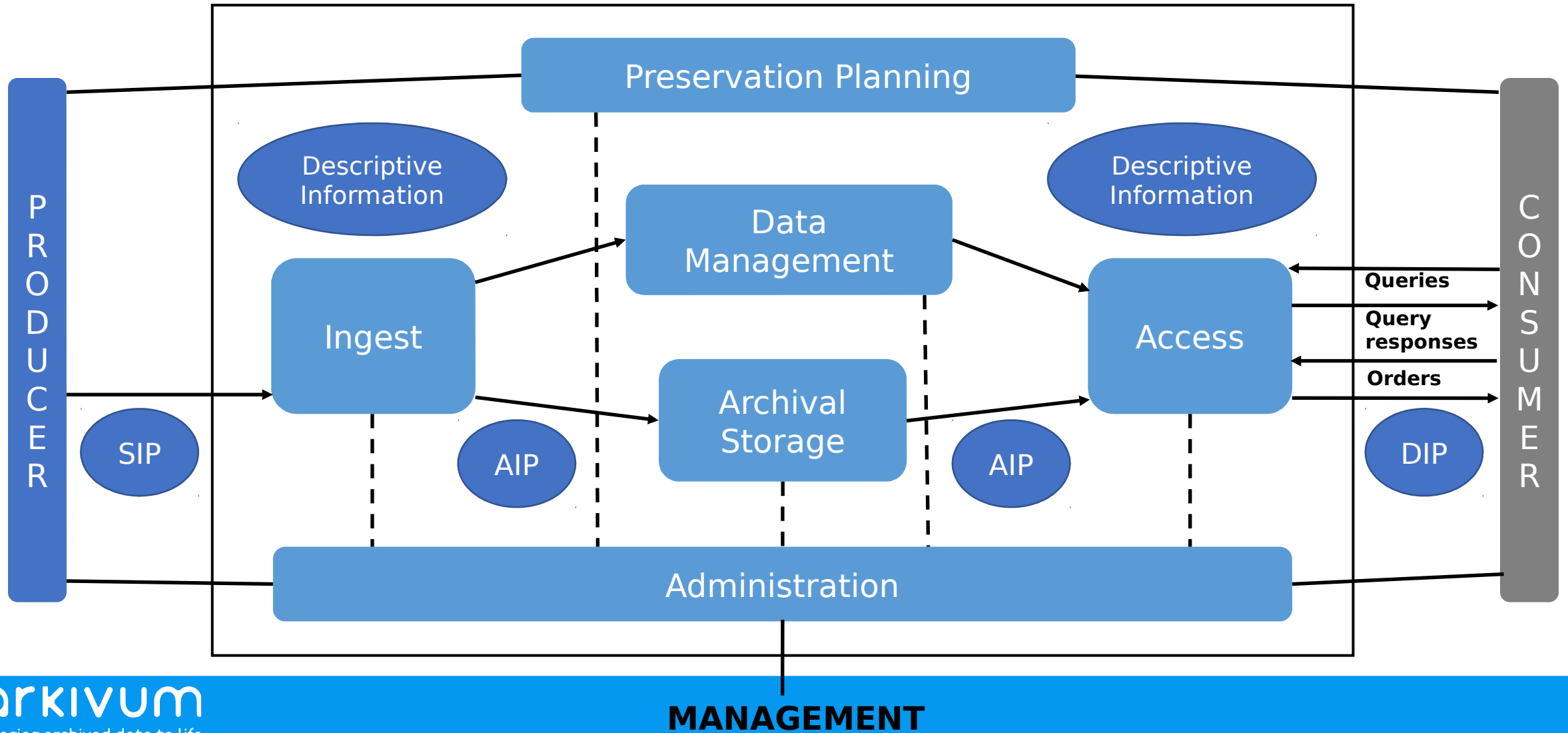
**Users can experience digital content in different contexts**





# Open Archive Information System (OAIS)

(ISO 14731:2012 - OAIS)

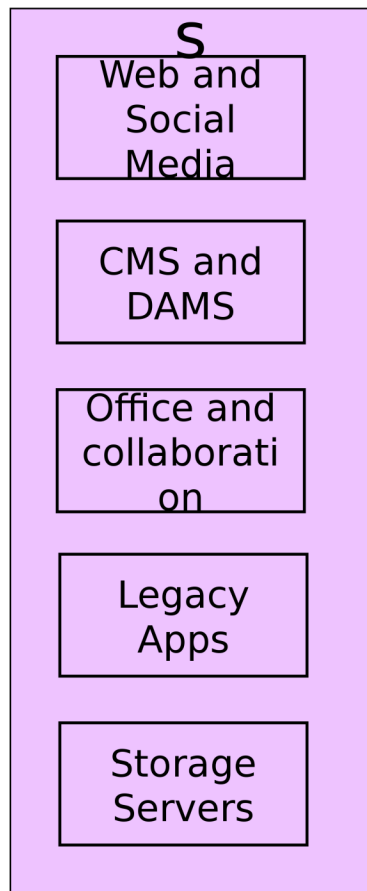


Integrate, transfer, assess and validate

Preserve & Safeguard

Describe, Discover, Embed, Access, Use

## Producer



Integration with content sources

Extraction and Transfer

Checks and Validation

Appraisal and Selection

Documentation and records

Submission Packaging

Identification

Characterisation

Transformation

Packaging (software, data, environments)

Index and Store

Search

Select and Arrange

Embed, export, link

Retrieve/Access

Interact/Reuse

## Consumers

Curation

Research

Public Access

Commercial

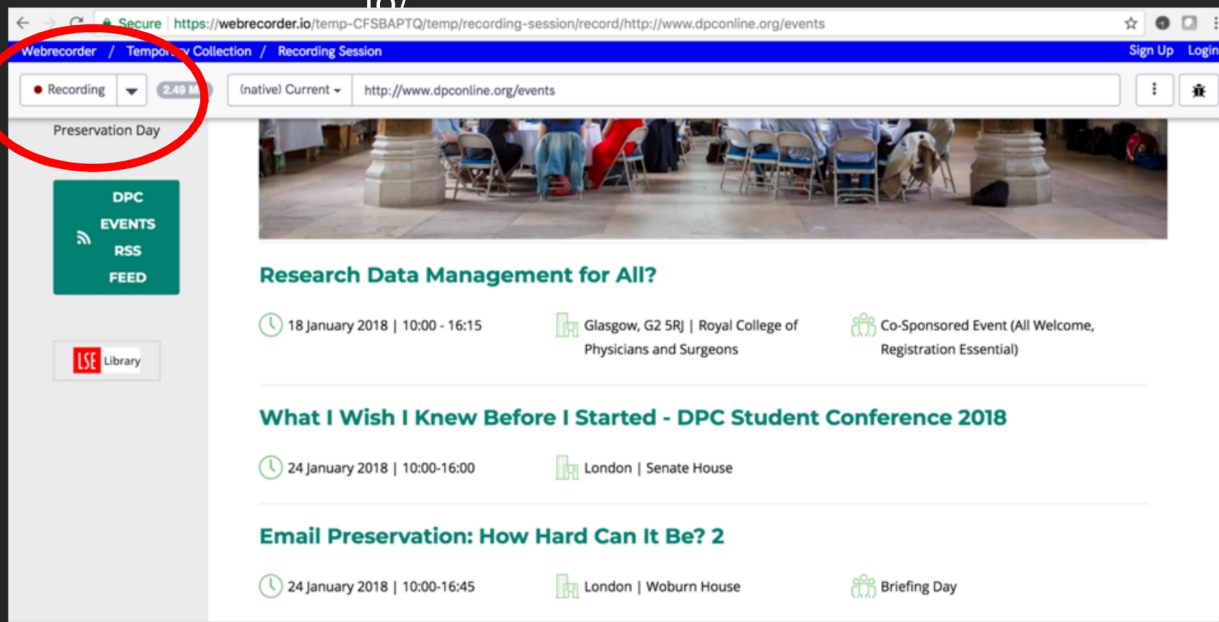
Education

Exhibition

Integration of DP tools and techniques

Integration with users' day-to-day environment and workflows

<https://webrecorder.io/>



Webrecorder / Temp Collection / Recording Session

Recording 2:49 M

(native) Current http://www.dpconline.org/events

Preservation Day

DPC EVENTS RSS FEED

LSE Library

### Research Data Management for All?

18 January 2018 | 10:00 - 16:15 Glasgow, G2 5RJ | Royal College of Physicians and Surgeons Co-Sponsored Event (All Welcome, Registration Essential)

### What I Wish I Knew Before I Started - DPC Student Conference 2018

24 January 2018 | 10:00-16:00 London | Senate House

### Email Preservation: How Hard Can It Be? 2

24 January 2018 | 10:00-16:45 London | Woburn House Briefing Day

## Task Force on Technical Approaches to Email Archives

About Working Groups Friends of the TF TF Documents Bibliography

### Managing Email for Preservation

Published March 1, 2018

People and organizations create and receive ever-increasing quantities of mail, which can easily accumulate into an undifferentiated morass of information. But good information management practices can mitigate this risk, making it more likely that valuable messages will remain findable, navigable, and useful. As a side benefit, good email management ensures that that digital files can be preserved and that future generations will be better able understand the activities of the people or organizations that left them behind

<http://www.emailarchivestaskforce.org/>

**ReproZip**

About

News

Documentation

GitHub Project

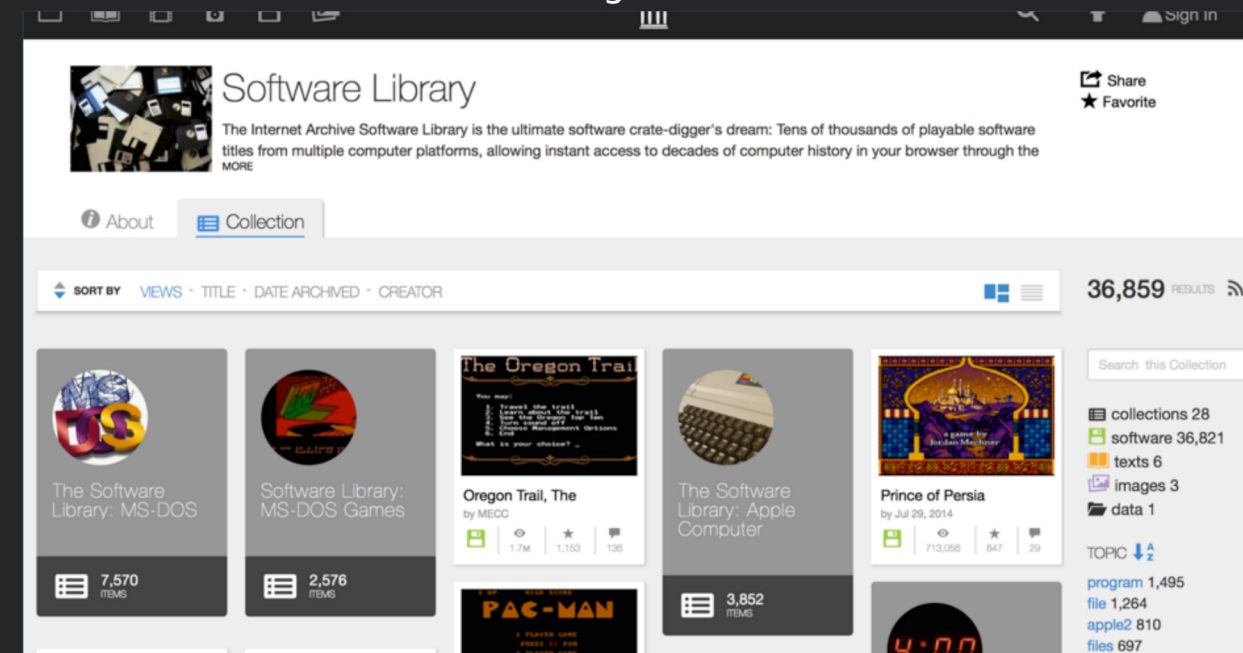
Examples

## Automatically pack your research to be run elsewhere!

ReproZip allows you to pack your research along with all necessary data files, libraries, environment variables and options.

Then anybody can reproduce the research on a different machine, without tracking down and installing the dependencies, or even having to run the same operating system!

<https://www.reprozip.org/>



Software Library

The Internet Archive Software Library is the ultimate software crate-digger's dream: Tens of thousands of playable software titles from multiple computer platforms, allowing instant access to decades of computer history in your browser through the MORE

About Collection

SORT BY VIEWS TITLE DATE ARCHIVED CREATOR

36,859 RESULTS

Search this Collection

collections 28  
software 36,821  
texts 6  
images 3  
data 1

TOPIC

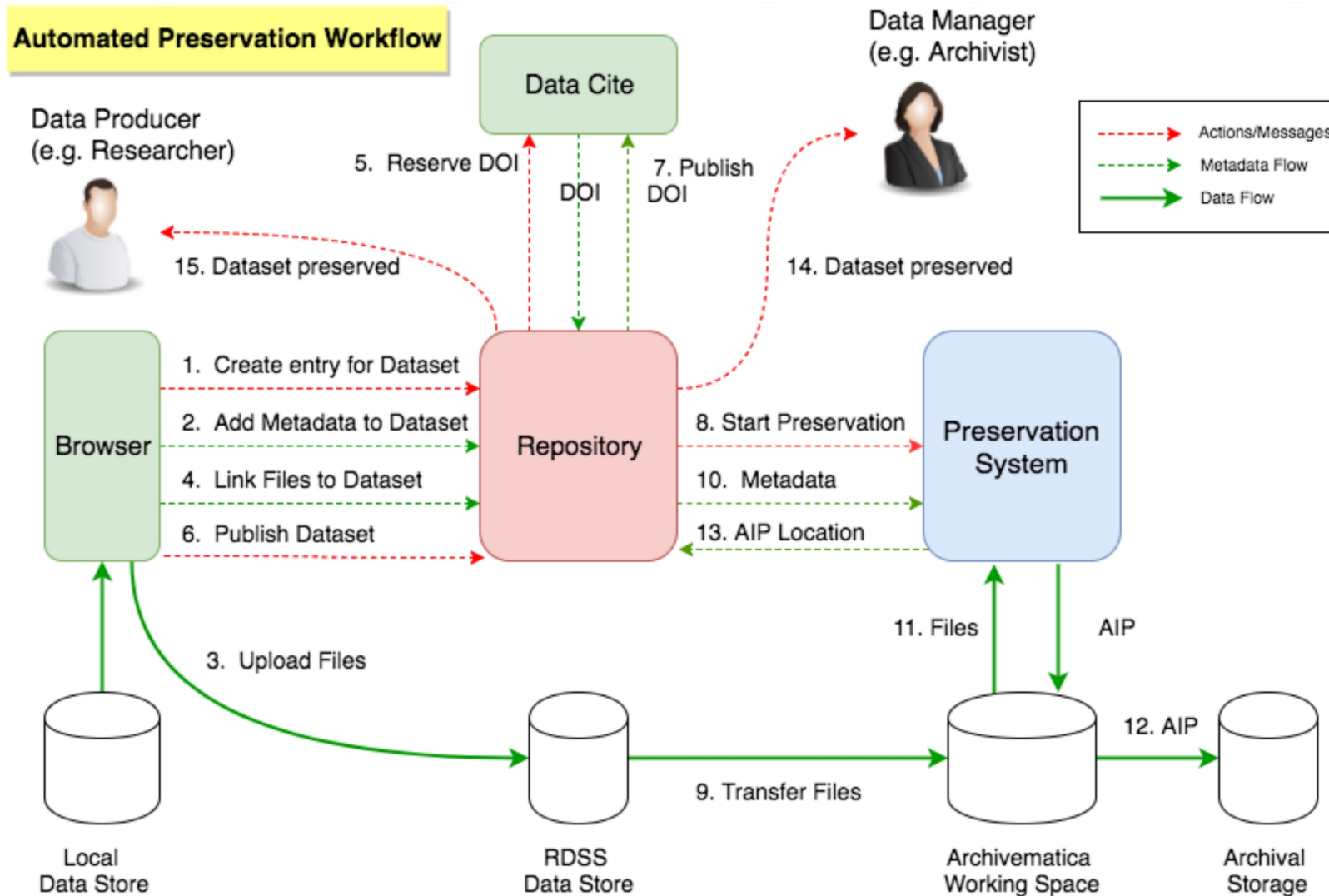
program 1,495  
file 1,264  
apple2 810  
files 697

<https://archive.org/details/softwarelibrary>



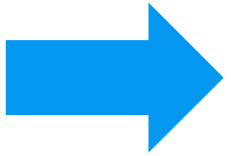


# Case Study: Jisc Research Data



# Making Digital Preservation easier in the RDSS

- Embed preservation into research publication workflows
- Not something that happens at the 'end of the line'
- Reduced manual intervention and specialist
- Preservation 'disappears' into day-to-day life



- Lower costs
- Increased throughput
- Reduced problems
- Less 'slips through the net'

- 220 UK higher education institutions
- 95,000 research staff
- 450 PB of data
- 5,000+ different file types
- DP and storage budget: 1-2%
- 0.5 – 5 dedicated staff per institution
- Cover everything: outreach, training, support, guidelines, policies, workflows, infrastructure, procurement ...





# Hidden Preservation!

## Jisc RDSS Use Case: Open Digital Preservation using Archivemata



Matthew Addis



Justin Simpson

<http://doi.org/ccz6>

PASIG\_11Sep2017\_DP\_JISC\_RDSS\_v5.pdf (16.07 MB)

1 / 3

MD5: 7cf770beec75f2f8a62e81898f8cd8f2 |

Cite

Download (16.07 MB)

Share

Embed

+ Collect

### Jisc RDSS Use Case: Open Digital Preservation using Archivemata

Version 4 12.09.2017, 14:01 by Matthew Addis, Justin Simpson

This presentation describes how Arkivum and Artefactual have worked together to provide an open digital preservation solution based on Archivemata for the Jisc

552 views

62 downloads

0 citations



PASIG\_11Sep2017\_DP\_JISC\_RDSS\_v5.pdf 16.07 MB

needed to publish & get DOI

**Title**

Jisc RDSS Use Case: Open Digital Preservation using Archivemata

**Authors**

Matthew Addis Justin Simpson

Add co-authors by name, full email or ORCID. Hit enter after each

**Categories**

Information Systems

**Item type** (what's this?)

Presentation

**Keyword(s)**

Archivemata digital preservation research data

research data management Jisc RDSS

PASIGOXFORD pasigoxford

Add keywords for easy discovery. Hit enter after each

**Description**

This presentation describes how Arkivum and Artefactual have worked together to provide an open digital preservation solution based on Archivemata for the Jisc Research Data Shared Service. Topics include the challenges and drivers for research data preservation in higher education institutions, the approach we've taken and solution we've developed, and future work that could be done to further collaborative digital preservation for research data.

The presentation was made at the Preservation and

**References**

<https://pasigoxford.org/>

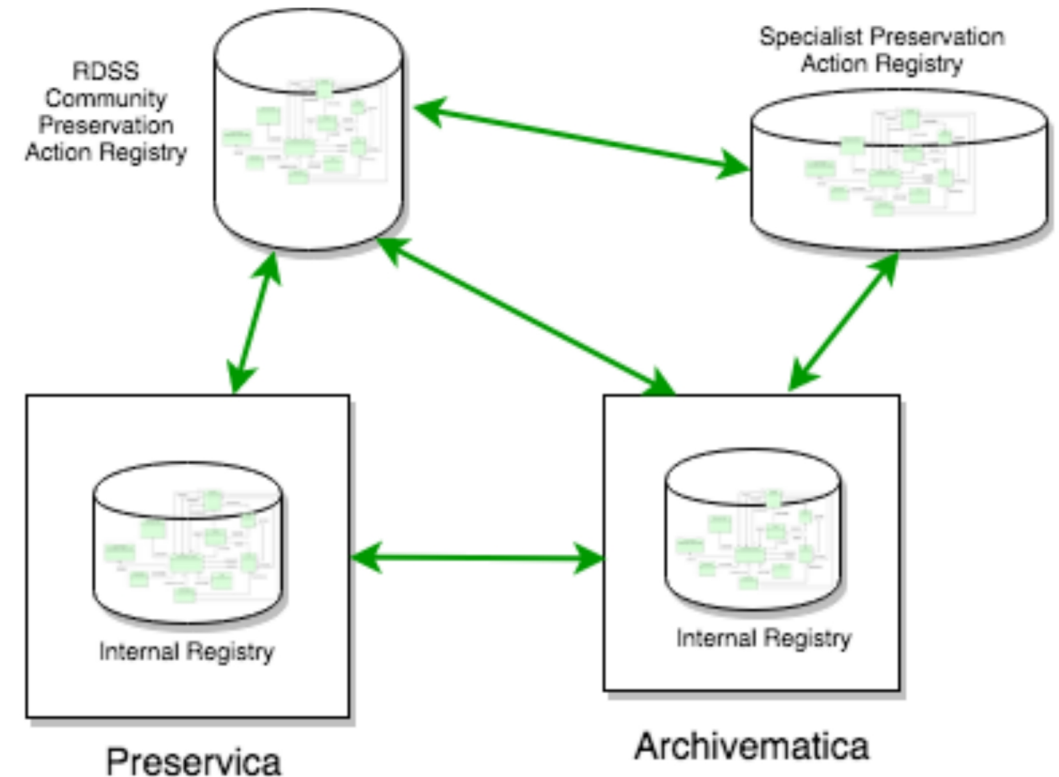
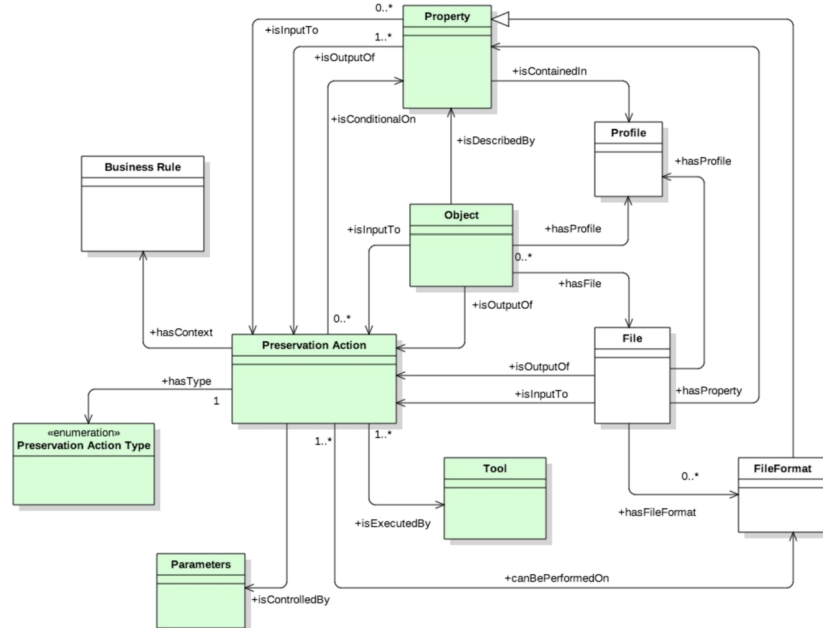
**Funding**

Cancel DOI Public **Save changes**



# Preservation Actions Registry

- What preservation actions to take
- How to implement them
- Capture and share knowledge
- Trusted set of rules and tools
- Executable by preservation systems

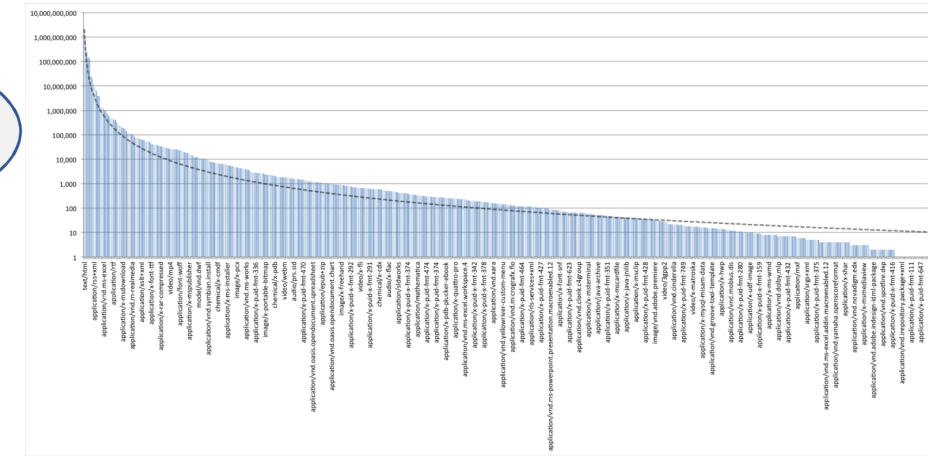


# An 'engineering approach' to Digital Preservation

Automatic Preservation

Semi-automated Preservation

Interactive Preservation



Volume of content

Majority of content, low cost

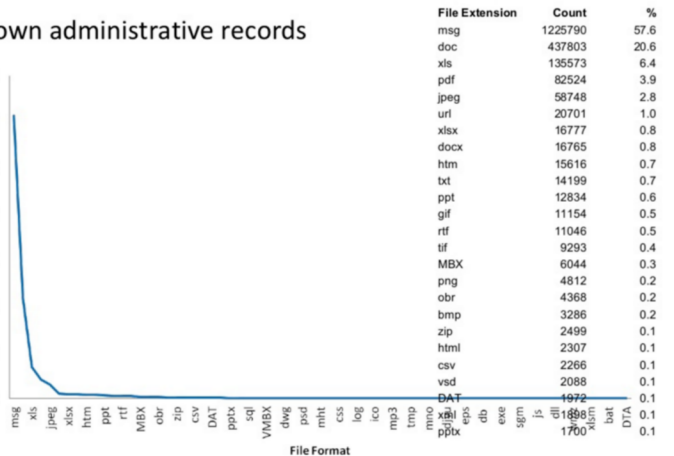
Minority of content, high cost

File types

## The File format Long Tail

- Our own administrative records

Number of files





# Experiences, authenticity, environments for reuse



[Home](#) [Emulation as a Service](#) [Use Cases / Demos](#) [Publications](#)

[twitter](#) [blog](#)

[EaaS in the Cloud](#) | [Whole System Preservation](#) | [Mass-curation of objects and systems](#)

(Re-)publication of Preserved, Interactive Content – Theresa Duncan CD-ROMs:  
Visionary Videogames for Girls



# A digital preservation future

**Digital Preservation will happen through integrated systems**

**Easier to capture, preserve, find and use content**

**Combine different techniques, tools and workflows**

**Open and flexible, avoid lock-in, easier to sustain**

**Digital Preservation will become more automated**

**Lower costs, scale-up**

**Focus valuable staff resources on the 'hard and valuable' content**

**Community sharing of knowledge and expertise**

**Users can interact with digital content in different contexts**







arkivum  
Bringing archived data to life



---

matthew.addis@arkivum.  
com



orcid.org/0000-0002-3837-  
2526



www.arkivum.com

---





