

# Preservation planning at the British Library

Michael Day

British Library  
Michael.Day@bl.uk

4C Workshop 6: Preservation Planning, London,  
18 November 2014

# Presentation outline

- The British Library context
- Approaches to preservation planning
- Collection profiling
- Format assessment
- Summing up

# British Library digital collections ...

**Chapter 1**

Alice was beginning to get very tired of sitting by her sister on the bank and of having nothing to do: once or twice she had peeped into the book her sister was reading, but it had no pictures or conversations in it, and

Archived November 2005 Archived May 2006 Archived June 2007 Archived March 2009 Archived March 2005 Archived November 2006 Archived November 2008 Archived May 2009

**The Manchester Guardian.**

Grid reference at centre: NY 900 351

**The Tragical History of HAMLET**

**ETHOS**  
**ELECTRONIC THESES ONLINE SERVICE**  
Opening access to UK theses

Dear Wendy  
thank you so much, you have made my day! as you are a hero of mine.  
I will be very careful and will include everything you wanted. People will really enjoy these,  
Best wishes  
Paula

# Strategic infrastructure

## Content Strategy

“Outside of legal deposit, we will prefer to connect to content except in circumstances where the connection is not technically feasible or when we wish to hold and preserve the materials for the long term”

## Digital Preservation Strategy

Improve our Repository

Manage risks across lifecycle

Embed digital sustainability

Collaborate!

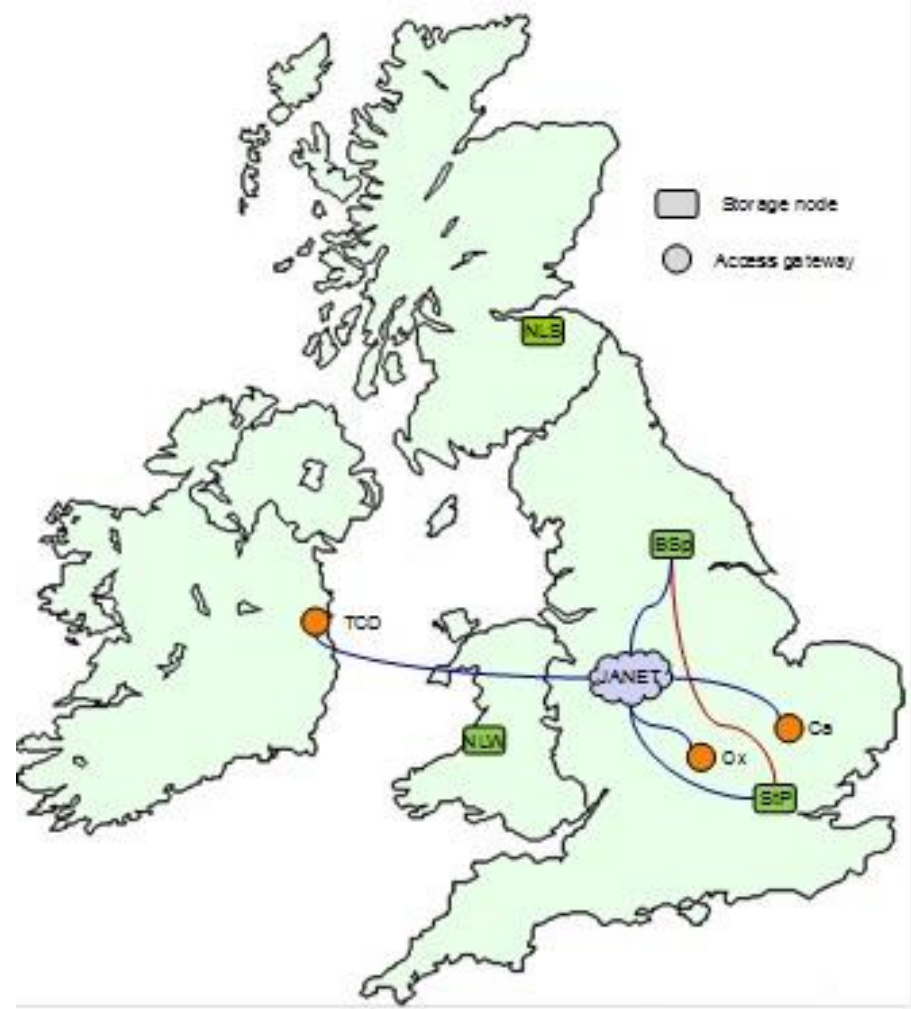
# Technical Infrastructure

## Storage nodes:

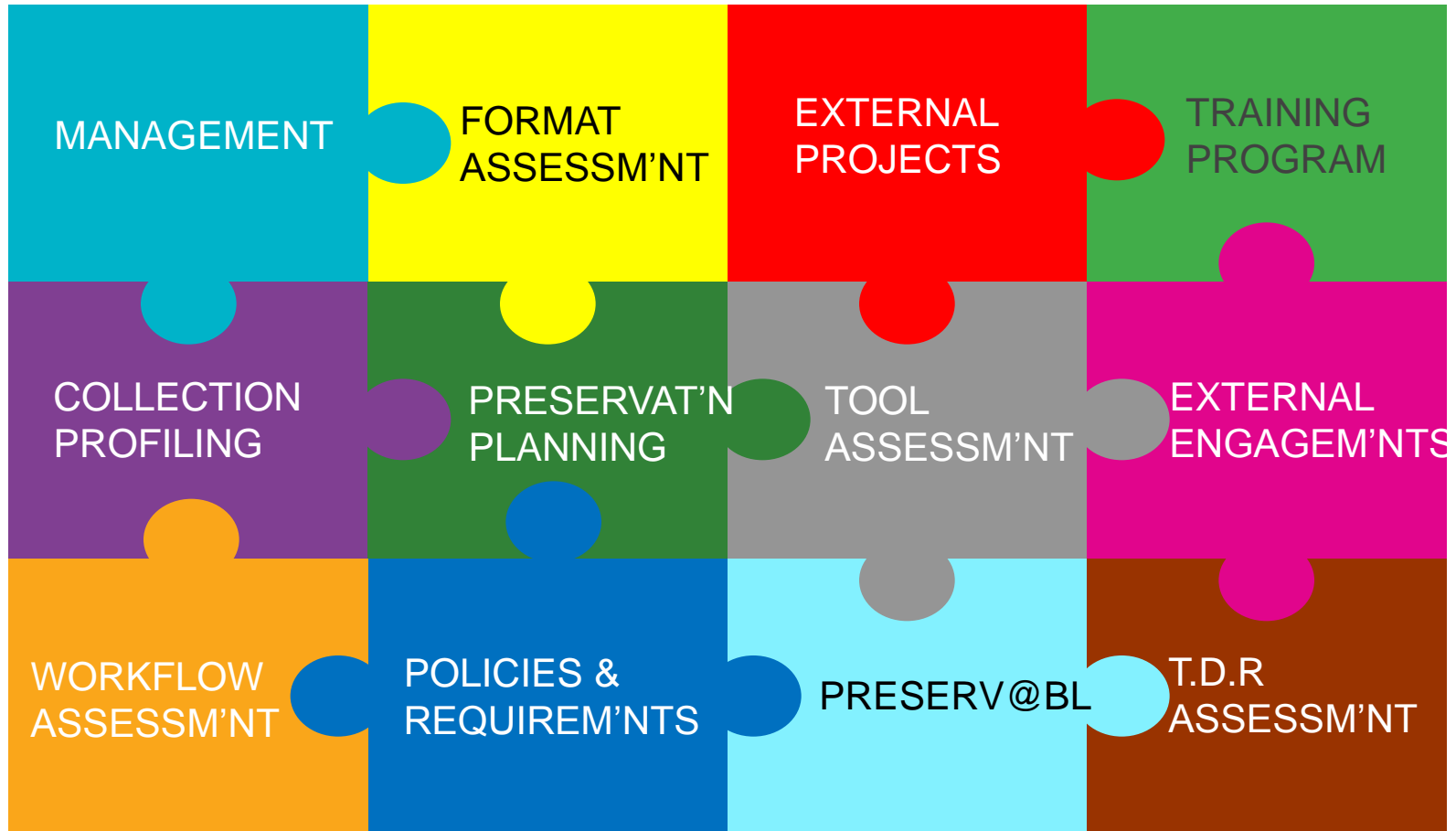
- BL St Pancras (STP)
- BL Boston Spa (BSP)
- National Library of Wales (NLW)
- National Library of Scotland (NLS)

## Access gateways:

- Bodleian Library (Ox)
- Cambridge University Library (Ca)
- Trinity College Library, Dublin (Tr)



# Programme of Work



[Maureen Pennock, iPRES 2014]

# Approaches to preservation planning

- Digital preservation is business as usual
- Principles for developing a preservation planning capacity:
  - Incremental (no need to do everything at once)
  - Practical (e.g. focused on institutional requirements)
  - Flexible
    - No one-size-fits-all solution (although needs to operate at scale)
    - Needs to cover a wide range of content types
    - Takes into account the whole life-cycle
    - Other stakeholders (e.g. legal deposit libraries)
  - Evidence-based (e.g., assessment-based activities)

# Collection profiling (1)

- What content do we have and what is important about it?
  - Digital Asset Register
- Collection profiling:
  - Documenting key knowledge about the Library's high-level digital collections
  - Considering preservation requirements / preservation intent for these collections
  - A mechanism for liaising with curators and collection specialists



# Collection profiling (2)

- Taxonomy of high-level collections:
  - No standard list of content types
  - Various lists are available on the Library website and catalogue, but they are inconsistent
  - Needed a pragmatic starting point for collection profiling:
    - Developed a new taxonomy based on the Library's existing digital asset register
    - Attempted to identify *logical* groupings, e.g. ignoring distinctions between digitised and 'born-digital' content, where possible
    - It is not perfect (and subject to revision)

# Collection profiling (3)

- The profile framework (has evolved over time):
  - Summary
  - What is this format, and what is it within the Library?
  - Acquisition and acquisition formats
  - Preservation Intent
  - Issues
  - Summary table of sub-collections
- A first step in an ongoing process of capturing preservation requirements

# Format assessments (1)

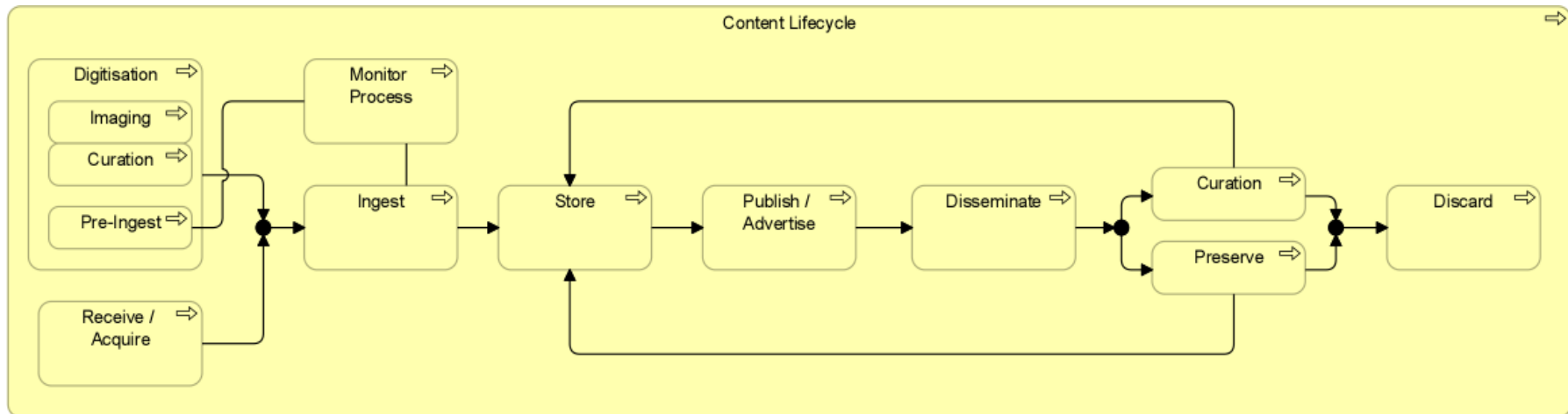
- Sustainability assessments of formats
  - Interacts with ongoing discussion on “preservation masters”
  - Less concerned with theoretical “obsolescence,” but aware of institutional factors in maintaining access to content
  - Principles:
    - Based, where possible, on real-world evidence (documented examples)
    - Clarity on purpose and intended audience
    - Not based on numerical scoring measures

# Format assessments (2)

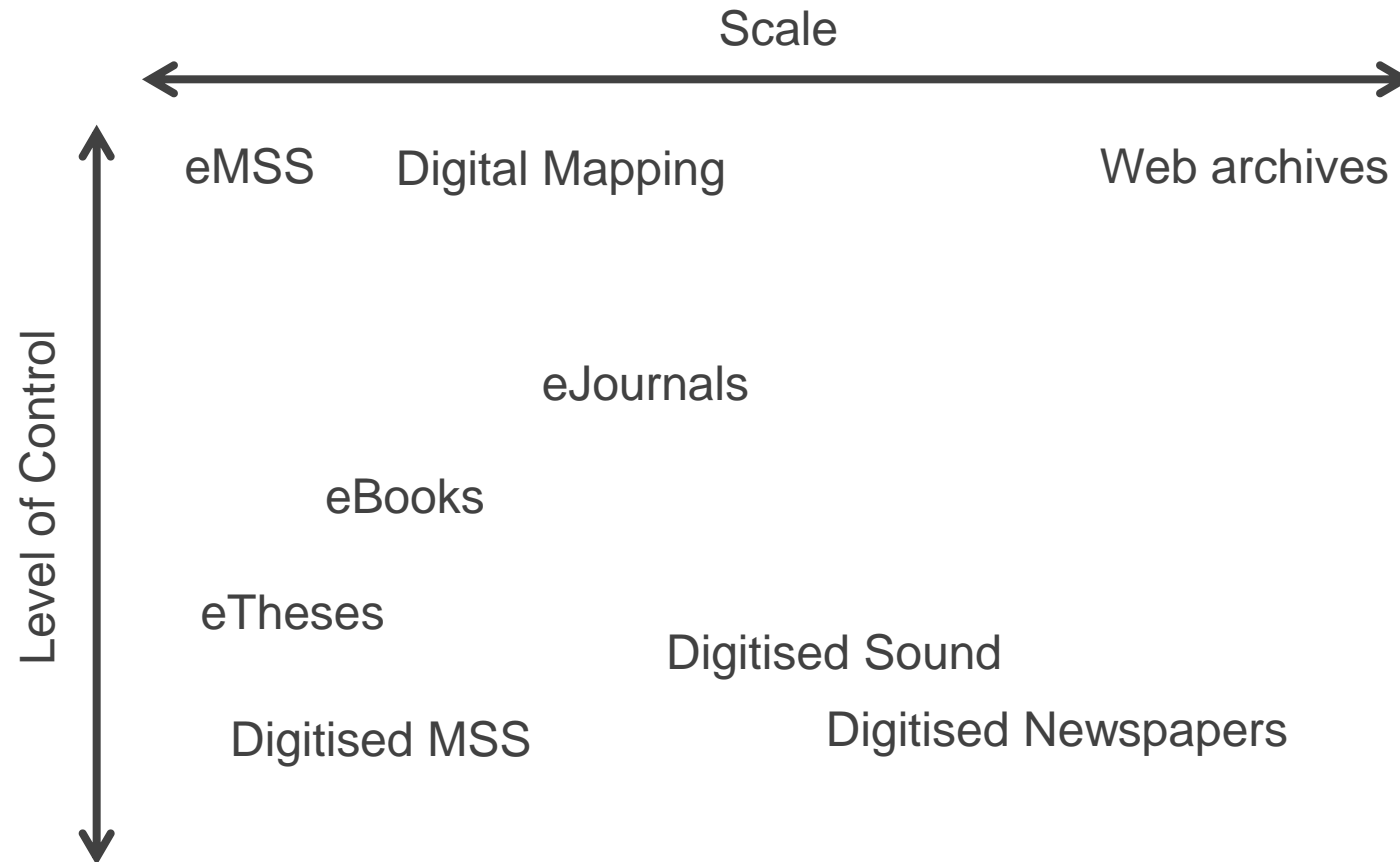
- Main criteria
  - Development Status
  - Adoption and Usage
  - Software Support
  - Documentation and Guidance
  - Complexity
  - Embedded or attached Content
  - External Dependencies
  - Legal Issues
  - Technical Protection Mechanisms
  - Other ...

# Developing preservation planning capacity

- Other tasks:
  - Tool assessment
  - Workflow assessment (considering content lifecycle)
  - Policies
  - Training



# Planning for both scale and complexity



# Summing up

- Preservation Planning in OAIS
  - “The OAIS functional entity which provides the services and functions for monitoring the environment of the OAIS and which provides recommendations and preservation plans to ensure that the information stored in the OAIS remains accessible to, and understandable by, and sufficiently usable by, the Designated Community over the Long Term, even if the original computing environment becomes obsolete.”
- There is a long way to go (but we *are* getting started)

Thank you



# Preservation planning at the Wellcome Library

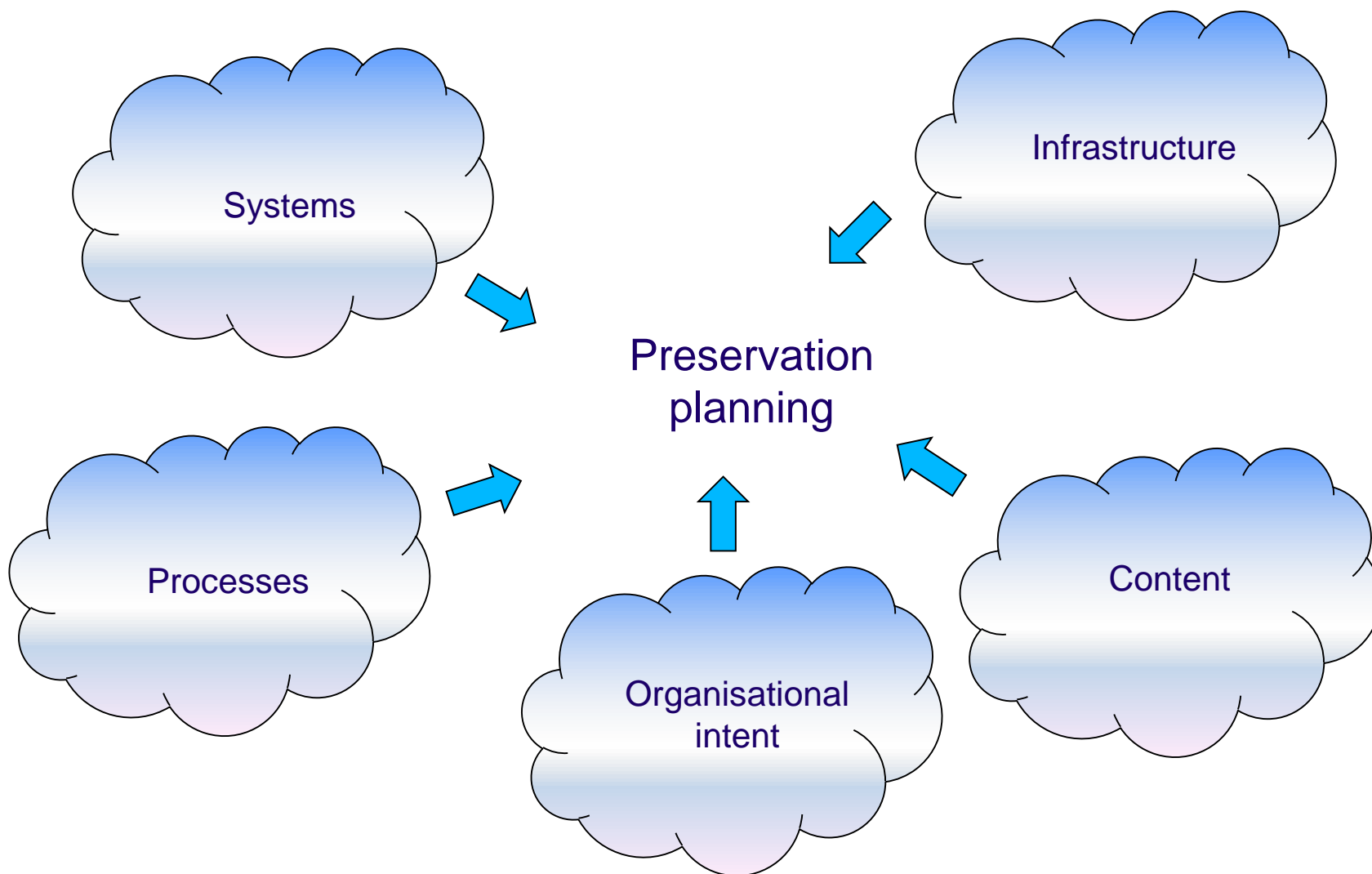
Investing in Opportunity: Policy Practice and Planning for a Sustainable Digital Future, November 2014

Dave Thompson  
Digital Curator, Wellcome Library

# The Wellcome Library has a published Preservation Policy

<http://wellcomelibrary.org/about-us/library-strategy-and-policy/preservation-policy/>





1. Preservation planning & digital content.
2. Systems & processes



# Understanding our content

- Built on characterisation reporting out of Preservica.
- Practical hands on investigation.
- Created a spread sheet that is updated monthly.
- Knowing what we have.



	A	B	C	D	E	F	G	H	I
	File format	Version	Format PUID	Frequency	% of BD collection	Approx date of creation	Migration pathway in Preservica?	Supported by vendor?	Rendering application(s)
1	Hypertext Mark-up Language	4.01	fmt/100	20138	51.894	1980s	Yes. To PDF 1.4 & PDFa	CERN / W3C	Web browser / HTML editor
2	Acrobat PDF 1.4 - Portable Document Format	1.4	fmt/18	2292	5.906	2001		No, Adobe	Adobe PDF Reader / Various open source PDF readers
3	JPEG File Interchange Format	1.01	fmt/43	2244	5.783		Yes. To PNG 1.0 GIF 89a PDF 1.3 PDFa PDF 1.4 & JPEG2000	Yes, Joint Photographic Experts Group	Image rendering / editing software / web browser
4	Windows Portable Executable		x-fmt/411	1500	3.865				Microsoft Windows
5	Tagged Image File Format		fmt/353	1356	3.494		Yes. To PDFa JPEG 1.01 PDF 1.4 PDFa GIF 1989a JPEG 1.02 JPEG2000 & PNG 1.0	Yes, Adobe	Image editing software
6	JPEG File Interchange Format	1.02	fmt/44	1157	2.981		Yes. To PNG 1.0 GIF 89a PDF 1.3 PDFa PDF 1.4 & JPEG2000	Yes, Joint Photographic Experts Group	Image rendering/editing software
7	Hypertext Mark-up Language		fmt/96	1230	3.170		Yes. To PDF 1.4 & PDFa	Yes, W3C, but depends on version	Web browser / HTML editing application
8	Plain Text File		x-fmt/111	1035	2.667			No	Any word processor / text editing tool
9	Microsoft Word for Macintosh Document	5	x-fmt/65	661	1.703	1991		Depends on version & network settings, Microsoft	Microsoft Word / Word free viewer / Libre Office / Open Office
10	Acrobat PDF 1.3 - Portable Document Format	1.3	fmt/17	653	1.683	2000	Yes. To XML 1.0 & PDFa	No, Adobe	Adobe PDF Reader Various open source PDF readers
11	Acrobat PDF 1.0 - Portable Document	1	fmt/14	635	1.636	Mid 1990s	Yes. To XML 1.0 PDF 1.3	No. Adobe	Adobe PDF Reader

	A	B	C	D	E	F	G	H	I
	File format	Version	Format PUID	Frequency	% of BD collection	Approx date of creation	Migration pathway in Preservica?	Supported by vendor?	Rendering application(s)
1									
2	Hypertext Mark-up Language	4.01	fmt/100	20138	51.894	1980s	Yes. To PDF 1.4 & PDFa	CERN / W3C	Web browser / HTML editor
3	Acrobat PDF 1.4 - Portable Document Format	1.4	fmt/18	2292	5.906	2001		No, Adobe	Adobe PDF Reader / Various open source PDF readers
4	JPEG File Interchange Format	1.01	fmt/43	2244	5.783		For 1.3 For 1.4 & JPEG2000	Experts Group	Image rendering / editing software / web browser
5	Windows			1500	3.865				Windows
6	Tagged			1356	3.494				Editing software
7	JPEG File Interchange Format	1.02	fmt/44	1157	2.981				Rendering/editing
8	Hypertext Mark-up Language		fmt/96	1230	3.170				Browser / HTML editing
9	Plain Text File		x-fmt/111	1035	2.667			No	Any word processor / text editing tool
10	Microsoft Word for Macintosh Document	5	x-fmt/65	661	1.703	1991		Depends on version & settings,	Microsoft Word / Word free viewer / Libre Office / Open Office
11	Acrobat PDF 1.3 - Portable Document Format	1.3	fmt/17	653	1.683				Adobe PDF Reader Various open source PDF readers
12	Acrobat PDF 1.0 - Portable Document	1	fmt/14	635	1.636	Mid 1990s	Yes. To XML 1.0 PDF 1.3	No. Adobe	Adobe PDF Reader

How many?

Percentage  
of our  
collection

Pink  
indicates  
falling  
numbers

Format

# Tells us...scope of our material

- Single dynamic point of information about our collection.
- Supports decision making around interventions.
- Shows us what is exceptional.
- Supports collection decisions.





# Further information

- Preservica report a good start.
- Added % of total collection for given fmt.
- Sources of further information.
- Indicators of growth/contraction.



# But it's not only about content

Not yet addressed;

- Need to consider systems, processes, IT infrastructure.
- How we process/manage content. Workflows etc.
- How we secure data.
- How we maintain infrastructure.
- Organisational stuff.

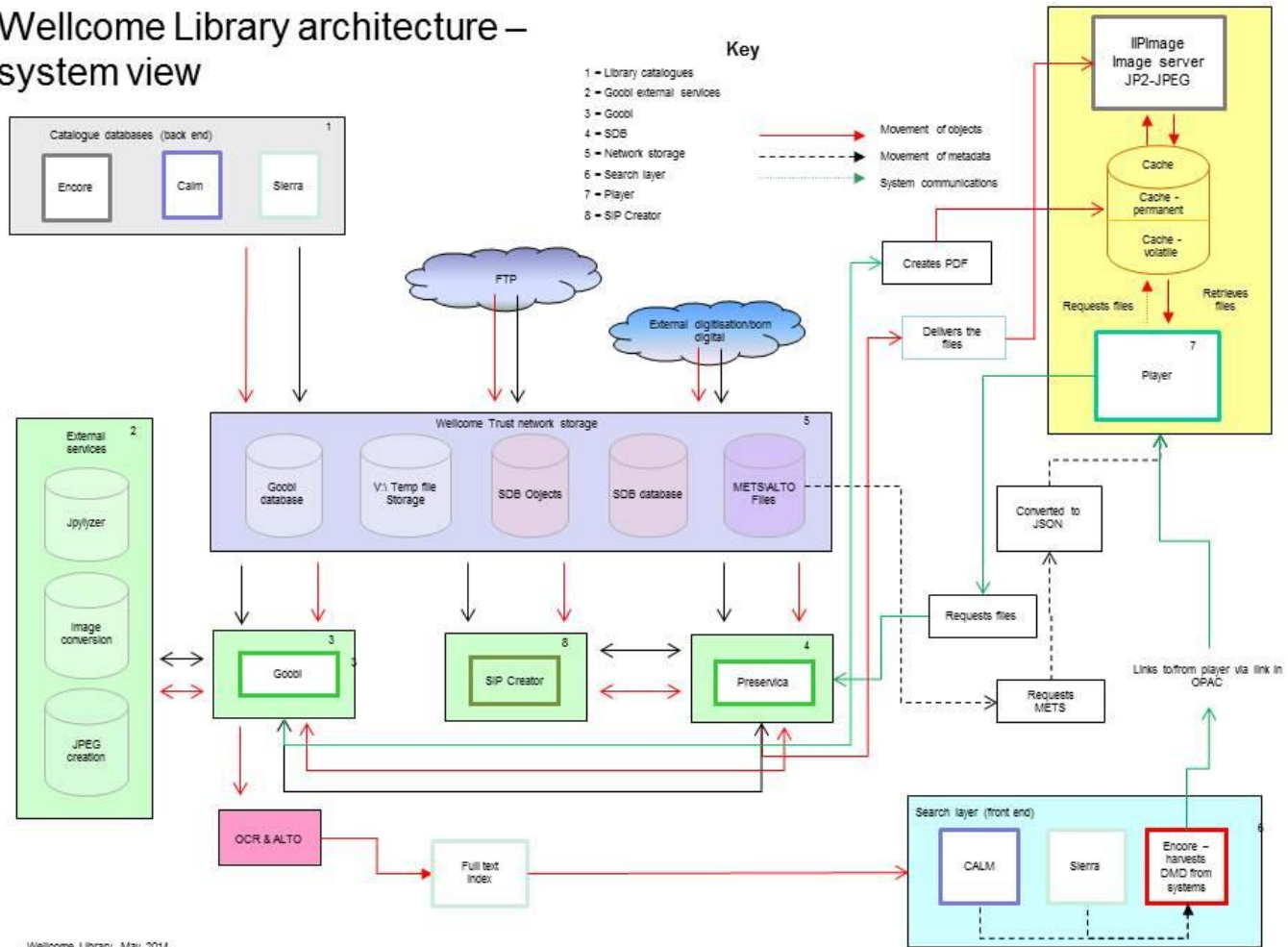
# Not as bad as we thought

- Have a finite set of tools, set of systems.
- Tools mostly workflow based, predictable & consistent.
- Clear set of processes depending on source/type of content.
- Clear set of roles/responsibilities.

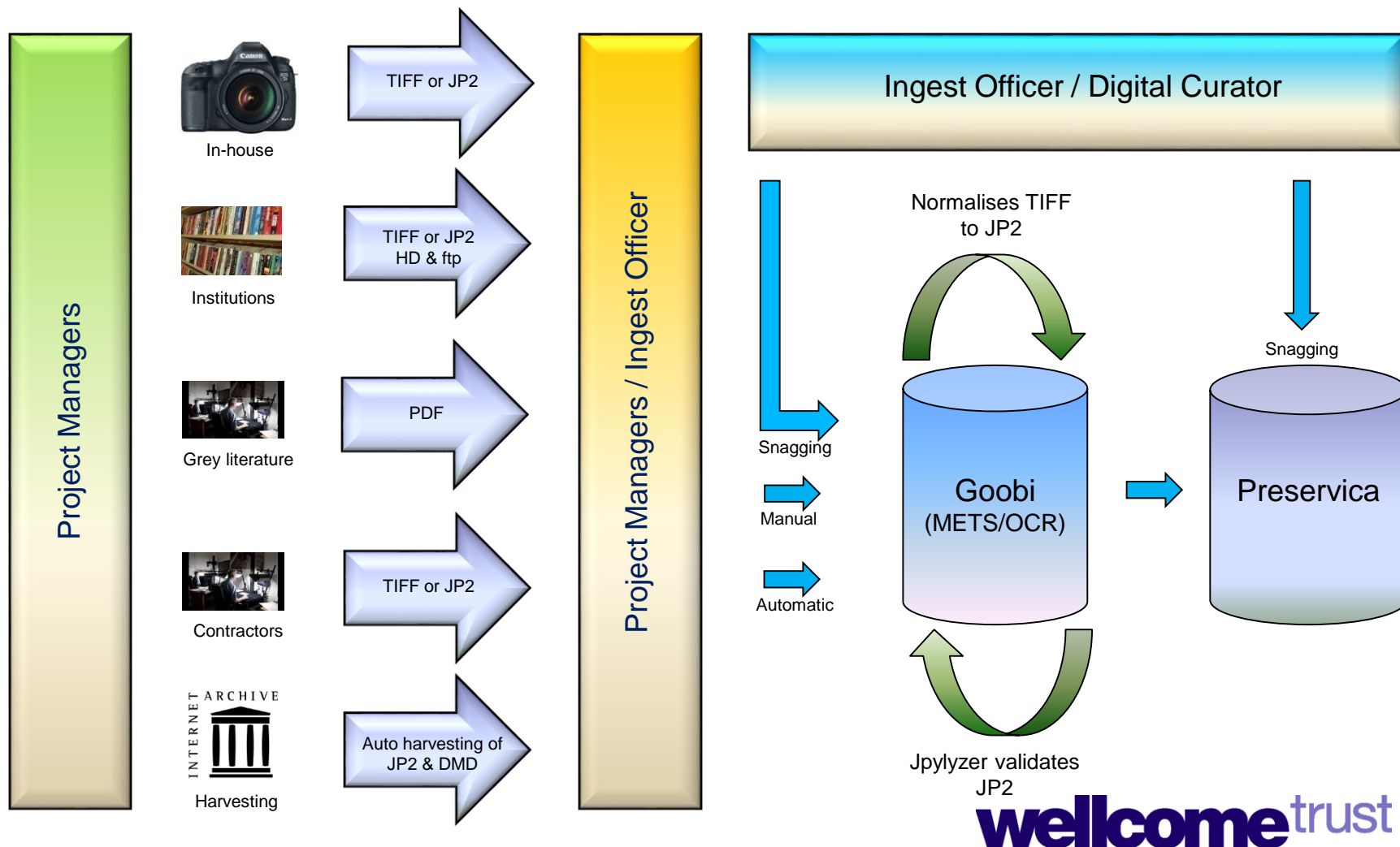


WOOHOO!!!

## Wellcome Library architecture – system view



# Sources of digitised content



# Standardised processes

Process title ▼▲		
[-] _wel_b2045692x		
Number	Title	Status
1	Automatic Import	
2	Download from internet archive with ITM	
3	Import data from internet archive	
7	Create derivatives	
10	Alto conversion	
12	Automatic METS edition	
50	SDB ingest	
60	Export to DDS	
80	PDF creation	
90	30 day delay	
100	Image removal	
150	Re-run METS editing	

Goobi.

Step Progress					
State	Name	Progress	Started	Finished	Messages
✓	Create Collection		14.11.14 01:21:37	14.11.14 01:21:38	
✓	Create XIP		14.11.14 01:21:38	14.11.14 01:22:56	
✓	Fixity Check		14.11.14 01:22:56	14.11.14 01:23:14	
✓	Content Integrity		14.11.14 01:22:57	14.11.14 01:23:02	
✓	Metadata Integrity		14.11.14 01:22:57	14.11.14 01:23:02	
✓	Characterise		14.11.14 01:23:14	14.11.14 01:27:20	<a href="#">View</a>
✓	Store Files		14.11.14 01:27:21	14.11.14 01:28:47	
✓	Store Metadata		14.11.14 01:28:47	14.11.14 01:32:26	
✓	Update Search Index		14.11.14 01:32:26	14.11.14 01:54:32	
✓	Create AMD		14.11.14 01:54:32	14.11.14 01:54:41	

Preservicea.

# What's missing?

- System/workflow documentation.
- Image processing workflow documentation.
- Clear alignment with organisational intent.
- Understanding the environment we have built.

# So, to wrap up...

- Very early days.
- Starting by understanding our content.
- Actually know more than we thought.
- Need to better (i.e. more) documentation of environment.



# Thank you

Questions now, questions later...?

Dave Thompson, Digital Curator  
Wellcome Library

[d.thompson@wellcome.ac.uk](mailto:d.thompson@wellcome.ac.uk) - @d\_n\_t

<http://wellcomelibrary.org/>

**wellcome**trust

# Turning Preservation Planning into Action

Chris Fryer | Senior Digital Archivist | Parliamentary Archives

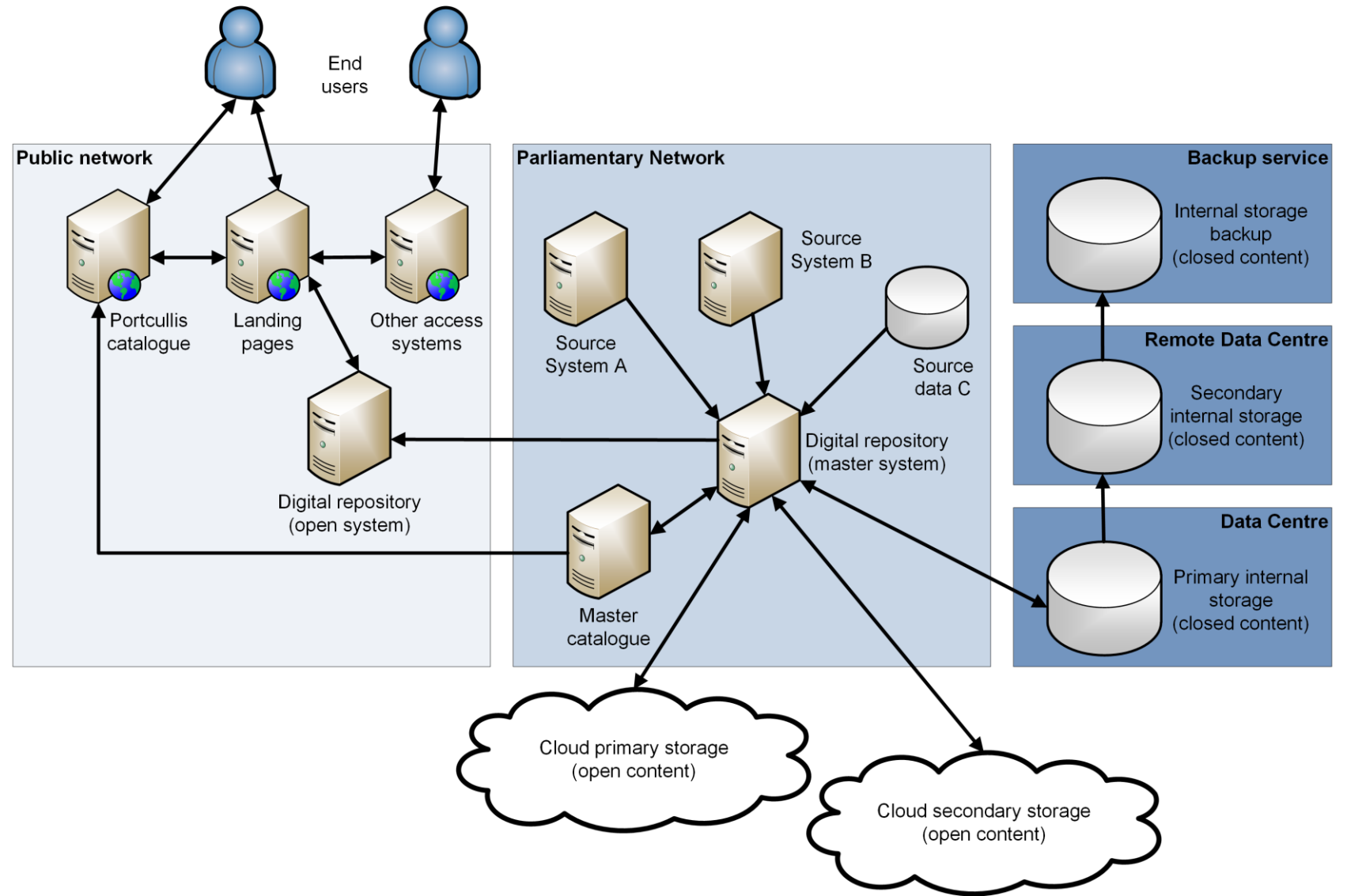


# Context and Goals

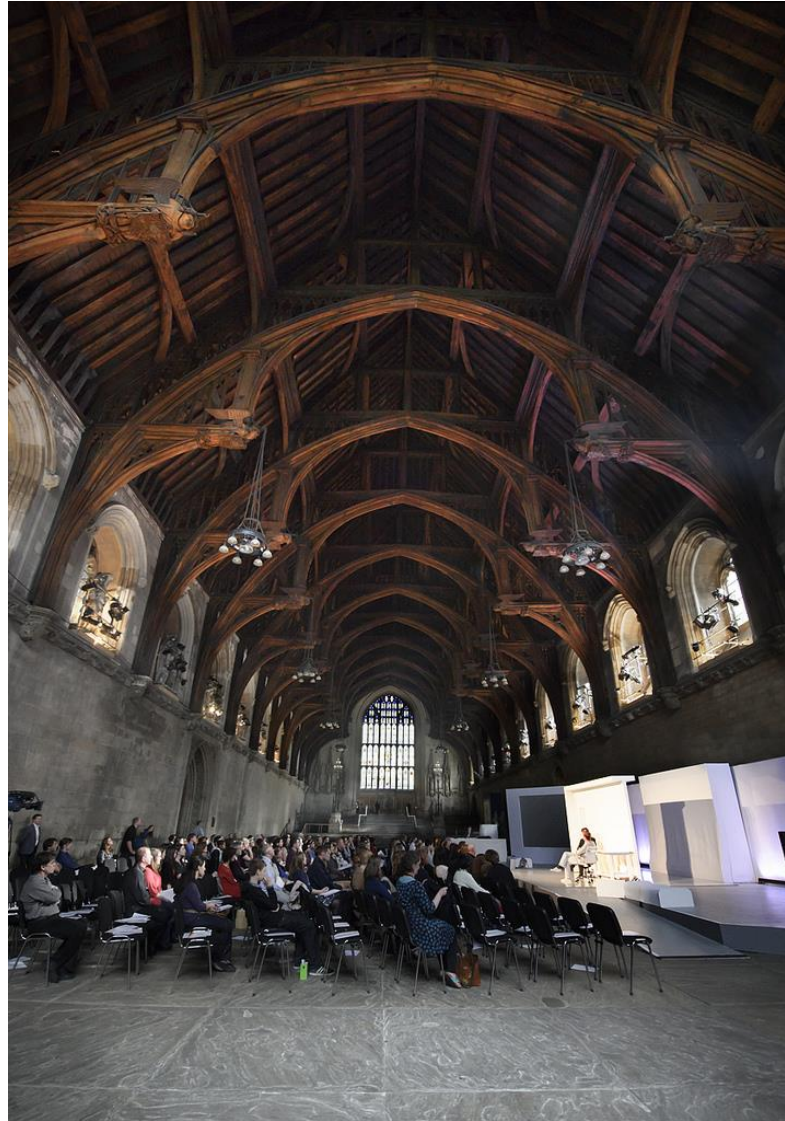




# Context



# Goals



1

- Survey existing best practice

2

- Define framework

3

- Review

4

- Put into action

5

- Evaluate and evolve

# Existing Best Practice



1) British Library:  
Digital Preservation  
Strategy



2) OCLC Research:  
Preservation Health  
Check



3) NLNZ: "Digital  
Preservation in Capable  
Hands"

# Findings

*Digital Preservation Strategy*

## | REASONS |

- Strategy revised and released in 2013
- Contemporary strategy
- Aims differ from Parliamentary Archives
- BL transitioning digital preservation work into BAU

## | RESULTS |

- Evidence led approaches
- Embed PP into corporate risk management
- Implement end to end workflows
- Consider sustainability in all system exercises

BRITISH  
LIBRARY

# Findings

*Preservation Health Check*

## | REASONS |

- New approach using metadata for monitoring
- PREMIS as an evidence base
- Threat assessment based on Simple Property-Oriented Threat (SPOT) model

## | RESULTS |

- Is such an approach is viable?
- Sensors, thresholds, and triggers
- Potential for automated assessment
- Work in progress





# Findings

*NLNZ Article*

## | REASONS |

- Risk assessment at NLNZ
- Challenges assumptions
- Reasoning behind risk assessment approach

## | RESULTS |

- Form of active, managed preservation
- Core ability to render and level of comfort
- Description of the universe of rendering possibilities
- Risk is situational



# Defining a Framework



# Risk Register

## | CORPORATE FRAMEWORK |



## | RISK HEAT MAP |

5) CATASTROPHIC					
4) MAJOR					
3) MODERATE					
2) MINOR					
1) INSIGNIFICANT					
Impact	1) RARE	2) UNLIKELY	3) POSSIBLE	4) LIKELY	5) ALMOST CERTAIN
	Likelihood				



# Outstanding Questions\*

\*lots, but not insolvable



# Going forward...

1) How can we maximise benefit of existing practice?



2) What threats are we trying to mitigate?



3) How can risk based approaches result in planning actions?



4) How can planning actions be quantified in terms of cost?



5) How can we benchmark and improve?



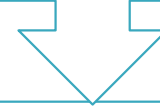
# Enough talk!

Let's get to work...

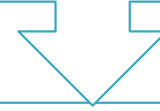


# More Questions...

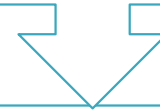
1) How does preservation planning fit in with the bigger picture?



2) How do I know what I'm doing is actually preservation planning?



3) How can I tell if the preservation planning activity is effective?



4) How can others – my organisation – know when I'm doing preservation planning?