



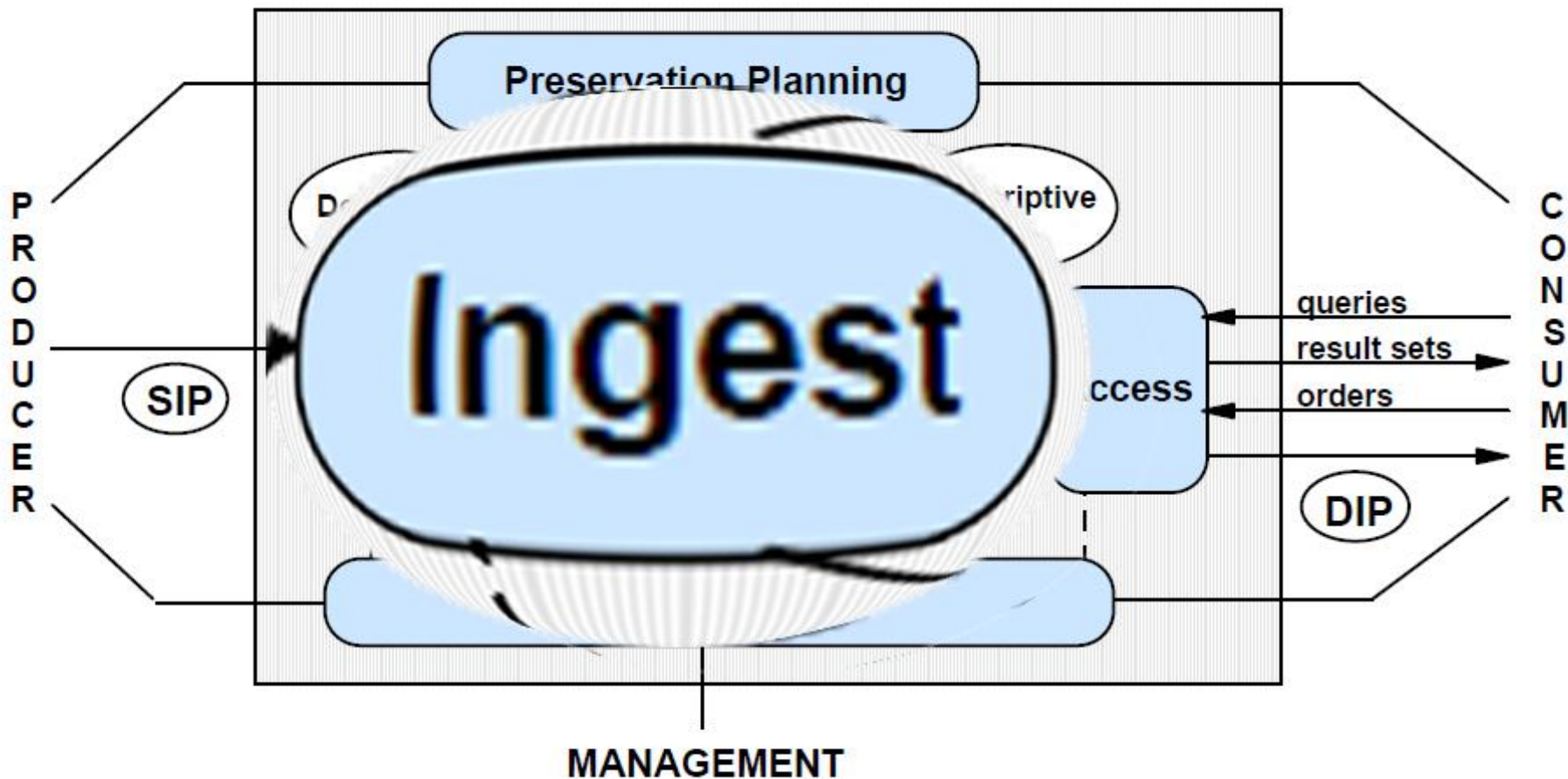
HOUSES OF PARLIAMENT
PARLIAMENTARY ARCHIVES

Making sense of digital collections

**What I wish I knew before I
started...**

Adrian Brown, 24 January 2012

Ingest



HOUSES OF PARLIAMENT
PARLIAMENTARY ARCHIVES

Things I wish I'd known #1

**Ingest accounts for up to 90%
of digital repository activity**



HOUSES OF PARLIAMENT
PARLIAMENTARY ARCHIVES

Things I wish I'd known #2

**Digital information is hugely
(and increasingly) diverse**



HOUSES OF PARLIAMENT
PARLIAMENTARY ARCHIVES

All Mail Folders

- [-] Mailbox - BROWN, Adrian
 - Deleted Items (5)
 - Drafts
 - [+] Inbox
 - Junk E-mail
 - Outbox
 - Sent Items
- [-] Search Folders
 - For Follow Up
 - Large Mail
 - Unread Mail
- [+] Archive Folders
- [-] Mailbox - Archives
 - Drafts [6]
 - Inbox2



23 Apr 2009 : Column 1597

House of Lords

Thursday, 23 April 2009.

11 am

Prayers—read by the Lord Bishop of Southwark.

Health: Tuberculosis

Question

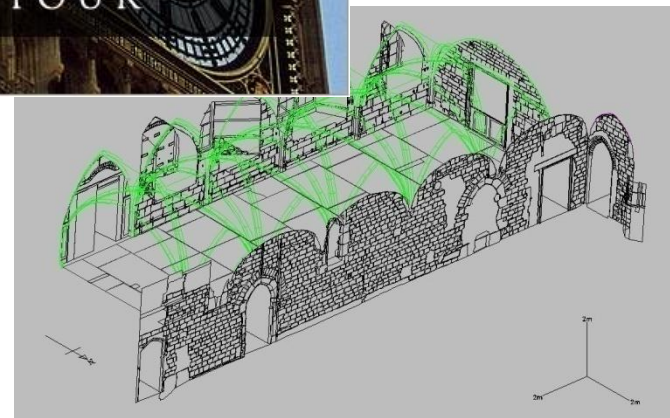
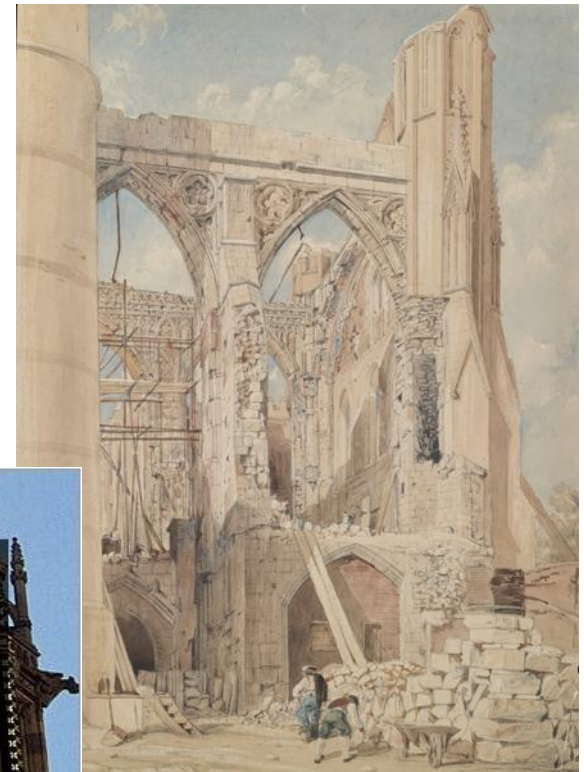
11.06 am

Asked By Baroness Sharples

To ask Her Majesty's Government what measures they are taking to combat the spread of tuberculosis in the United Kingdom.

Baroness Thornton: My Lords, the Government are committed to tackling TB, and in 2004 the Chief Medical Officer for England published an action plan. Our key strategy is early detection and completion of treatment. To implement this, we launched an information pack designed to help NHS commissioners secure best services to meet local needs, and we are funding an awareness-raising programme for the public and professionals to help earlier identification of TB.

Baroness Sharples: My Lords, I thank the noble Baroness for that Answer. Does she agree that one in three of the world's population is infected with latent tuberculosis, that not sufficient funds are available to see



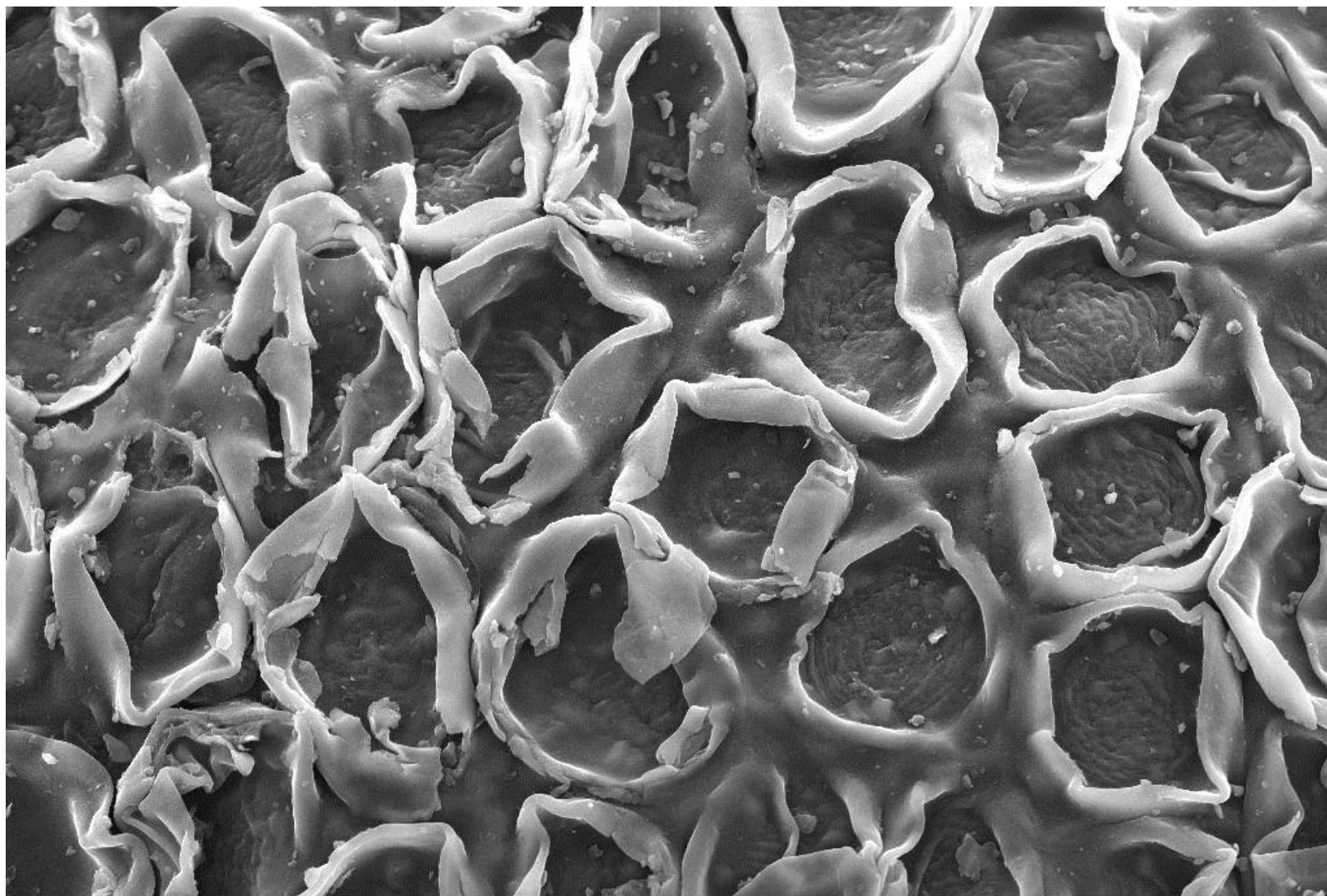
HOUSES OF PARLIAMENT
PARLIAMENTARY ARCHIVES



HOUSES OF PARLIAMENT
PARLIAMENTARY ARCHIVES



HOUSES OF PARLIAMENT
PARLIAMENTARY ARCHIVES



MAG = 421 X

EHT = 25.00 kV

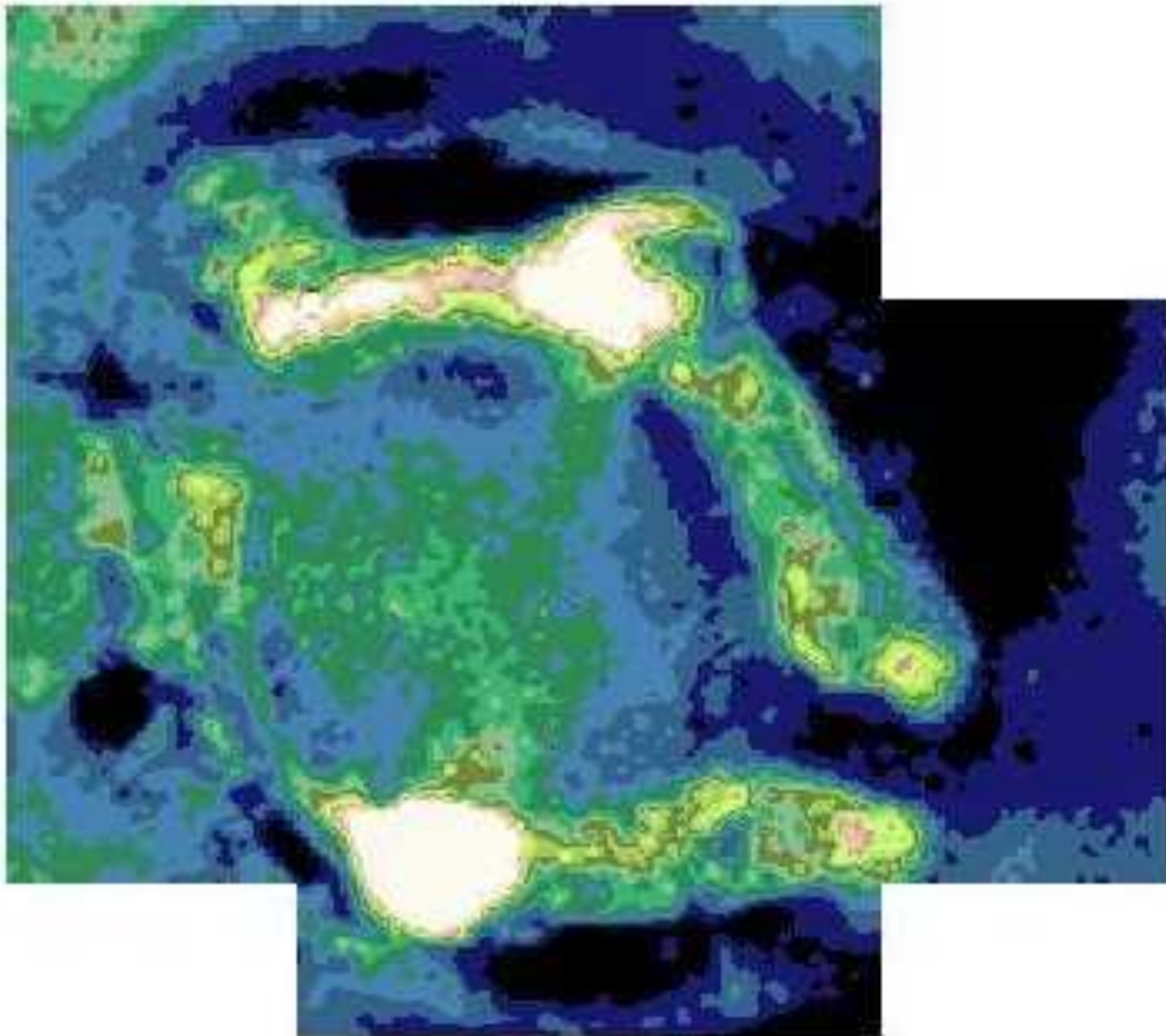
100µm*

Detector = SE1

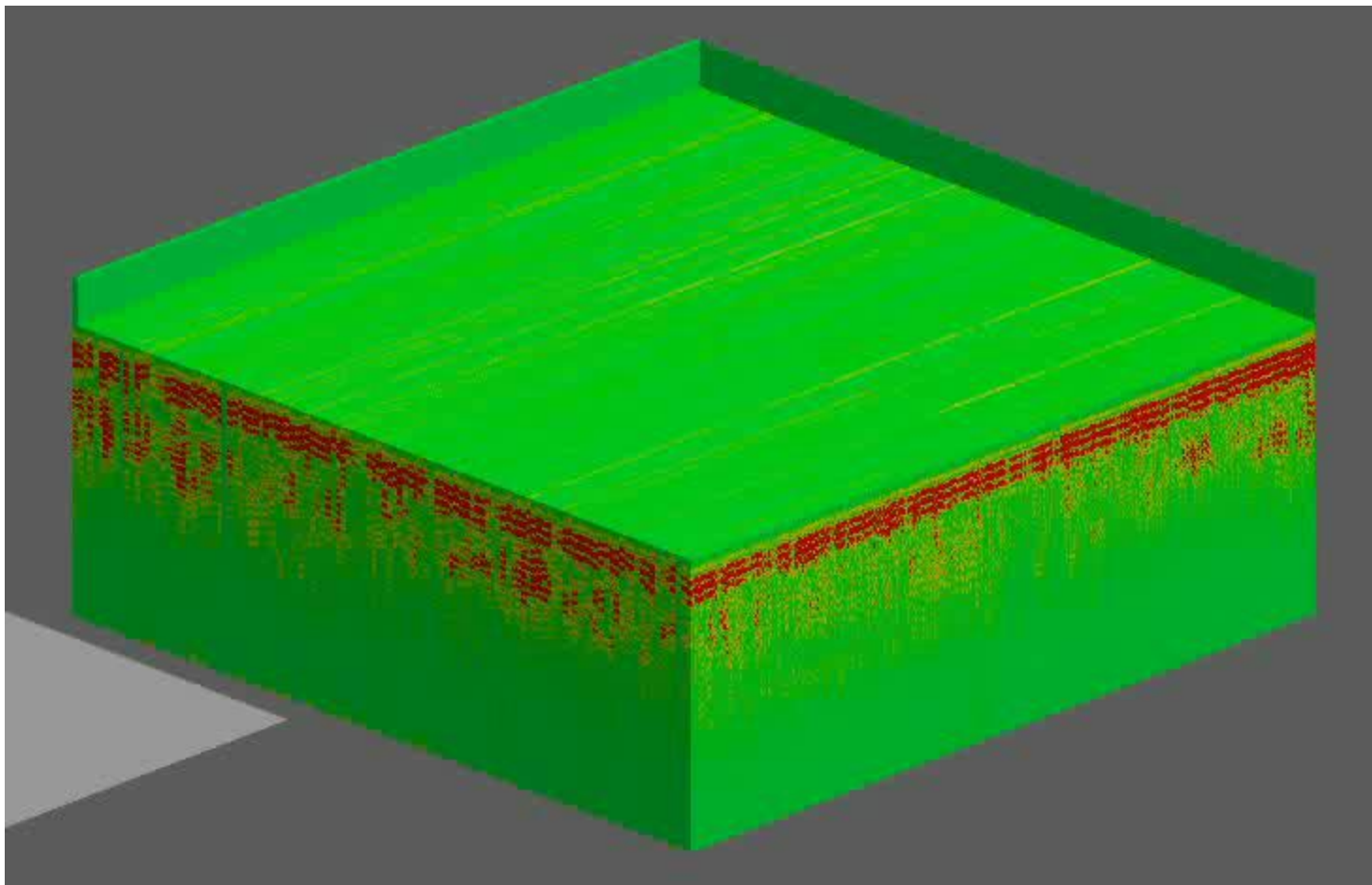
Date :7 Dec 2000



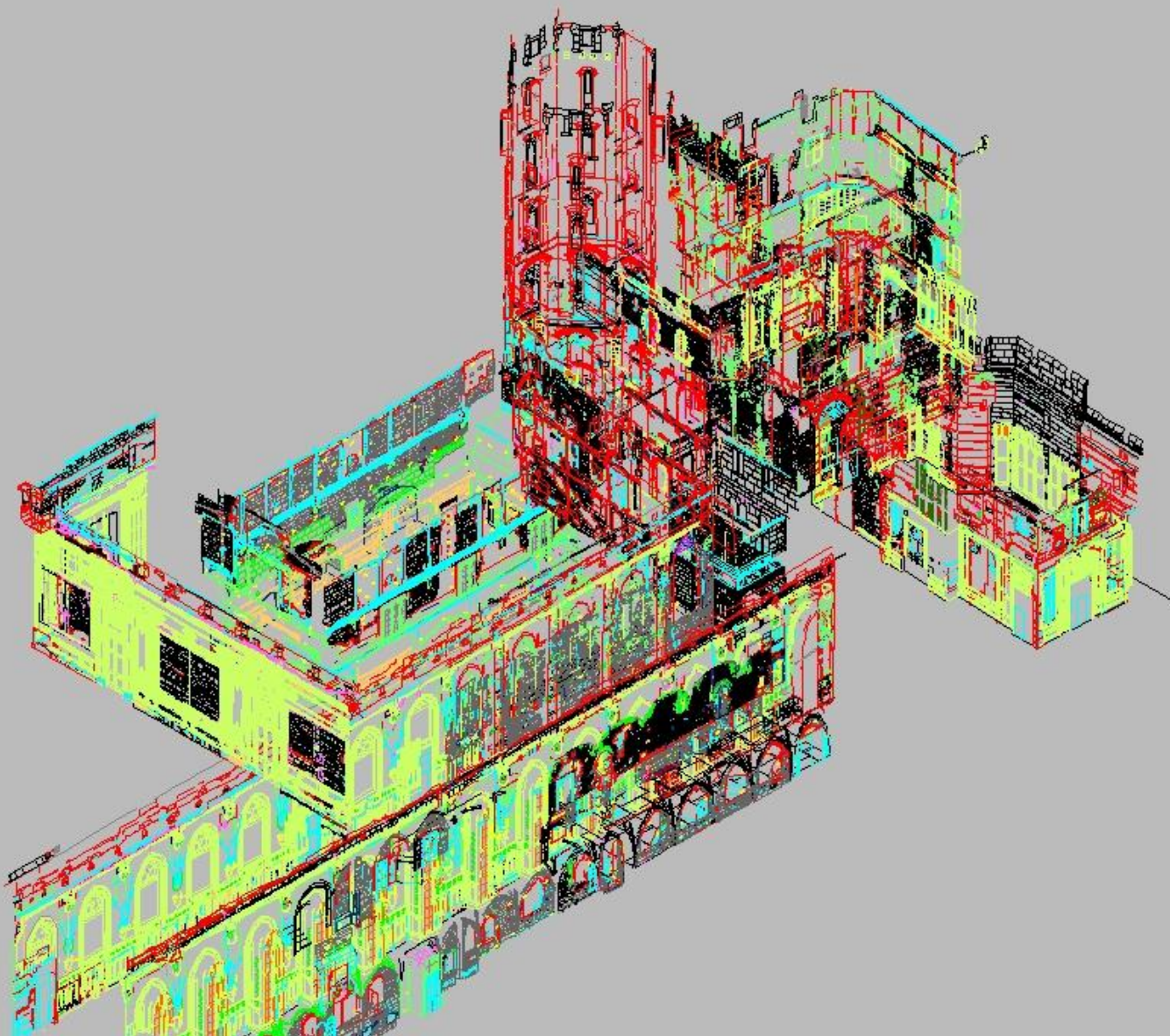
HOUSES OF PARLIAMENT
PARLIAMENTARY ARCHIVES



HOUSES OF PARLIAMENT
PARLIAMENTARY ARCHIVES



HOUSES OF PARLIAMENT
PARLIAMENTARY ARCHIVES



HOUSES OF PARLIAMENT
PARLIAMENTARY ARCHIVES

facebook

Sign Up UK Parliament is on Facebook



UK Parliament

Wall Info RSS/Blog FriendFeed Events

UK Parliament Just Fans

UK Parliament We've redesigned the Parliament website - have a look at the new look at: <http://www.parliament.uk/> - have a look at the new look at: <http://is.gd/mQo> 05 April at 08:41

flickr

Home The Tour Sign Up Explore

UK Parliament Sets

Royal Gallery

Thumbnails Detail Comments



5 photos | 61 views

Items are from between 25 Jul 2008 & 02 Apr 2009.

Feed - Subscribe to the set "Royal Gallery"

You Sign in Create Your Free Account
Explore Places Last 7 Days This Month Popular Tags The Commons Creative Commons Search
Help Community Guidelines The Help Forum FAQ Sitemap Help by Email

Flickr Blog | About Flickr | Terms of Use | Your Privacy | Copyright/MP Policy | Report Abuse

Accessibility A-Z Index Glossary Help Contact Us

www.parliament.uk



Thursday 23rd April 2009

Search

Home Visiting About Parliamentary Business MPs, Lords and Offices Get Involved Topical Issues Education

Welcome to the United Kingdom Parliament

Features on the Parliament website



Autumn school visits to Parliament

Call to book a school visit to Parliament for the autumn term 2009 - phone lines will open at 9am on 29 April



UKParliament

Reading transcript of this morning's questions to Ministers on energy and climate change. Today in the Commons: <http://is.gd/u5XJ>

About 1 hour ago from web

Watching Daily Mail editor: Paul Dacre, give evidence to Cmtee

Parliamentary Business



What's On Parliamentary News Topical Issues Bills and Legislation

Latest news from the House of Commons and House of Lords



Local Democracy Bill: report stage
Local government and development - Lords completes

YouTube

English
Videos | Channels | Community | Upload



Videos | Playlists | Groups | Subscriptions

UK Parliament

Subscribe



UKParliament
Style: Broadcaster
Joined: 17 May 2007
Last Sign In: 3 days ago
Subscribers: 385
Channel Views: 58,919

This is a pilot YouTube channel for the United Kingdom Parliament.

Website: <http://www.parliament.uk/>
Live and archive footage: <http://www.parliamentlive.tv/>
Twitter: <http://twitter.com/ukparlia...>
Facebook: <http://www.facebook.com/pag...>
Flickr: <http://www.flickr.com/photos...>
FriendFeed: <http://friendfeed.com/ukpar...>
Blog: <http://iparliamentlabs.wordpress...>
Easy YouTube player: <http://tiny.cc/ukeasy-you...>



Hey there! UKParliament is using Twitter.

Twitter is a free service that lets you keep in touch with people through an exchange of quick, frequent answers to one simple question: What are you doing? Join today to start receiving UKParliament's updates.

Join today!

Already using Twitter from your phone? [Click here.](#)

Home UK Parliament
Location Westminster, London
Web <http://www.parliament.uk/>
Bio Keeping an eye on government, debating laws, approving laws. This official Twitter feed is produced by the web team in Parliament.

5,604 5,741
Following followers

Updates 906

Favorites



UK Parliament Labs

Production blog of the UK Parliament Web Centre

Subscribe to feed Home About this blog

Search

Google

Search the Parliament website with our Google search gadget XML

EMAIL ALERTS - SUBSCRIBE!

Subscribe to UK Parliament Labs by email

Subscribe to all UK Parliament's RSS feeds by email

UK PARLIAMENT TWITTER

UKParliament: Reading

Parliament Tweetminster widget

17/04/2009 in Social-networking, twitter | Tags: tweetminster, iframe, widget | by Alistair Reid | No comments

We worked with the good folks from Tweetminster recently to put together a custom iFrame widget that we could use on the Parliament website.

This involved styling the widget Tweetminster already make available so that it fitted with the look of our new redesign, and specifying what Twitter feeds it would use. We wanted the widget to be "Parliament on Twitter" - so it uses our own @ukparliament feed, any MPs or Peers that use Twitter (are there any Peers using Twitter yet?), but no parliamentary candidates, government departments or political parties.

The hope is that using this widget will further promote how Parliament and its Members are trying to engage with social media tools and with the public. During recess it should also give an insight into the work of MPs in their constituencies. Although, we have included a disclaimer because the Web Centre team is responsible only for the content of @ukparliament updates, not



HOUSES OF PARLIAMENT
PARLIAMENTARY ARCHIVES



PARLIAMMENTARY ARCHIVES

[Shop All Departments](#) Search

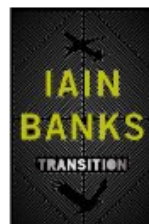
GO




Basket

[Wish List](#)[Adrian's Amazon.co.uk](#)[Page You Made](#)[Recommended For You](#)[Rate These Items](#)[Improve Your Recommendations](#)[Your Profile](#)[Learn More](#)[Adrian's Amazon.co.uk](#) > **Recommended for you**(If you're not Adrian Brown, [click here](#).)**Just For Today**[Browse Recommended](#)**Recommendations**[Baby](#)[Books](#)[DIY & Tools](#)[DVD](#)[Electronics & Computing](#)[Garden & Outdoors](#)[Health & Beauty](#)[Home & Garden](#)[Jewellery](#)[Lighting](#)[MP3 Downloads](#)[Music](#)[PC & Video Games](#)[Shoes & Accessories](#)[Software](#)These recommendations are based on [items you own](#) and more.view: **All** | [New Releases](#) | [Coming Soon](#)[More results](#) 

1.

**Transition**

by Iain Banks (Sep 3, 2009)

Average Customer Review:  (21)

In stock


RRP: £18.99

Price: £9.49[30 used & new](#) from £5.22☐ I own it ☐ Not interested  Rate this itemRecommended because you said you owned **Complicity (Abacus Paperback)** and more ([Fix this](#))

2.

**Zima Blue**

by Alastair Reynolds (April 30, 2009)

Average Customer Review:  (6)

In stock

RRP: £18.99

Price: £12.28[24 used & new](#) from £6.96

Using SPARQL With Geographic Data

The SPARQL endpoint for geo data is available at <http://os.services.tso.co.uk/geo/sparql>.

Clear Text

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT * WHERE {
  ?s ?p ?o
}
LIMIT 10

```

- ☒ Plain Text
☐ SPARQL-XML
☐ JSON

Sample Queries

- This query returns a list and names of Westminster Constituencies, and their containing county electoral divisions.

```

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX admingeo: <http://data.ordnancesurvey.co.uk/ontology/admingeo/>
PREFIX spatialrelations: <http://data.ordnancesurvey.co.uk/ontology/spatialrelations/>

```

```

SELECT ?constituency ?constituencyname ?cddivision ?name
WHERE
{
  ?constituency rdf:type admingeo:WestminsterConstituency .
  ?constituency foaf:name ?constituencyname .
  ?cddivision rdf:type admingeo:CountyElectoralDivision .
  ?cddivision foaf:name ?name .
  ?constituency spatialrelations:contains ?cddivision .
} LIMIT 10

```

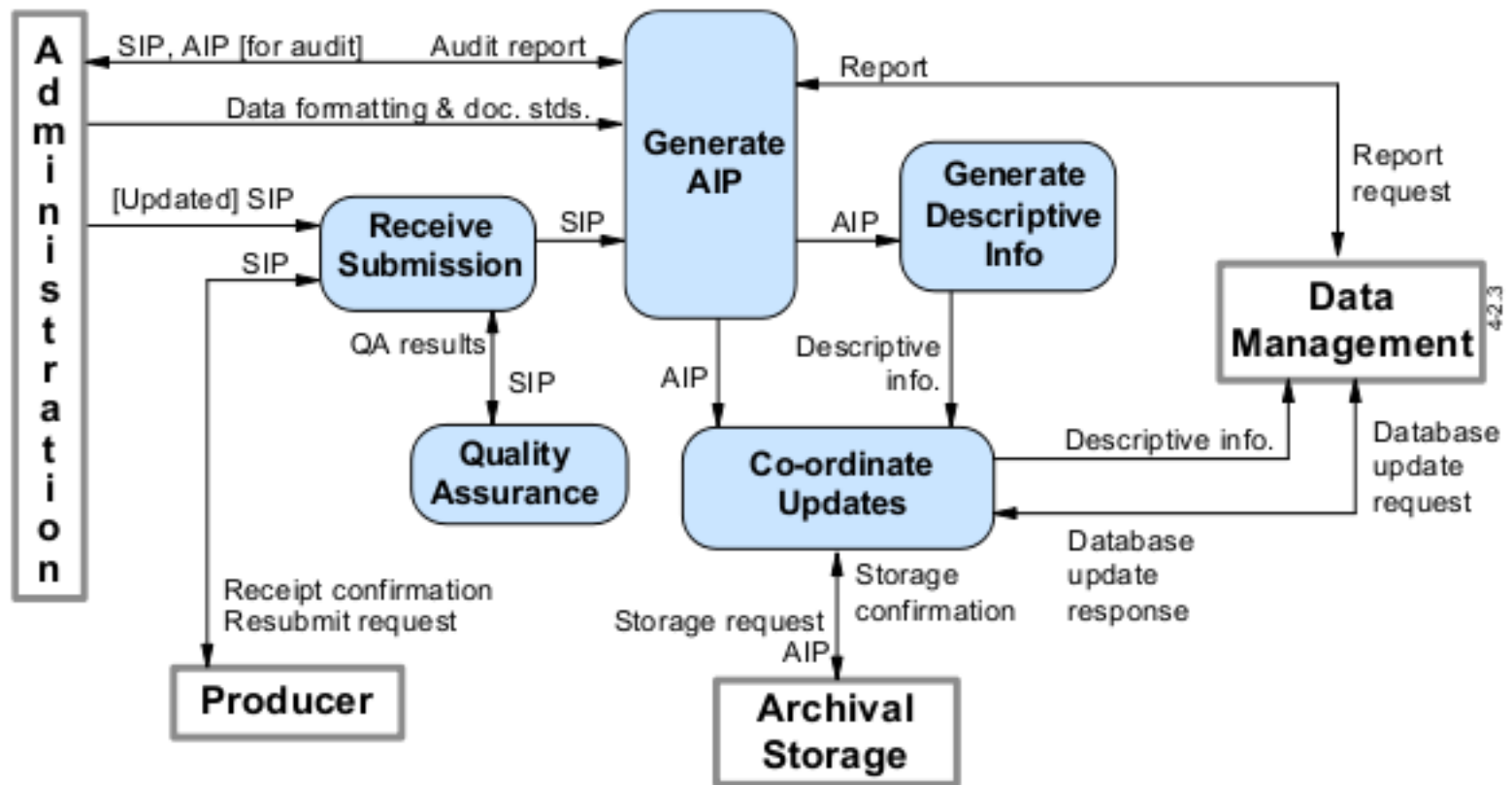


Things I wish I'd known #3

**You rarely get nice, standard,
complete SIPs**

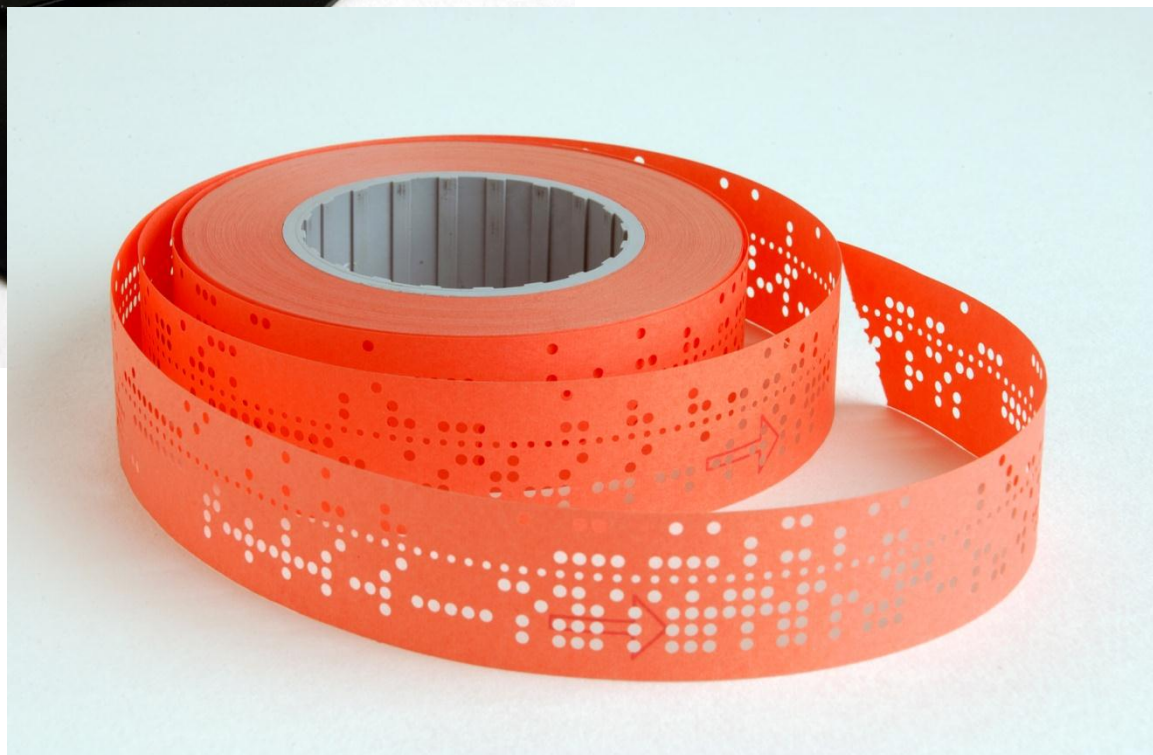


HOUSES OF PARLIAMENT
PARLIAMENTARY ARCHIVES





HOUSES OF PARLIAMENT
PARLIAMENTARY ARCHIVES



HOUSES OF PARLIAMENT
PARLIAMENTARY ARCHIVES

2008 in SPIRE on SPIRE_PRODUCTION - Microsoft Outlook

File Edit View Go Tools Actions SPIRE Help

Send/Receive Search address books

Mail

Favorite Folders

Mail Folders

All Mail Items

- BOARDS AND GROUPS
- COMPLIANCE
- CONTROL
- DISPOSAL
- INTELLECTUAL PROPERTY
- MONITORING
- PLANNING AND STRATEGY
- POLICY MAKING
- PROCEDURES AND GUIDANCE
- PROGRAMME MANAGEMENT
- PROJECT MANAGEMENT
- <PROJECT NAME>
- HL - DIS - PARLIAMENTARY
- ARCHIVES IT PORTFOLIO
- BFI AUDIO-VISUAL CON
- CATALOGUE APPLICATION
- CLASSIFICATION SCHEM
- CLOSED RECORDS ON C
- DESIGNATIONS
- DIGITAL PRESERVATION
- PROJECT BOARD
 - 2008
 - 2009
 - 2010
 - 2011
 - PROJECT CLOSURE
 - PROJECT DELIVERY
 - PROJECT DOCUMENT
 - PROJECT REVIEW
 - PROJECT START UP
 - DIGITISATION
 - SHARED DRIVE DETOX
 - UNESCO MEMORY OF THE WORL
 - VITAL RECORDS PROGRAMME
 - WAR MEMORIALS
 - REPORTING
 - RISK MANAGEMENT
 - SECURITY
 - INFORMATION SYSTEMS AND TECHNOLOG
 - LEGAL SERVICES

2008-2

Fri 25/11/2011 16:52

N, Adrian

2008-1

Fri 25/11/2011 16:52

N, Adrian

2008-3

Fri 25/11/2011 16:52

N, Adrian

2008-3

Fri 25/11/2011 16:52

n Board Minutes 2008-1

N, Adrian

2008-3

Fri 25/11/2011 16:52

n Board Minutes 2008-2

N, Adrian

2008-3

Fri 25/11/2011 16:52

n Board Minutes 2008-3

N, Adrian

Agenda Meeting 2008- 2.doc

DIGITAL PRESERVATION
PROJECT BOARD

2nd Meeting

30 Sept 2008, 11am

HL Director of Information Services' Office

AGENDA

1. Apologies
2. Minutes of Last Meeting
3. Highlight Report
4. Communications Plan – for approval
5. AOB
6. Date of Next Meeting – 15th December 2008, 2.30pm

To-Do Bar

November 2011

Mo	Tu	We	Th	Fr	Sa	Su
31	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30	1	2	3	4
5	6	7	8	9	10	11

UCL slides
15:30 - 17:00

Divert phone
Wed 9:30 - 10:00

Mid-year review - LA
Wed 10:00 - 11:00
Rm 21; BROWN, Adrian

Arranged By: Due Date

Type a new task

- Update S&A...
- Plan update...
- Digitisation ...
- Digitisation ...
- Parliamentar...
- Volunteers
- a new worki...
- PQAF Digital...
- AGIMO issu...
- Fwd:
- Digitisation ...
- Mapped Re...
- Workflows ...
- In Summary
- RE: notes an...
- SDB USM pr...
- SDB USM pr...
- *NEW* Yoga...
- Digitisation ...
- Command a...

This Week

- PICTs Cloud ...
- Digitisation ...
- Notes actio...

6 Items

Office Communicator DRAFT revised Auth... 2008 in SPIRE on SPI... Hampshire MDO ev... Adrian Brown UCL 2...

15:56



Things I wish I'd known #4a

**Digital workflows are highly
sensitive to variation**



HOUSES OF PARLIAMENT
PARLIAMENTARY ARCHIVES

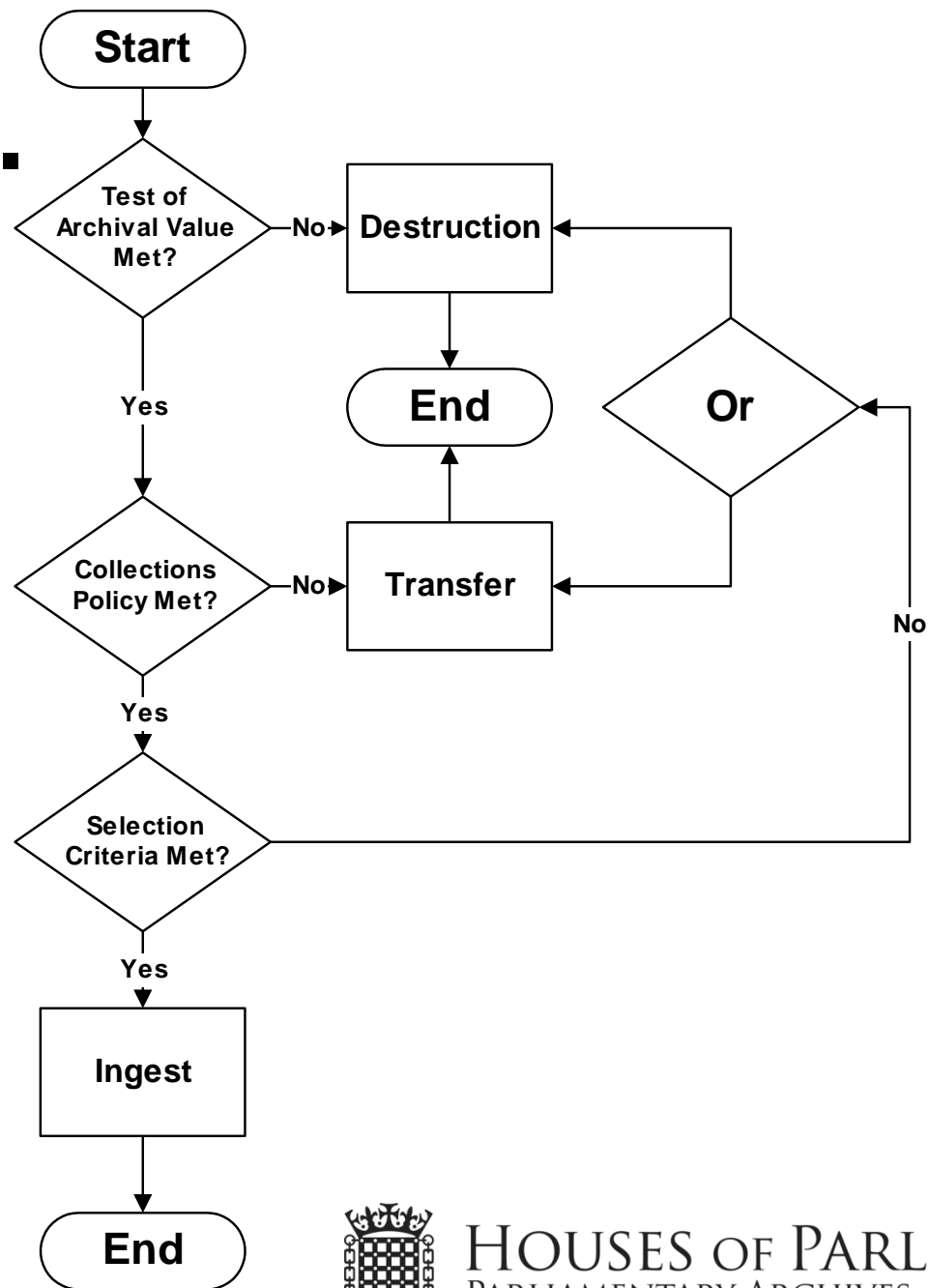
Things I wish I'd known #4b

**QED: There are as many ingest
workflows as there are
archivists to design them**

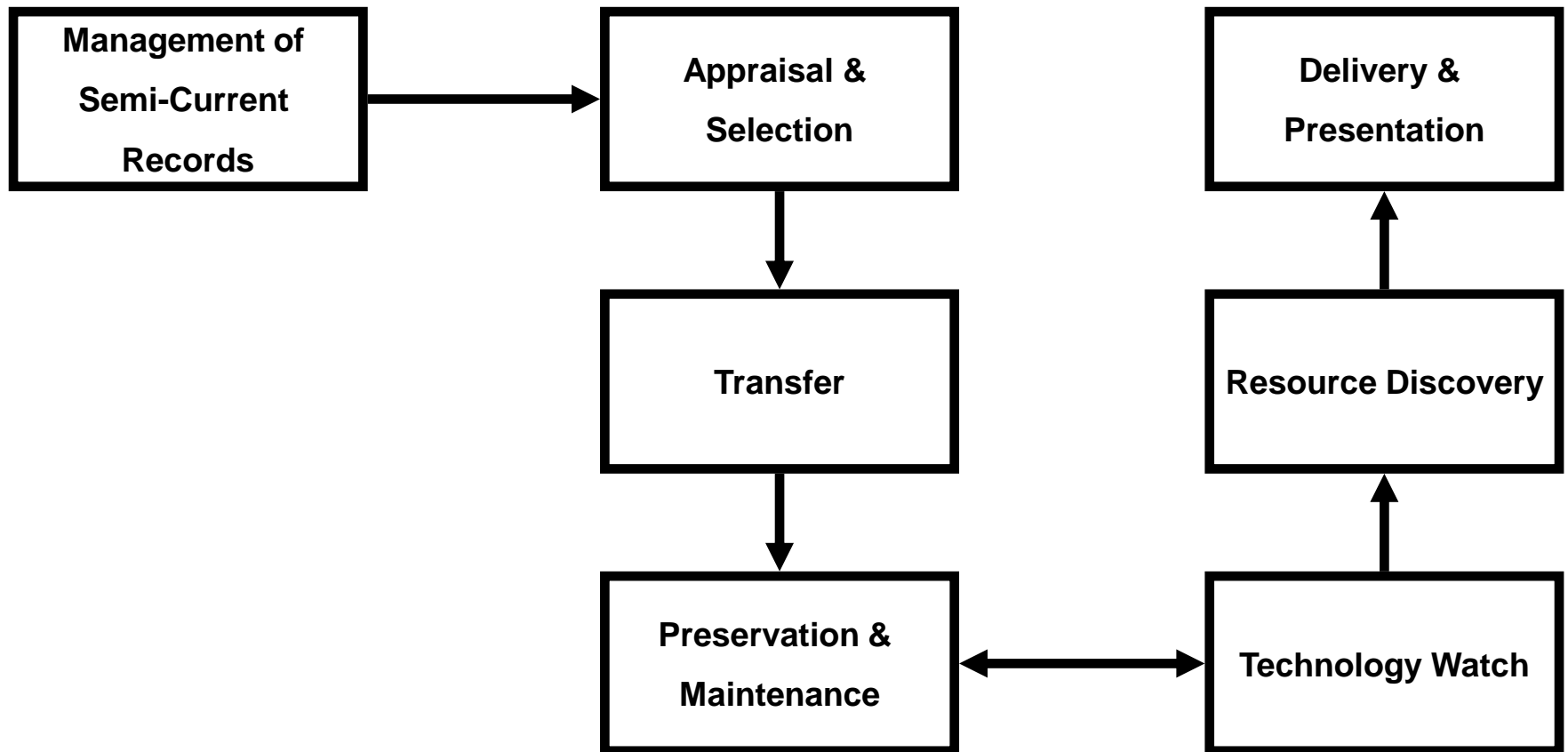


HOUSES OF PARLIAMENT
PARLIAMENTARY ARCHIVES

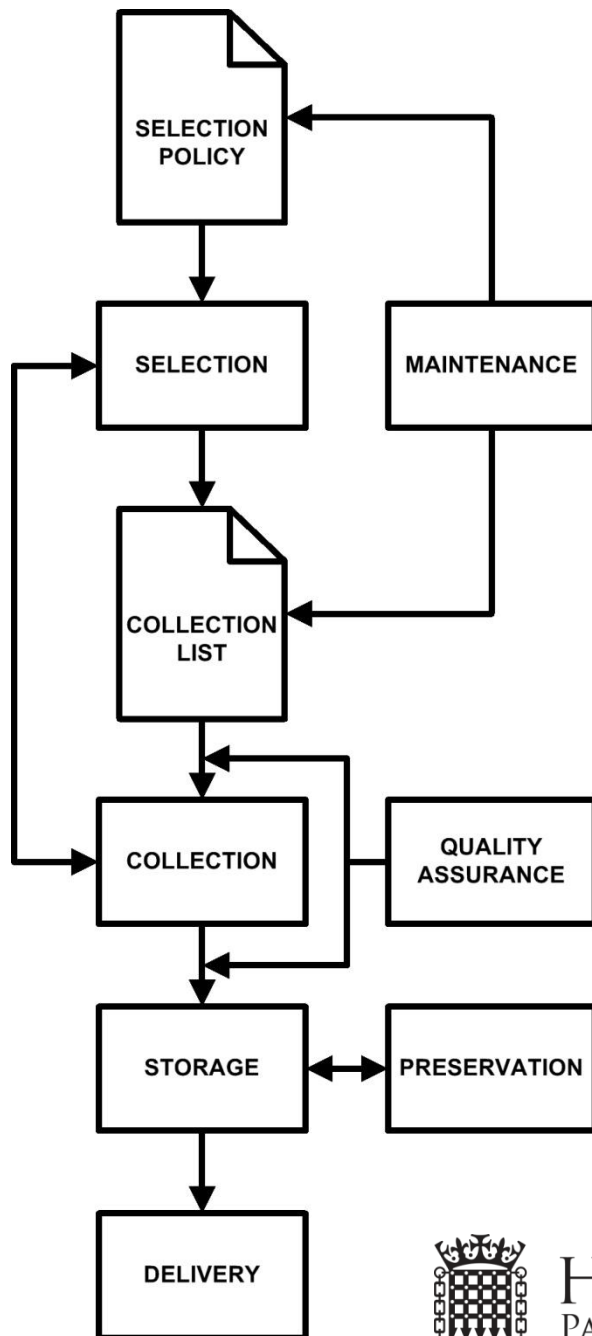
Workflows...



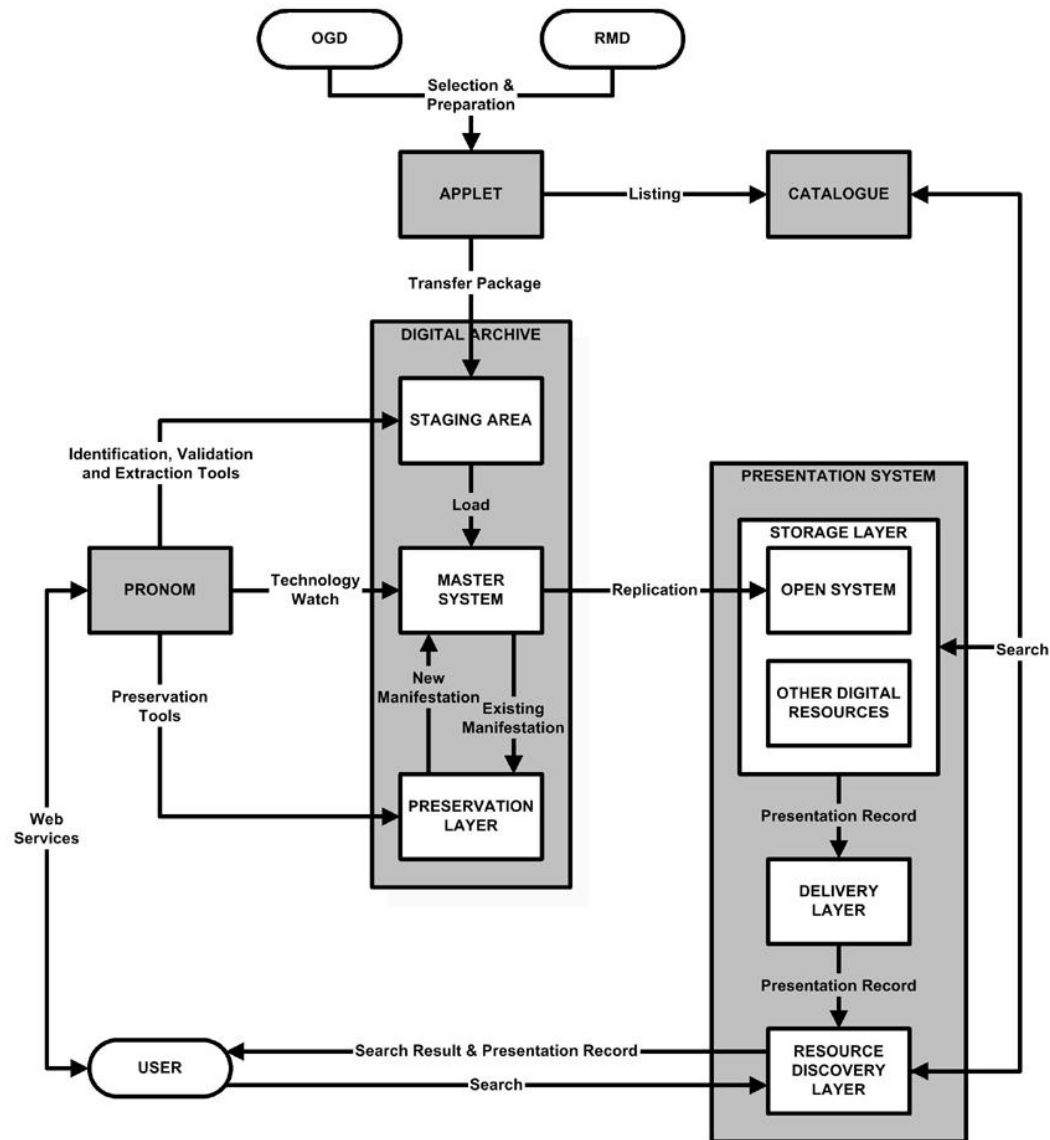
■ ■ ■



■ ■ ■

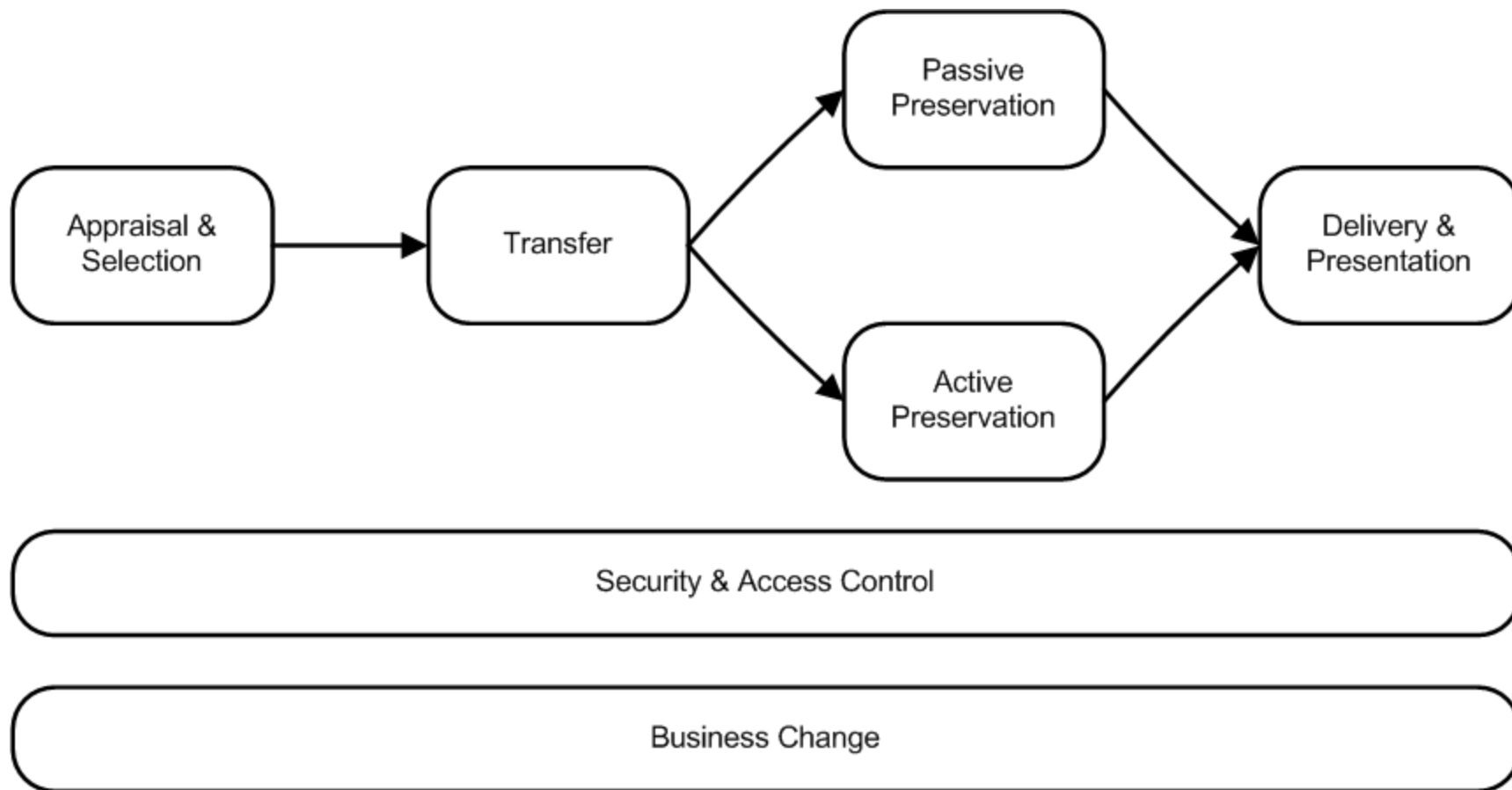


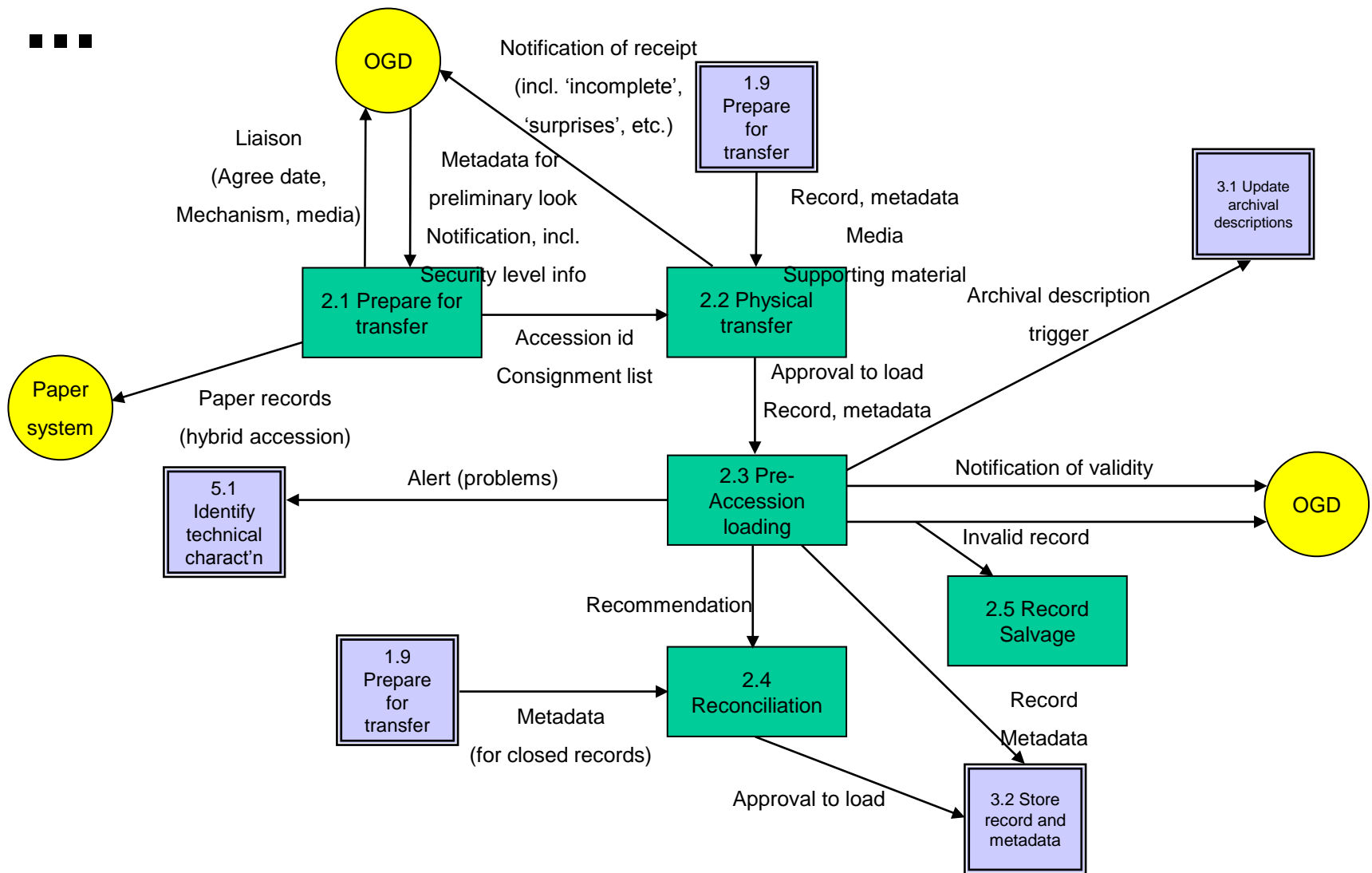
HOUSES OF PARLIAMENT
PARLIAMENTARY ARCHIVES



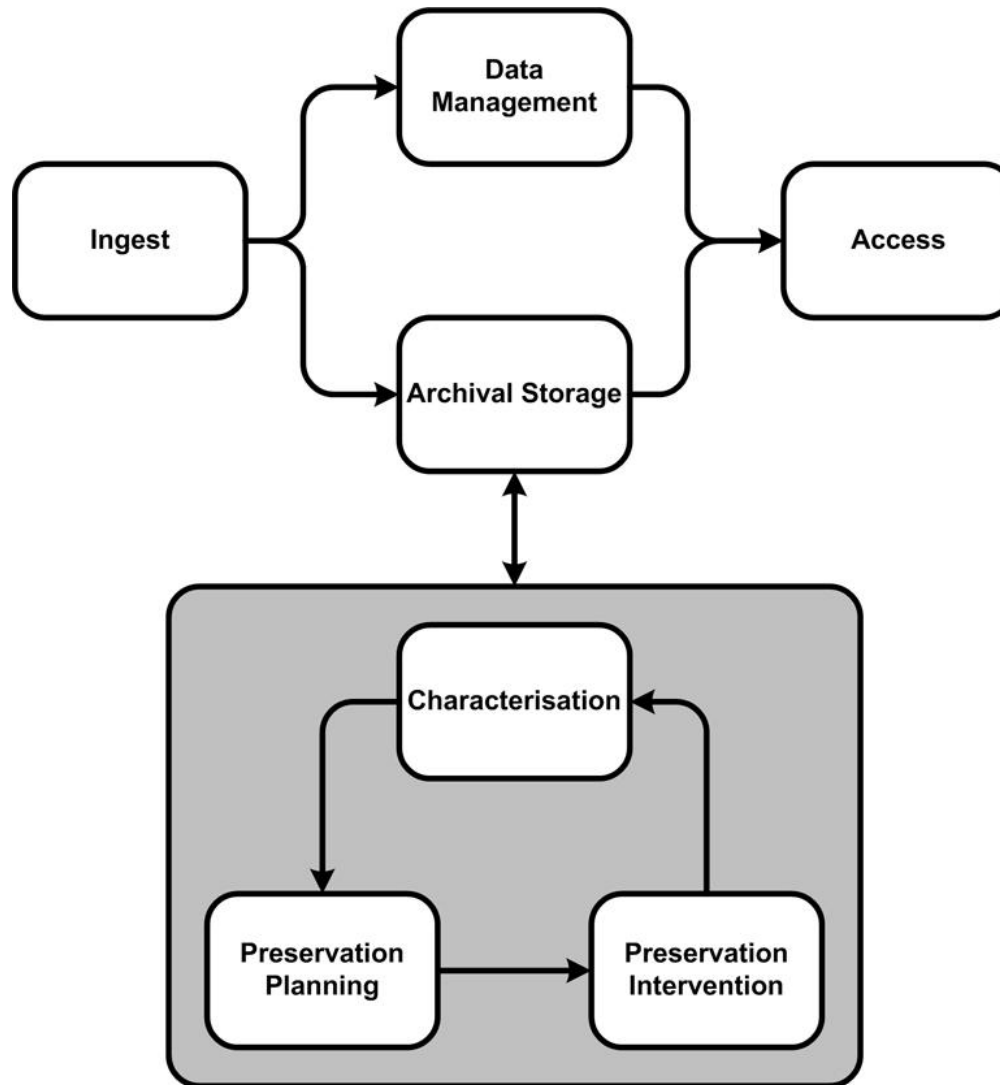
HOUSES OF PARLIAMENT
PARLIAMENTARY ARCHIVES

■ ■ ■

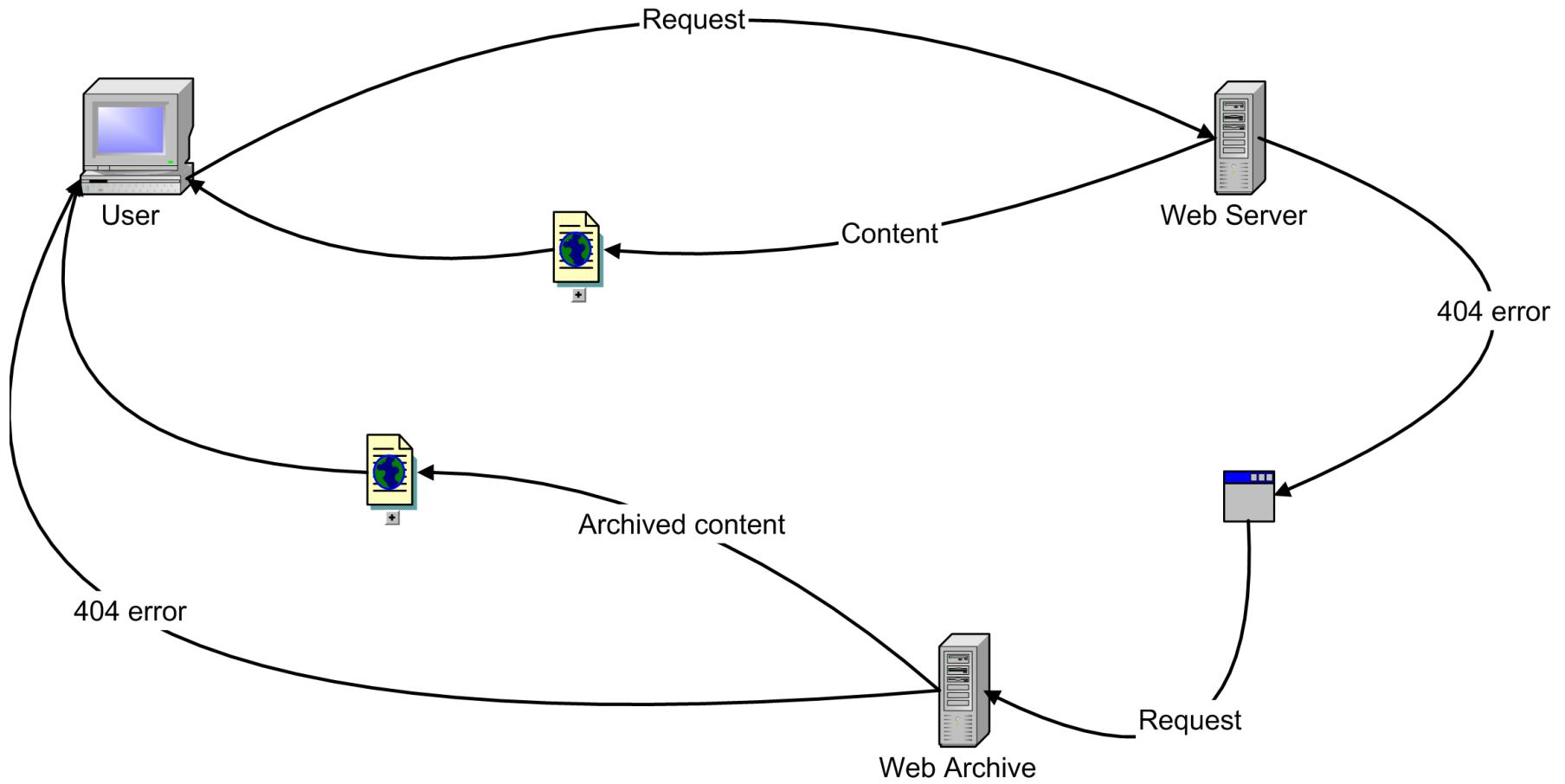




■ ■ ■



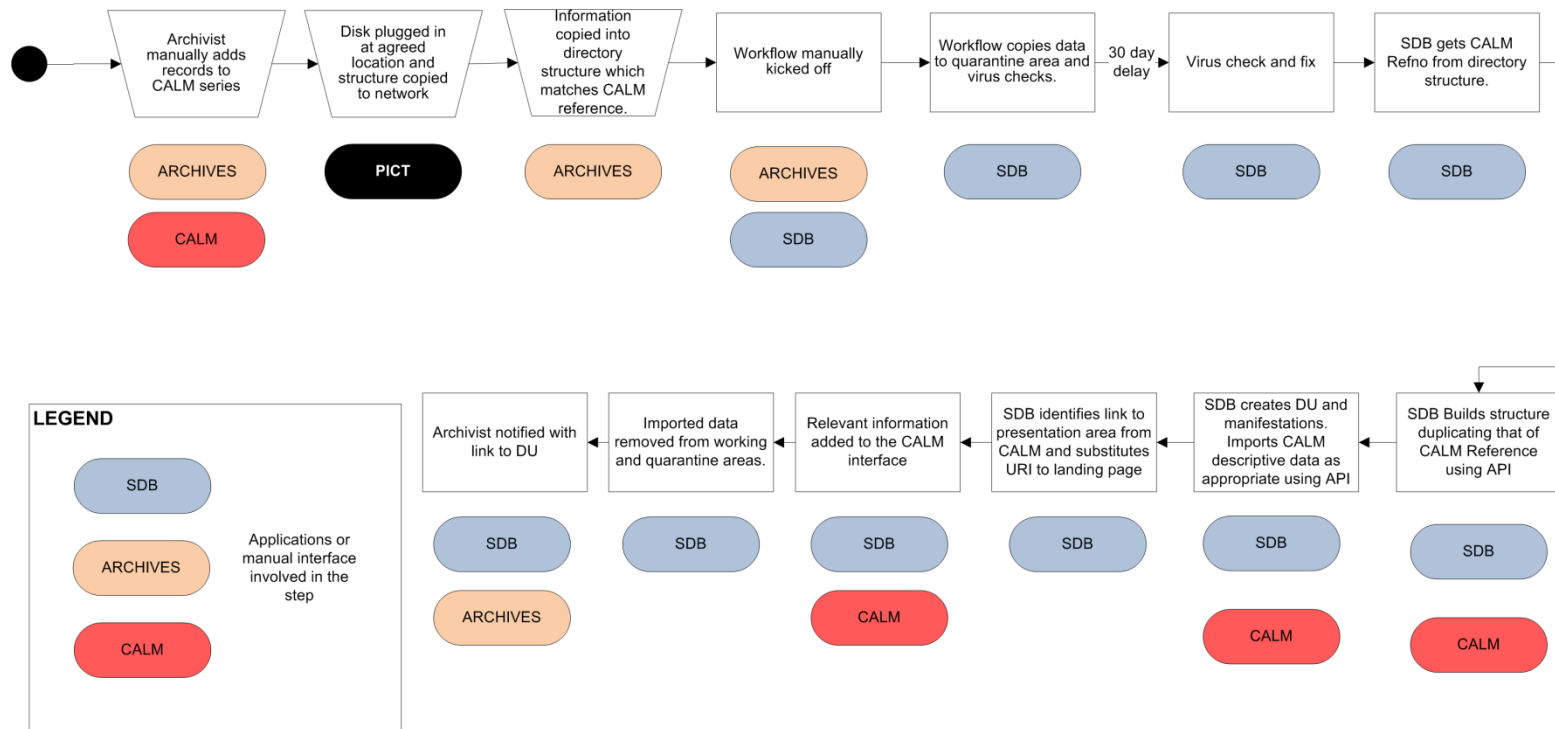
HOUSES OF PARLIAMENT
PARLIAMENTARY ARCHIVES



HOUSES OF PARLIAMENT
PARLIAMENTARY ARCHIVES

...etc...

Ingest Process for Web Archive Data Type

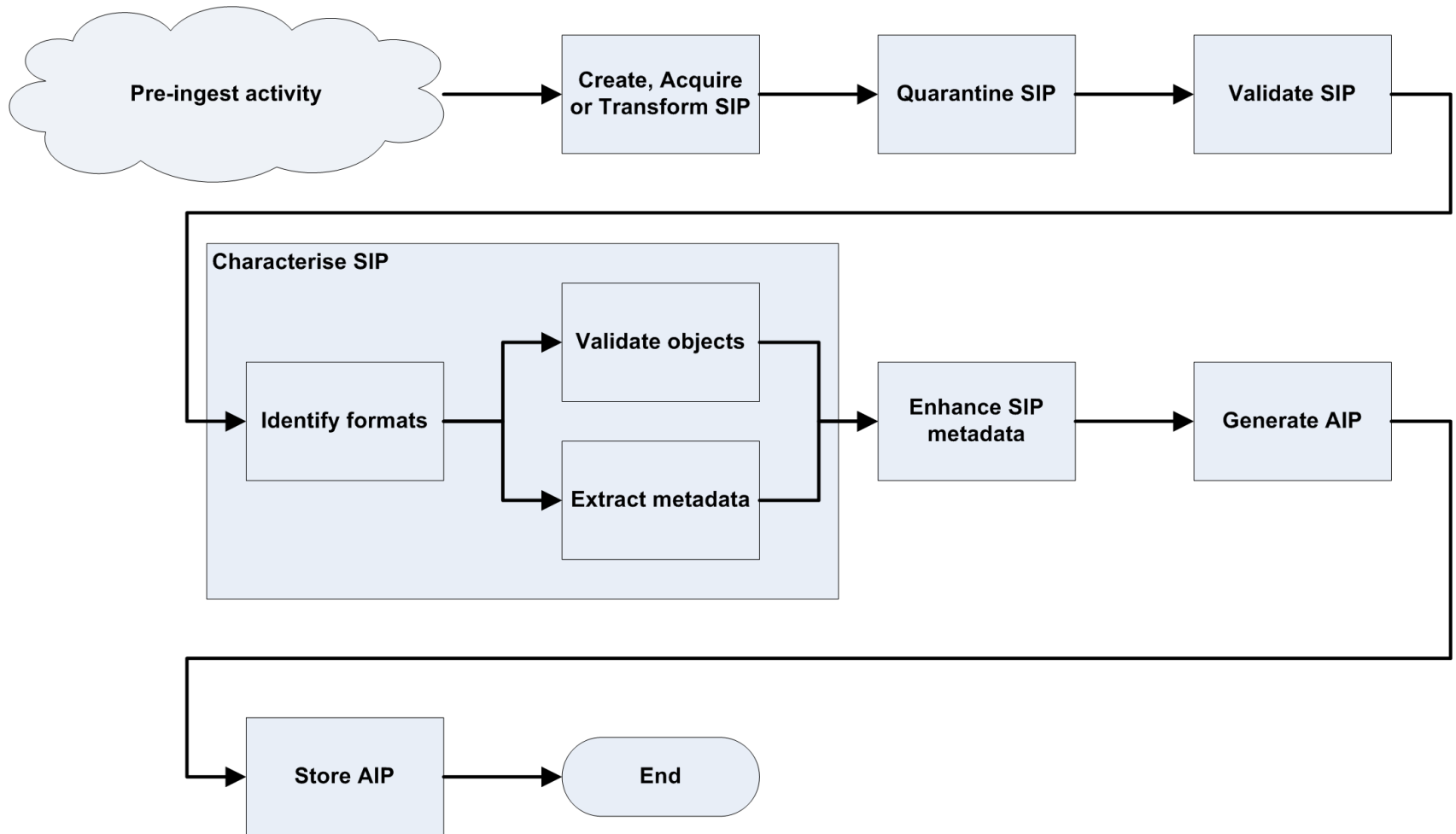


These are all GOOD things!



HOUSES OF PARLIAMENT
PARLIAMENTARY ARCHIVES

Ingest workflow

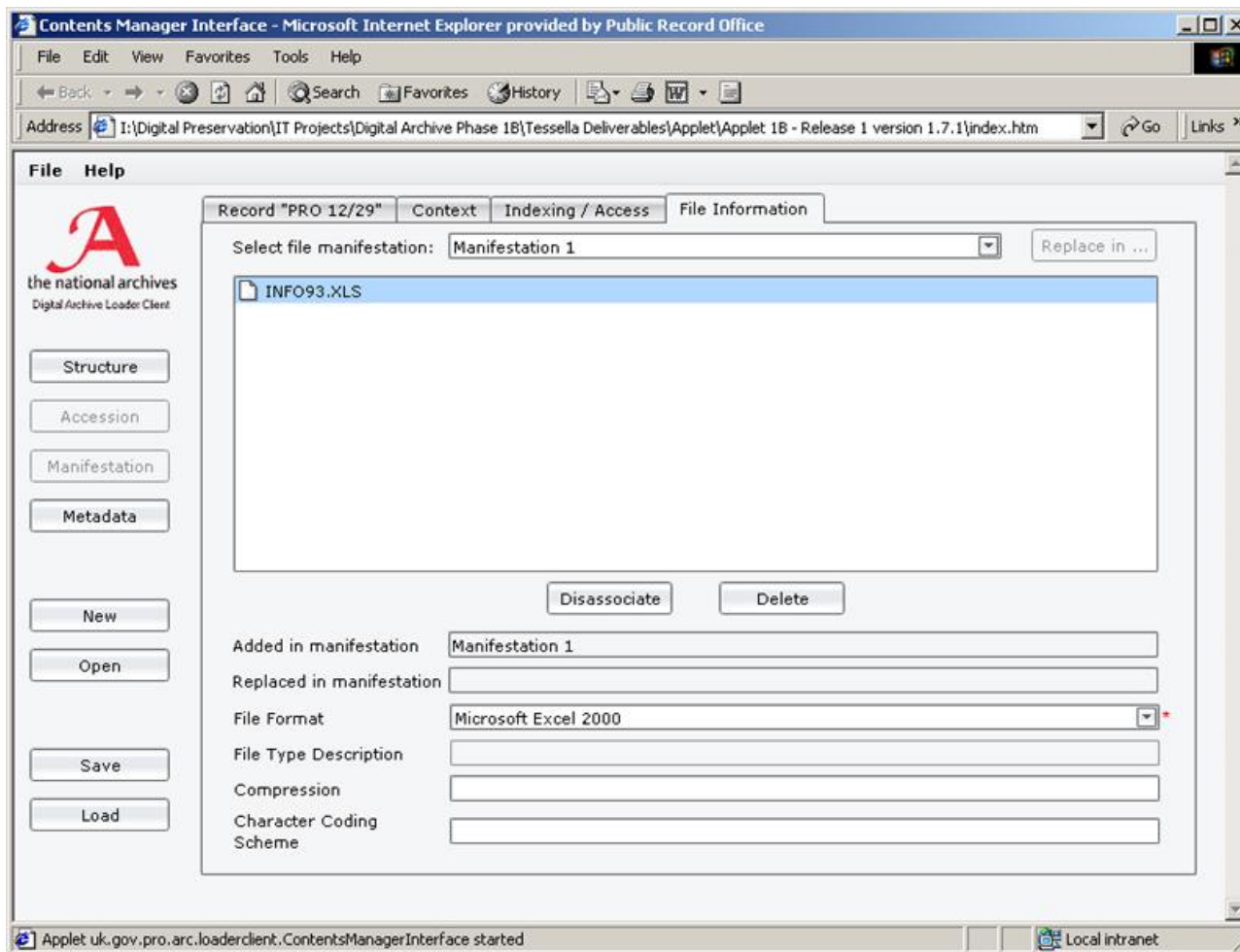


Things I wish I'd known #5a

**We can automate the boring
stuff**



HOUSES OF PARLIAMENT
PARLIAMENTARY ARCHIVES













HOUSES OF PARLIAMENT
PARLIAMENTARY ARCHIVES

[Start](#)
[Waiting](#)
[Running](#)
[Completed](#)
[Reports](#)
[Manage](#)

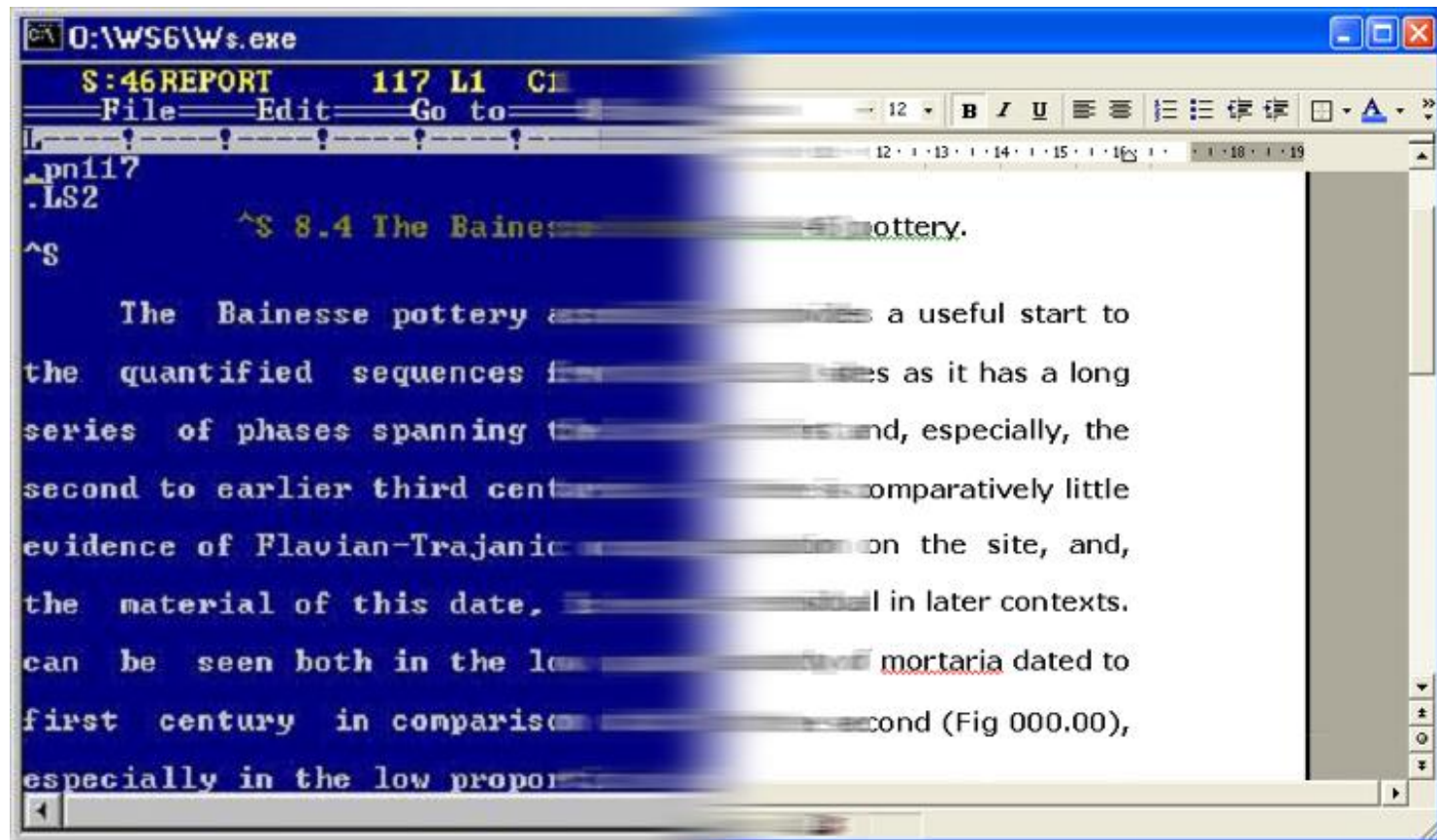
Workflow Details

Workflow Context	Full ingest workflow with...
Workflow Definition	Full Ingest Workflow (Manual Selection, Store Metadata File, Thumbnails)
Workflow ID	19
Workflow State	Completed
Date Started	04.10.11 11:19:00
Date Finished	04.10.11 11:21:37
Number of Files	26
Total Size	2 MB
Collection Code	CALM/GH1
Submission name	CALM Export Test
Top Level Record	Spire Ingest 2

Step Progress

State	Name	Progress	Started	Finished	Messages
	Select	<div></div>	04.10.11 11:19:00	04.10.11 11:19:07	
	Copy XIP Package	<div></div>	04.10.11 11:19:07	04.10.11 11:19:11	
	Virus Check	<div></div>	04.10.11 11:19:11	04.10.11 11:19:33	
	Fixity Check	<div></div>	04.10.11 11:19:33	04.10.11 11:19:35	
	Metadata Integrity	<div></div>	04.10.11 11:19:35	04.10.11 11:19:57	
	Content Integrity	<div></div>	04.10.11 11:19:57	04.10.11 11:20:19	
	Characterise	<div></div>	04.10.11 11:20:19	04.10.11 11:20:51	View
	Store Files	<div></div>	04.10.11 11:20:51	04.10.11 11:20:55	
	Store Metadata	<div></div>	04.10.11 11:20:55	04.10.11 11:21:18	
	Store Metadata File	<div></div>	04.10.11 11:21:18	04.10.11 11:21:25	





Things I wish I'd known #5b

**A digital archivist's role is to
manage the exceptions**



HOUSES OF PARLIAMENT
PARLIAMENTARY ARCHIVES

Things I wish I'd known #6

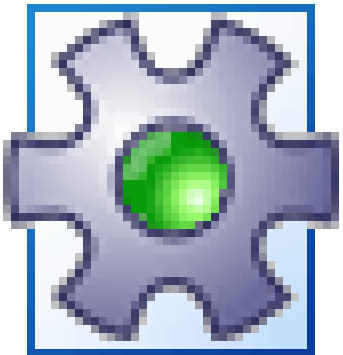
**Characterisation is one of the
most important tasks**



HOUSES OF PARLIAMENT
PARLIAMENTARY ARCHIVES

Characterisation

- Identification
 - *What is it?*
- Validation
 - *Is it really it?*
- Property extraction
 - *What can we measure about it?*
 - *Technical and significant properties*



Preservation planning

[?] Sample Records

[What are sample records and why do I need them?](#)

Description of sample records: PDF sample

Sample Record

Full name: Test012.pdf

Short name: Test012.pdf

Has data: ☒ 0.97 MB [download](#)

Original technical environment:

Description:

[Remove sample](#)

Object Format

PUID: fmt/16

Name: Portable Document Format

Version: 1.2

Mime-type: application/pdf

Extension: pdf

[Identify format](#)



[View Characteristics](#)



[View FITS Characteristics](#)



XCDL Description:



[Describe samples in XCDL](#)

[What is XCDL?](#)

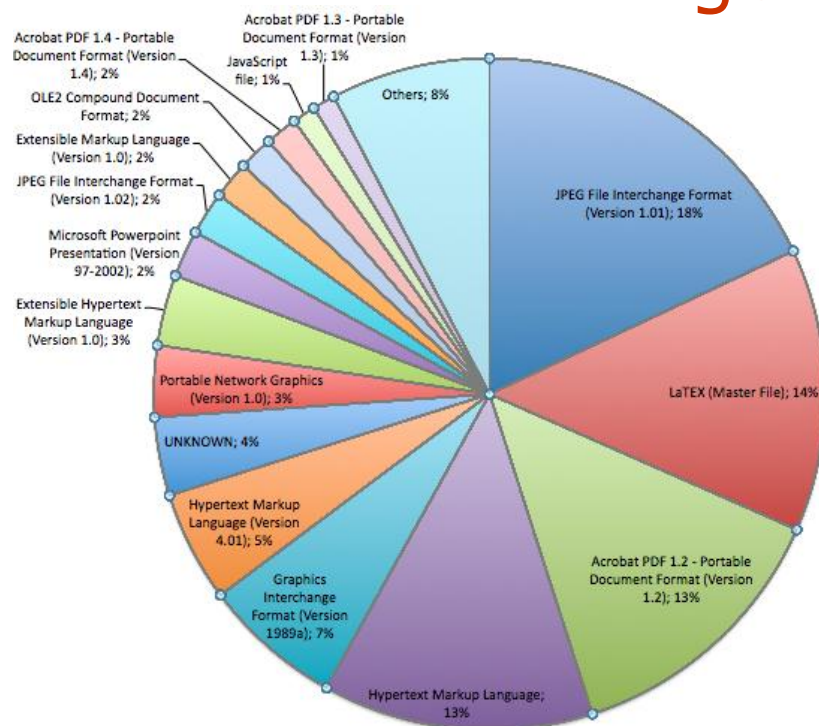


HOUSES OF PARLIAMENT
PARLIAMENTARY ARCHIVES

Collection profiling

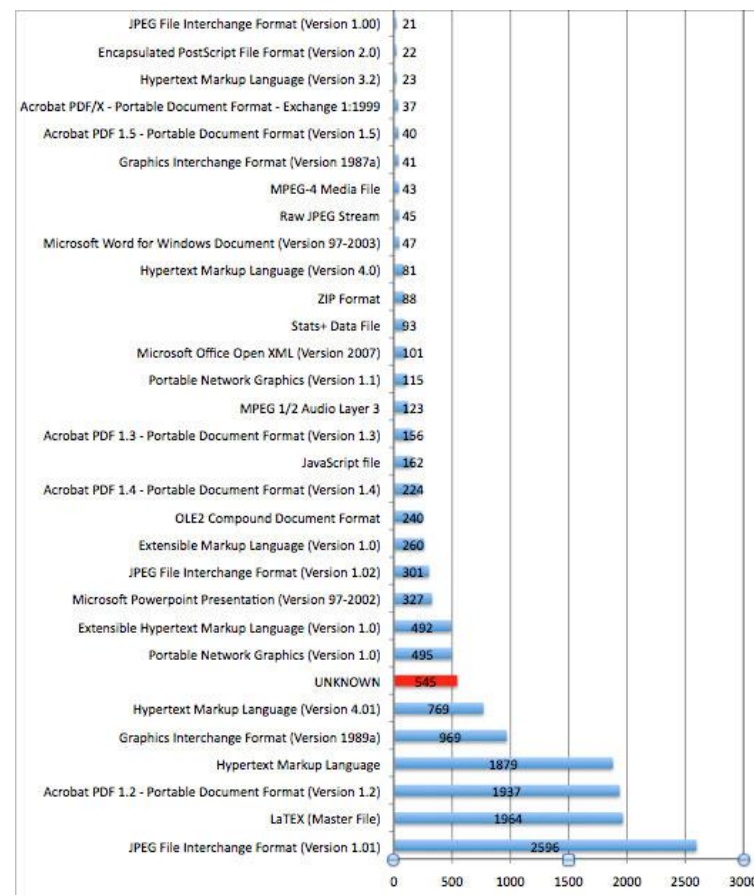
Formats

Age

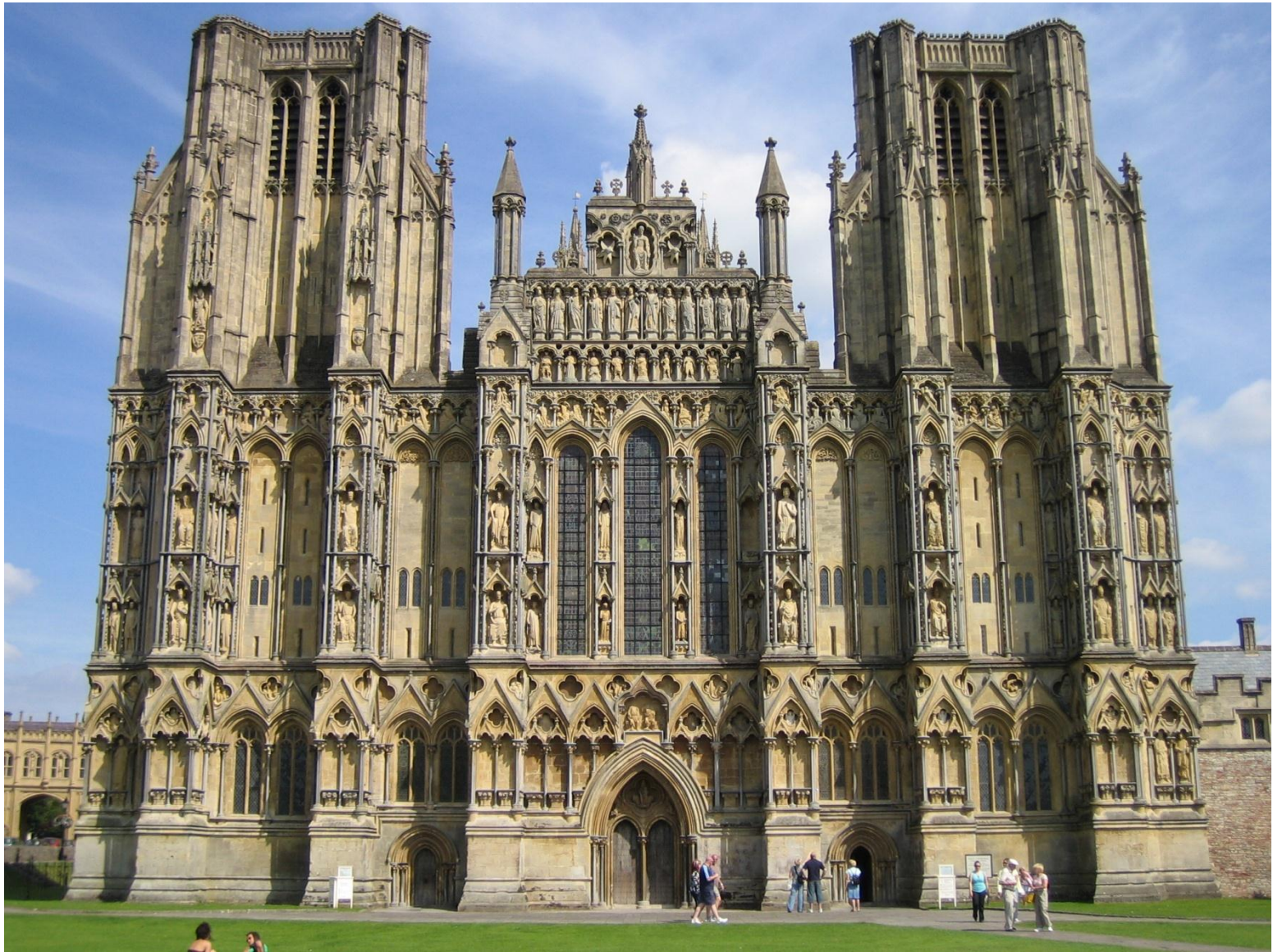


Volume

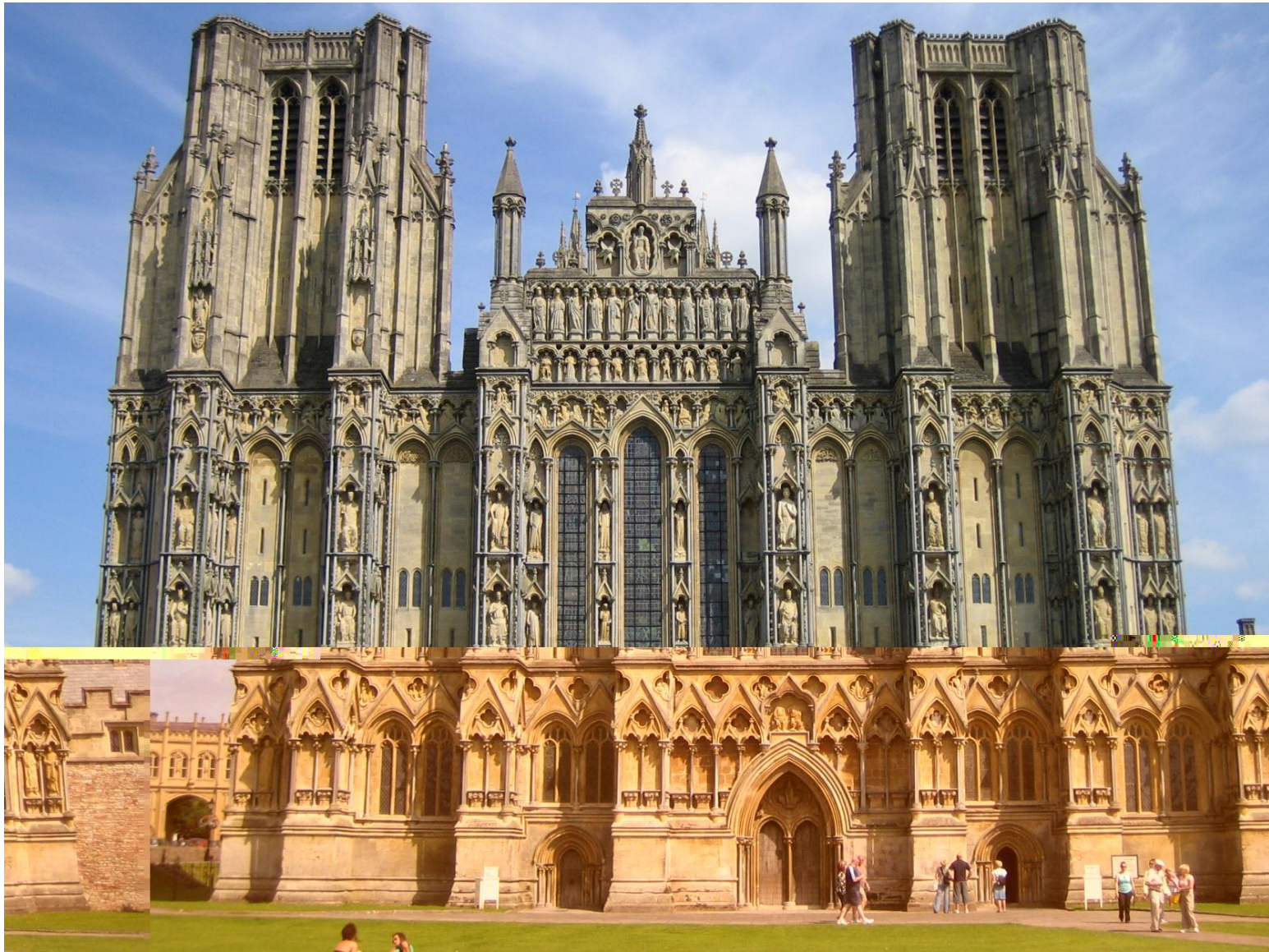
Usage



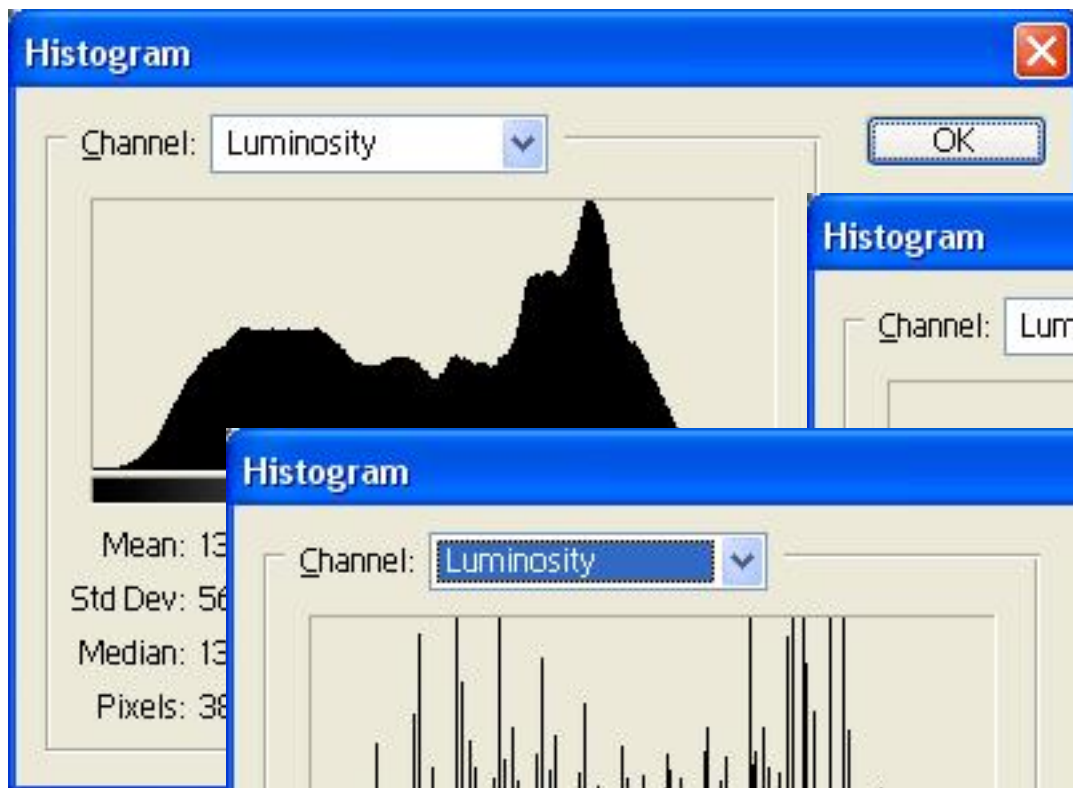
HOUSES OF PARLIAMENT
PARLIAMENTARY ARCHIVES



HOUSES OF PARLIAMENT
PARLIAMENTARY ARCHIVES

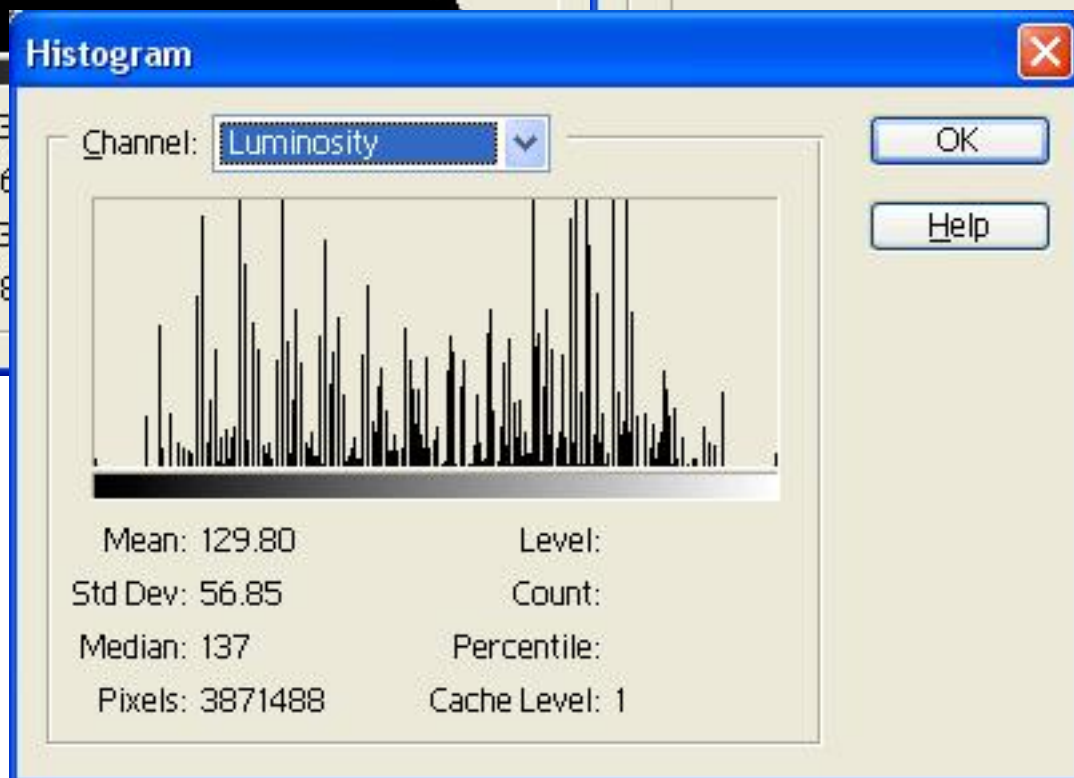
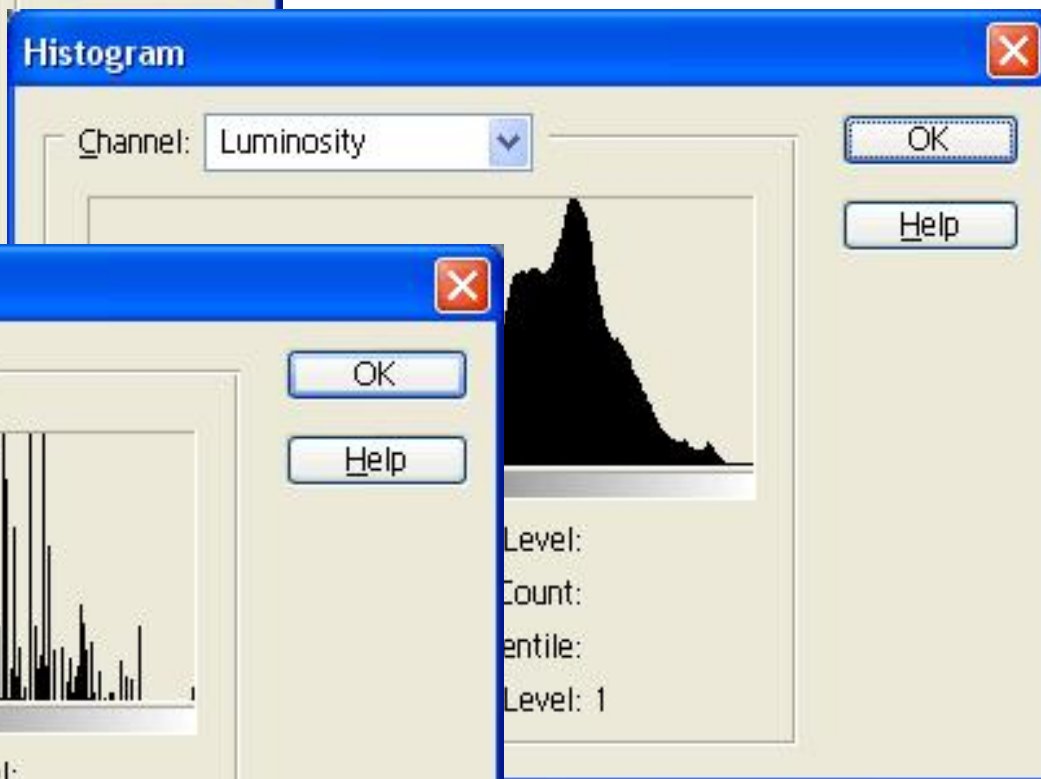


HOUSES OF PARLIAMENT
PARLIAMENTARY ARCHIVES



BMP

PNG



GIF



HOUSES OF PARLIAMENT
PARLIAMENTARY ARCHIVES

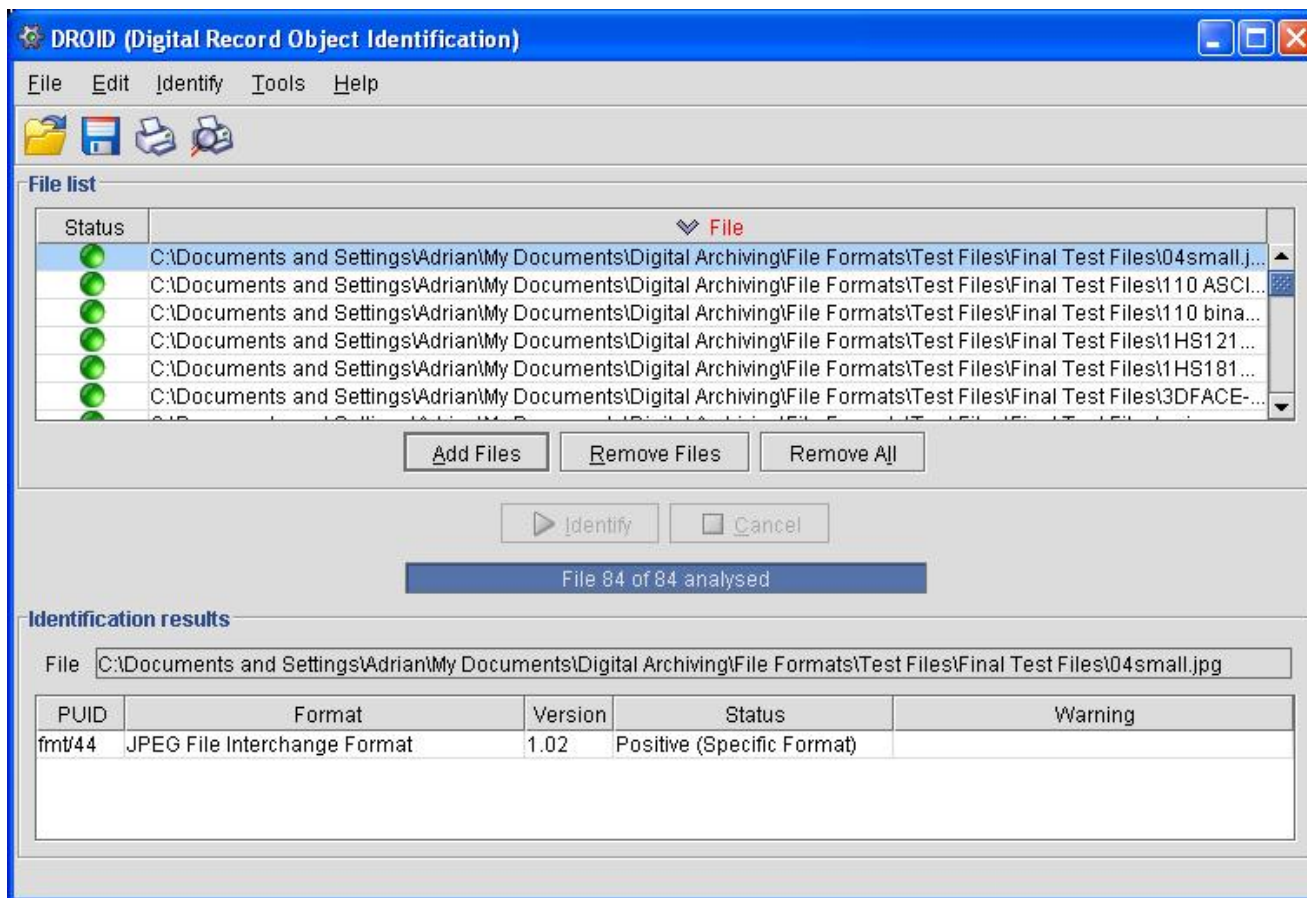
Things I wish I'd known #7

**Characterisation is very
amenable to automation**

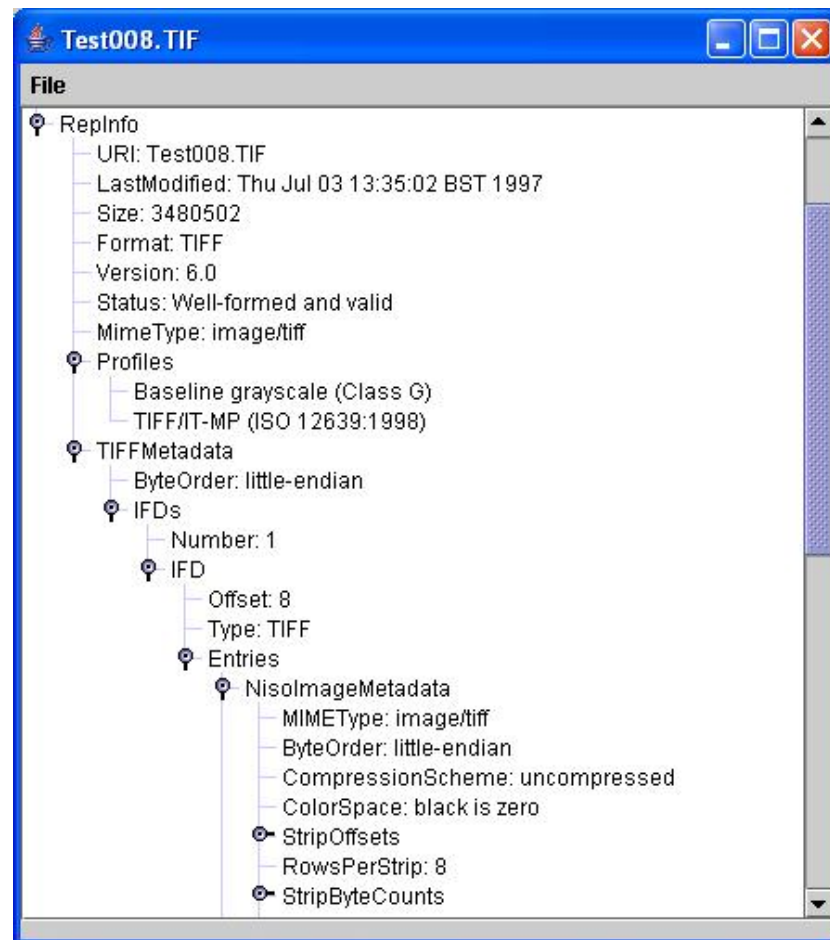
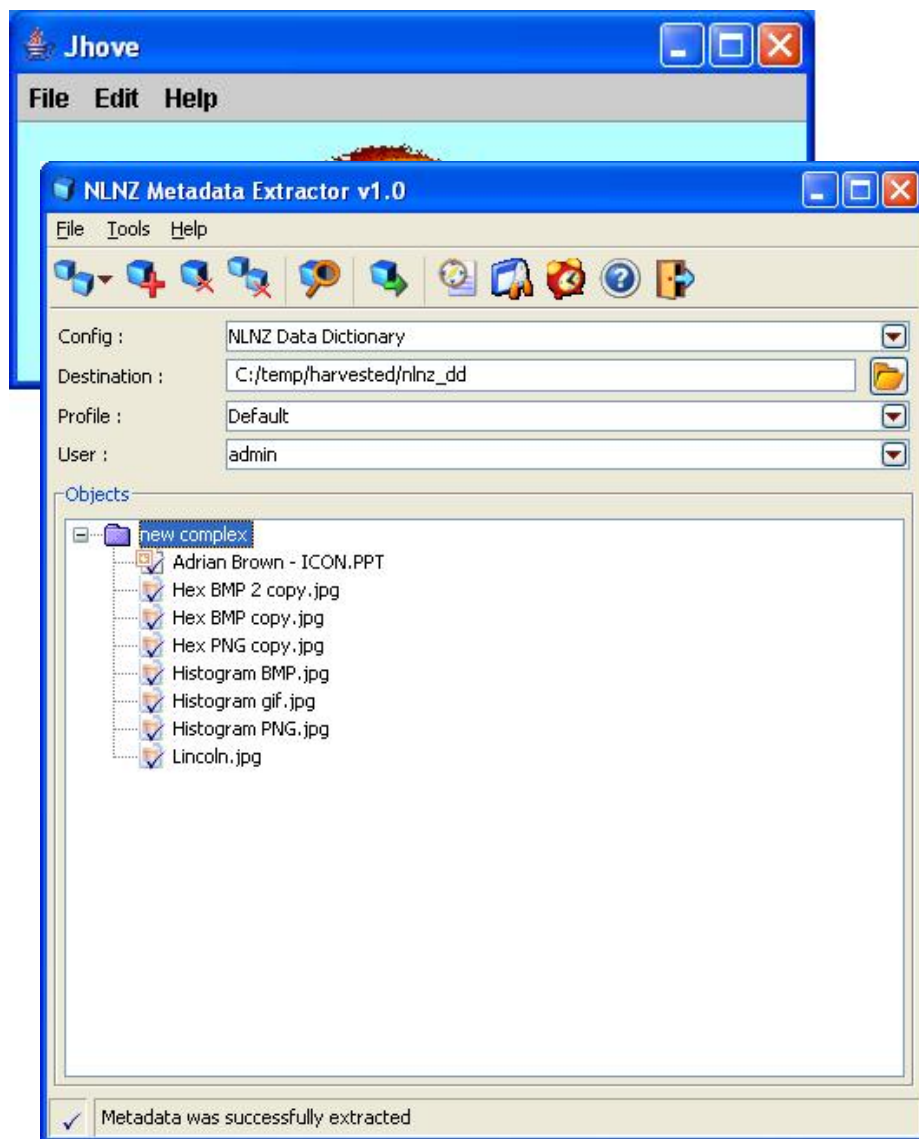


HOUSES OF PARLIAMENT
PARLIAMENTARY ARCHIVES

Format identification



Validation and property extraction



Things I wish I'd known #8a

**We can automatically extract
vast amounts of metadata
about digital objects**



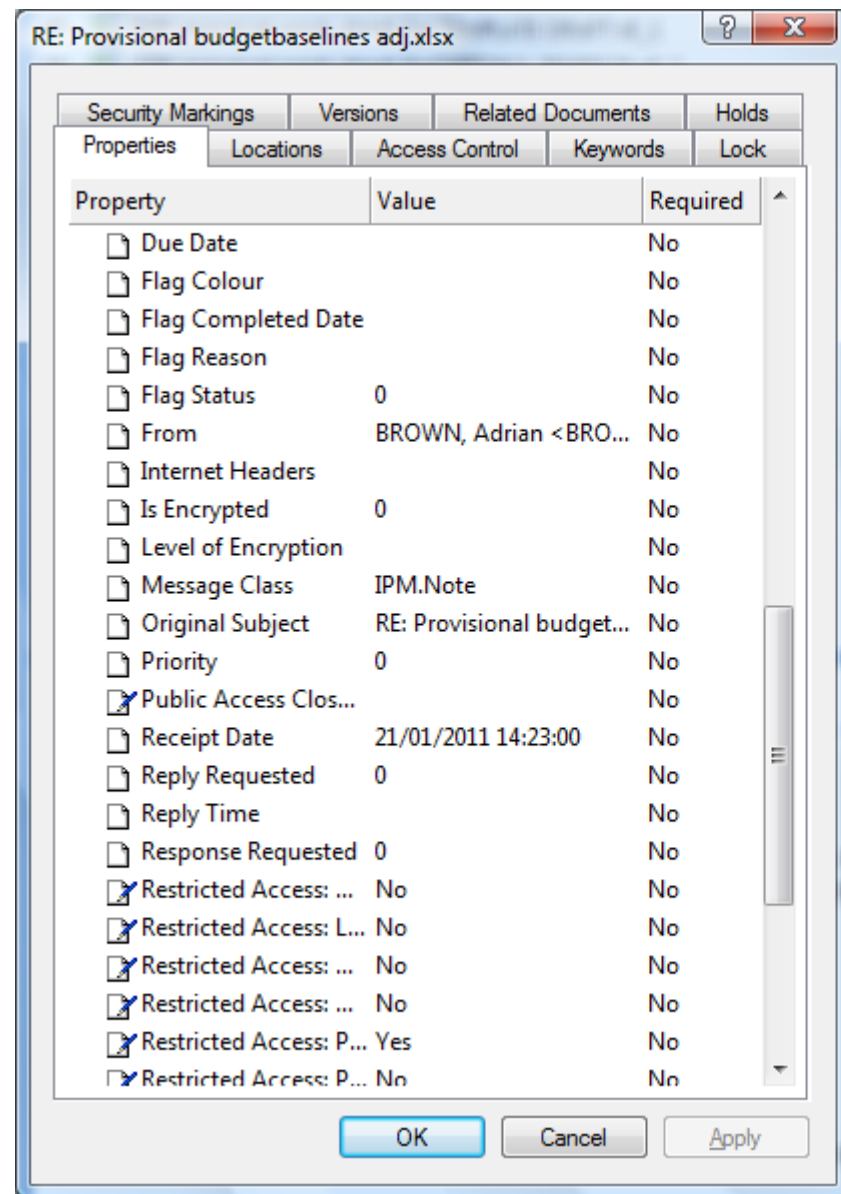
HOUSES OF PARLIAMENT
PARLIAMENTARY ARCHIVES

Things I wish I'd known #8b

We probably shouldn't



HOUSES OF PARLIAMENT
PARLIAMENTARY ARCHIVES



HOUSES OF PARLIAMENT
 PARLIAMENTARY ARCHIVES

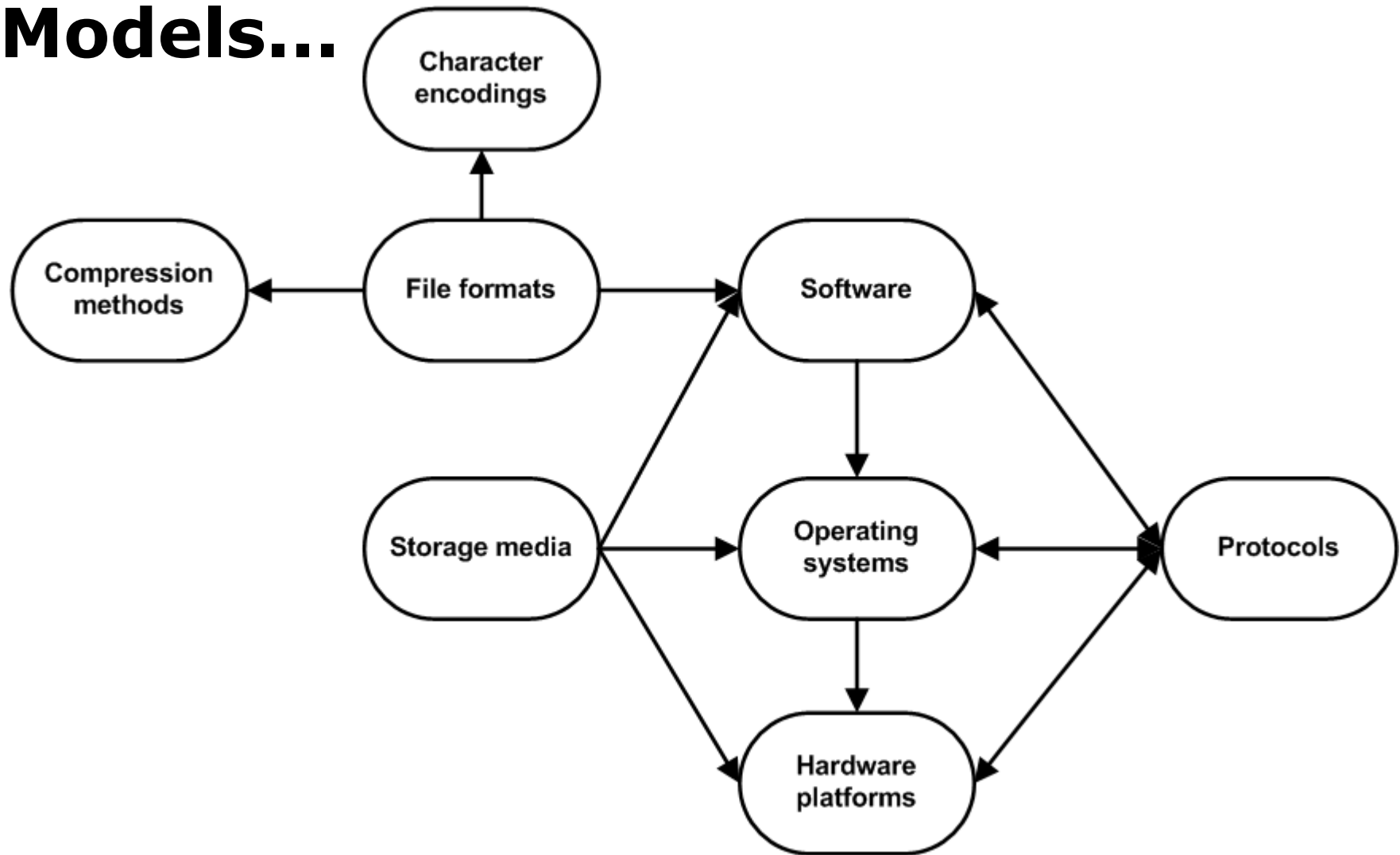
Things I wish I'd known #9

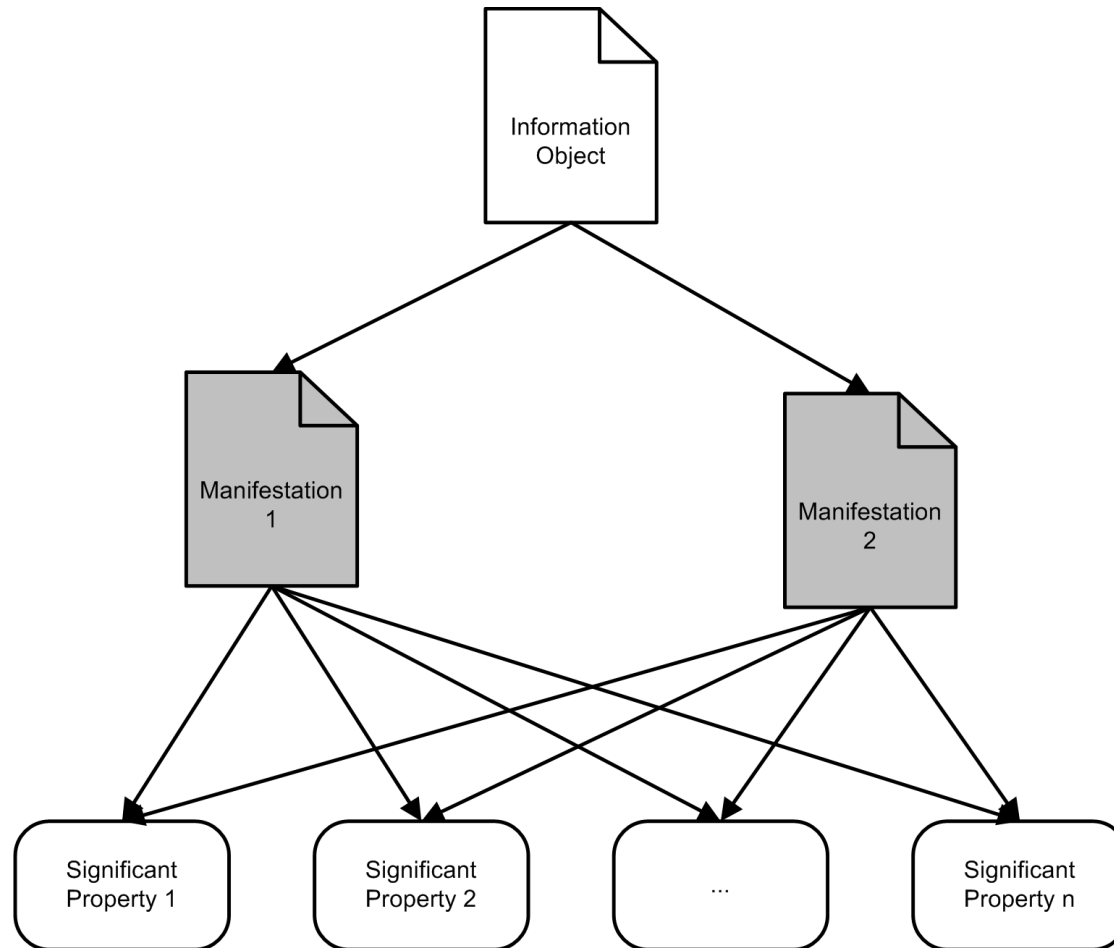
Don't try to model the world

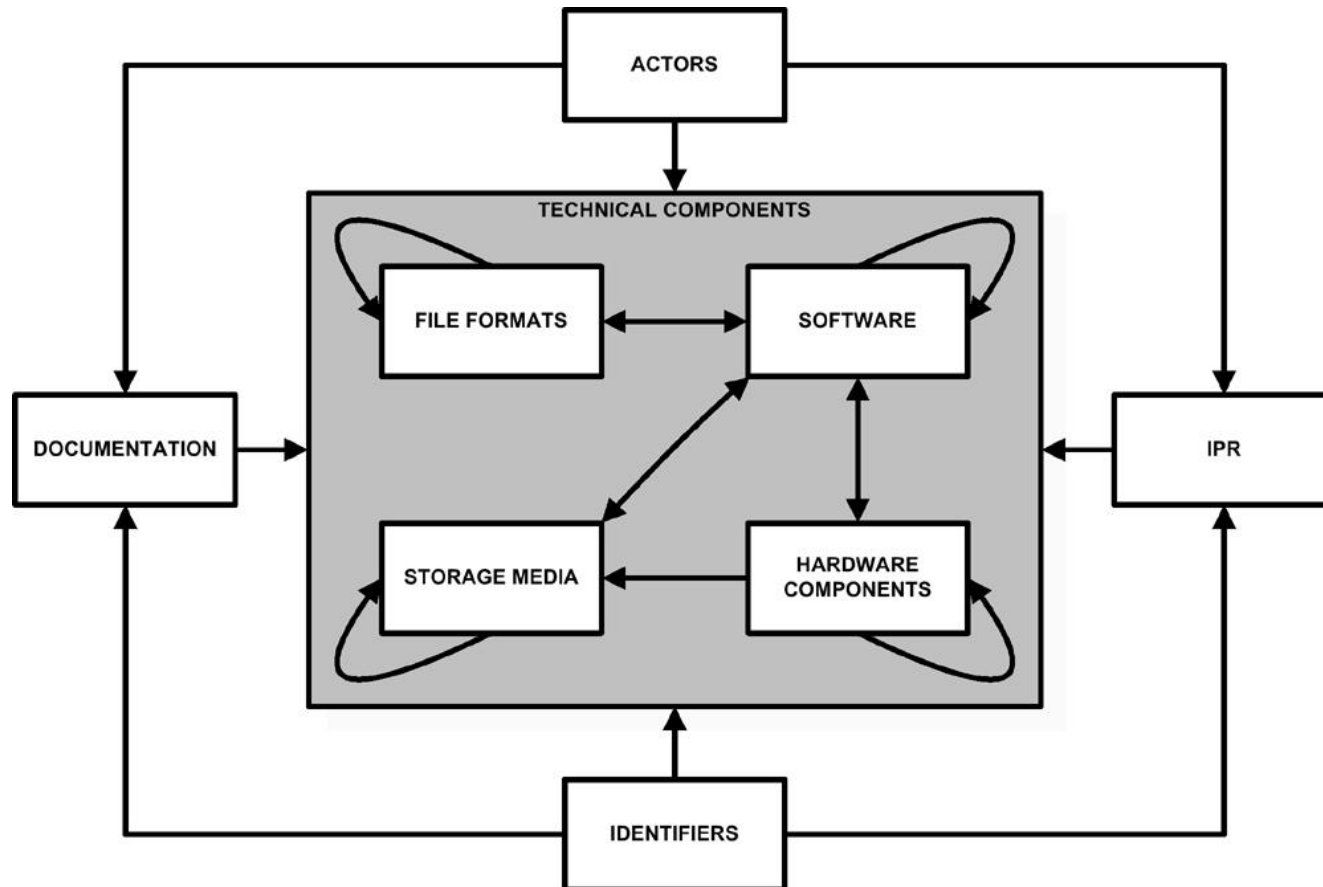


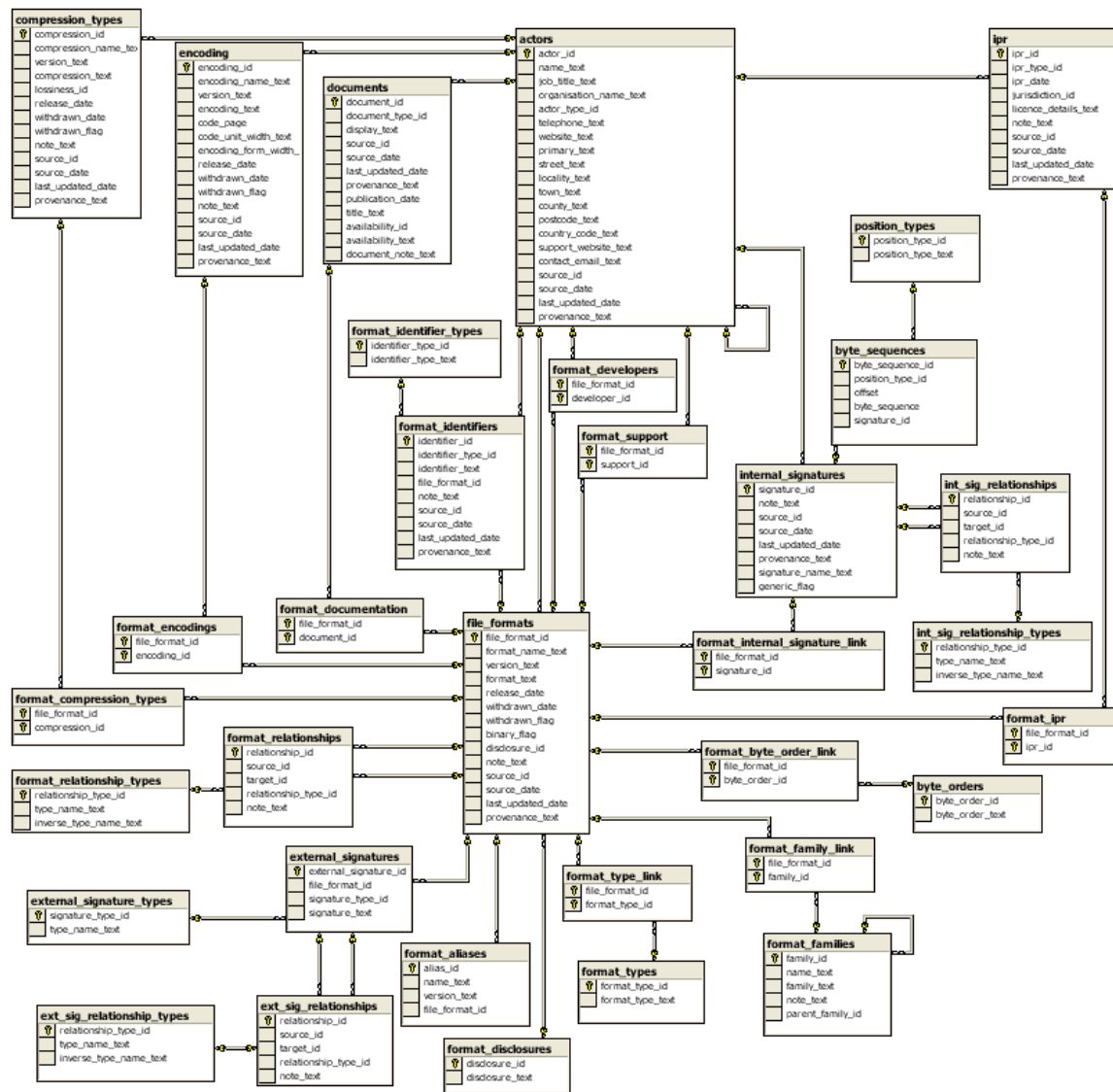
HOUSES OF PARLIAMENT
PARLIAMENTARY ARCHIVES

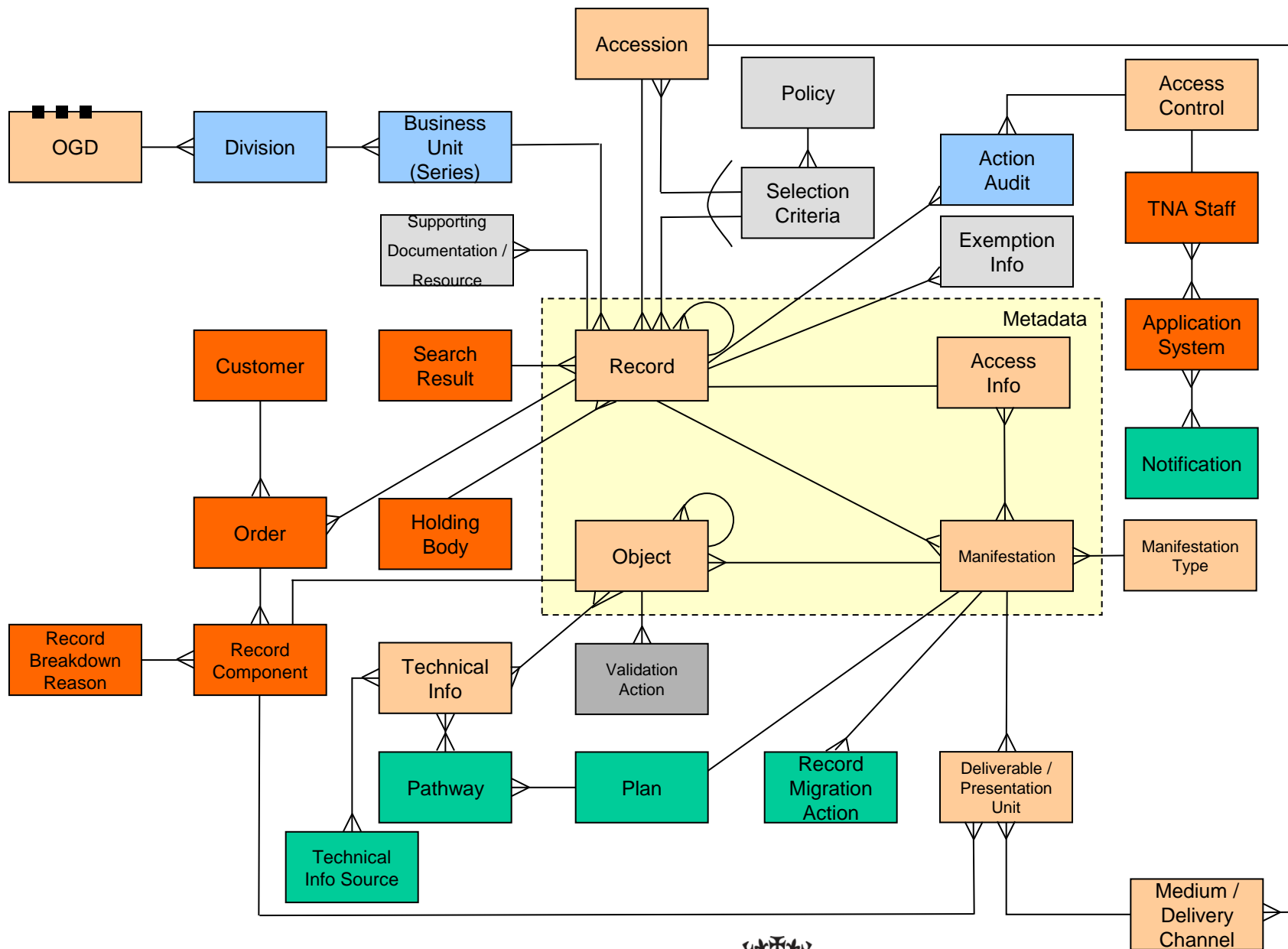
Models...



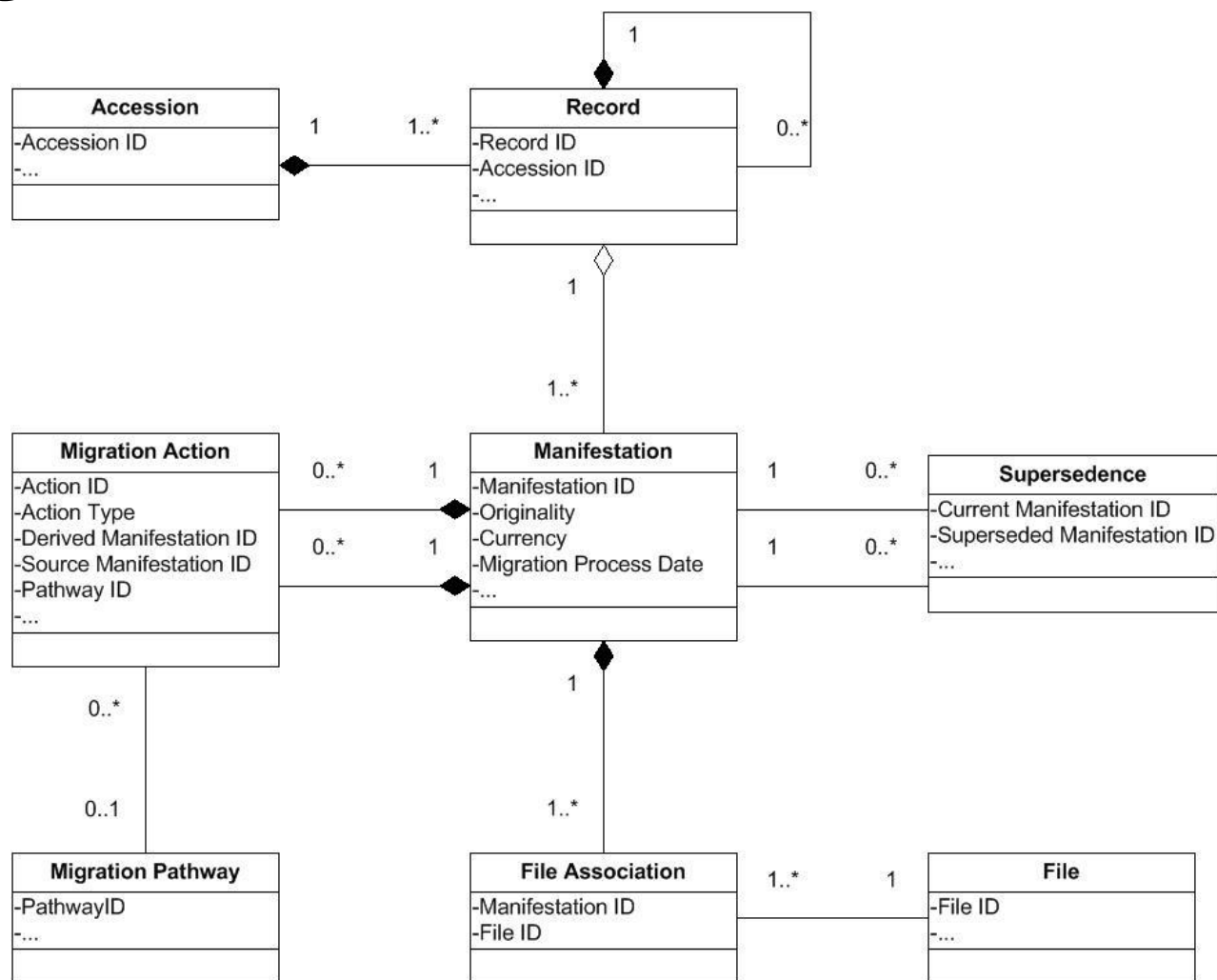








...etc



Things I wish I'd known #10

You don't have to solve every problem yourself - collaborate



HOUSES OF PARLIAMENT
PARLIAMENTARY ARCHIVES

Collaboration



Collaboration



**Archives & Records
Association**
UK & Ireland



HOUSES OF PARLIAMENT
PARLIAMENTARY ARCHIVES

Collaboration



HOUSES OF PARLIAMENT
PARLIAMENTARY ARCHIVES

Questions?

Web: <http://www.parliament.uk/archives>

Twitter: @UKParlArchives, @realAdrianBrown



HOUSES OF PARLIAMENT
PARLIAMENTARY ARCHIVES