

Digital Preservation In Practice

Ed Fay

London School of Economics

EMAIL: e.fay@lse.ac.uk

TWITTER: [@digitalfay](https://twitter.com/digitalfay)



British Library of Political
and Economic Science

We do not have a long-term
digital preservation strategy

We do not have a long-term
digital preservation strategy

...and we're perfectly
ok with that!

This does mean....

That we are agnostic about
the 'final' solution

(We don't even necessarily think there is/has to be one...)

That we are happy to investigate,
experiment and change what we are doing

This doesn't mean...

We are doing nothing!

LSE

Library

British Library of Political
and Economic Science

We have:

**Decided we want to be
preserving digital stuff**

We are:

Starting to take the first steps

Long-term means *long*-term

- Digital preservation is about preserving *access* at a given point in time
- Right now we have no critical collections but, given time, we will

Long-term means *long*-term

- We don't need:
 - a 'complete solution'; right now
- We do need:
 - to start thinking, getting clear about the problem, and talking about it
- We also need to start doing something so it doesn't become too late

BORN-DIGITAL

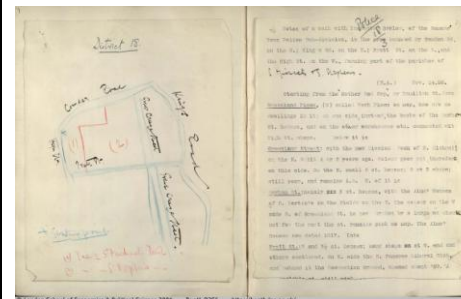


DIGITAL ARCHIVES

DIGITISATION



INSTITUTIONAL REPOSITORY



British Library of Political and Economic Science

What have we done?

- Collections audit (a spreadsheet)
- Risk assessment (DRAMBORA)
- ...
- User requirements analysis (ongoing, for curators as well as end-users)

DMID Digital Collections Scoping Sep-09						
Collection	Identifier	Format	Size (MB)	Media	Location	Accretion/Attrition
Exam papers		PDF	1126	Server/Tape Backup	https://library-2.lse.ac.uk/protected-exam/	
Research publications	LSE RO	PDF Word/data	4710 ?	Server/Tape Backup Server/Tape Backup	http://eprints.lse.ac.uk http://eprints.lse.ac.uk	
e-Theses	LSE TO?					
Data Resource Archive		text/PDF/SB\$	~5000?	Server	lib-6	stable (once it has been
Paddy Ashdown diaries	M3348	Word/html/en	1628	CD (4)	Archives	Expecting more Ashd

Risk Number 1

Risk Name:	Loss of trust or reputation
Risk Description:	One or more stakeholder communities have doubts about the repository's ability to achieve its objectives.
Risk Areas:	Personnel, Management & Admin Procedures Operations & Service Delivery
Vulnerability(ies):	* An irrecoverable loss of digital objects provokes community concerns about the repository's competence
Consequence(s):	* Credibility of the organisation as a location for deposit of digital objects, and associated funding, reduced.
Relationship(s):	No Relationships Established
Risk Owner(s):	Senior Manager

Tools: DRAMBORA

- Risk assessment
- Why?
- Start the conversation
- Make the problems clear to all stakeholders (curators, technical specialists, senior managers)
- Not for detailed functional analysis
- <http://repositoryaudit.eu/>

What are we doing?

- Creating a place to store our digital objects
 - 'repository core' for object storage (redundancy, backups) and identification
- Creating a way to ingest/accession objects
 - workflow for object characterisation, checksums, quarantine/virus check

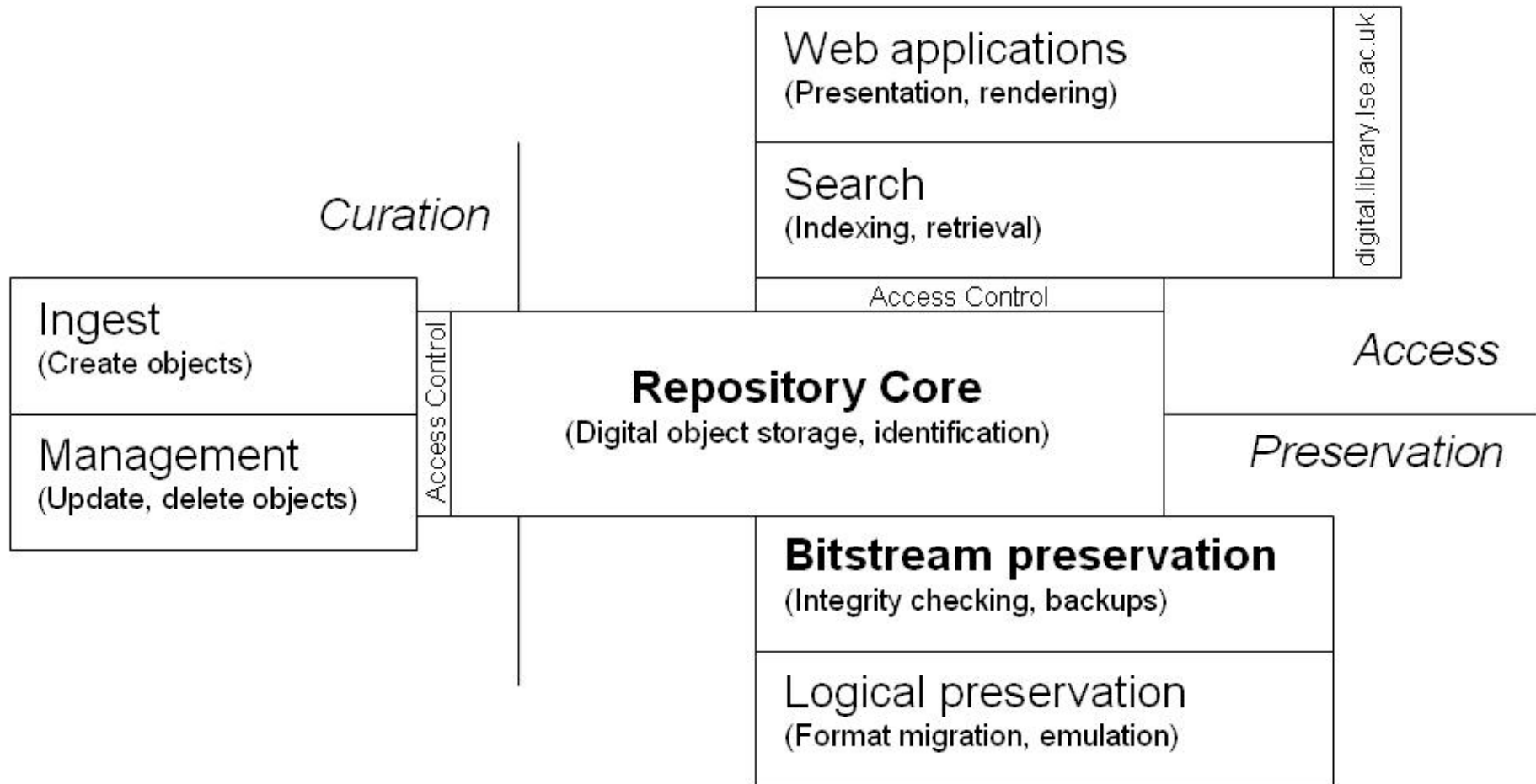
Tools: Archivematica

- Workflow tool
- Why?
- For 'ingesting' digital objects
- Bundles tools for:
 - quarantine/virus check (filesystem, ClamAV)
 - checksum creation/verification (MD5)
 - format characterisation/validation (FITS which packages DROID, JHOVE, NZ Metadata Extractor ...)
- <http://archivematica.org/>

What will we be doing?

- Building management interfaces
- Building access interfaces
- Developing a logical preservation approach
 - strategy and policies
- Building logical preservation functionality
 - implementing tools for migration or emulation

LSE Digital Library



LSE Digital Library: Design Principles

- Flexible—we can hold a range of different types of digital collection
- Extensible—we can adapt to changing collections and user requirements
- Modular—we can replace components without disrupting other functions

Tools

- DRAMBORA
 - **used** for high-level risk assessment – as a tool for starting the conversation
- Archivemata
 - **used** to characterise our collections and to assist in producing a more detailed risk profile for further analysis
- Fedora/Hydra repository
 - **will be used** to store/manage all our digital collections
- Planets/Plato
 - **will(?) be used** to help us plan our long-term strategy

Guiding principles

- Openness: standards, technologies
- Transparency: clear, documented decisions and processes
- Engagement: bringing everyone along with us (senior managers, depositors, colleagues across the library)

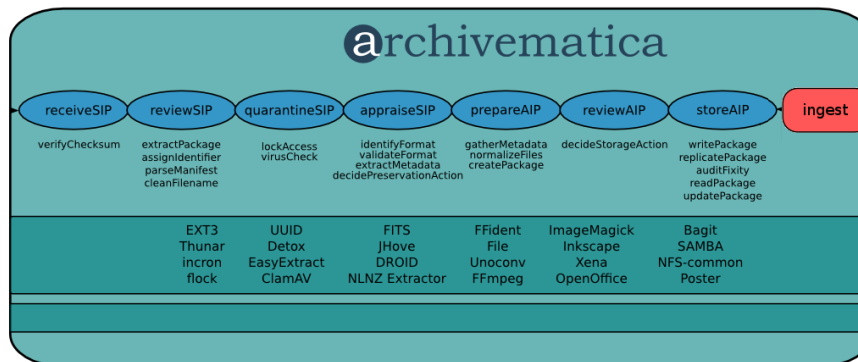
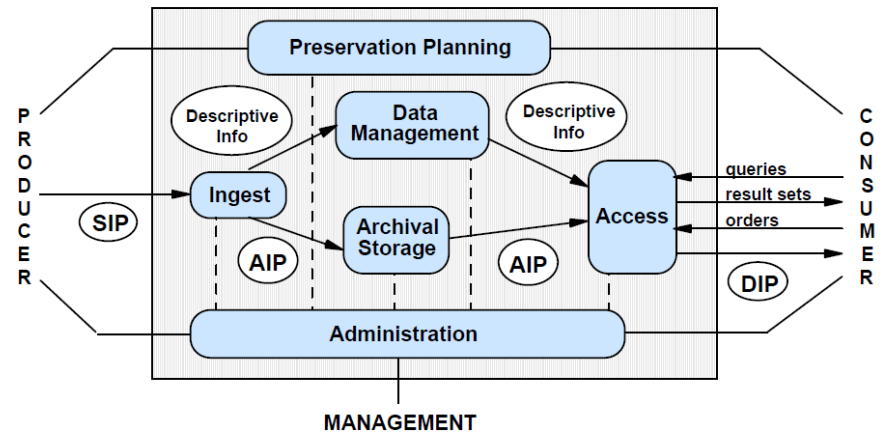
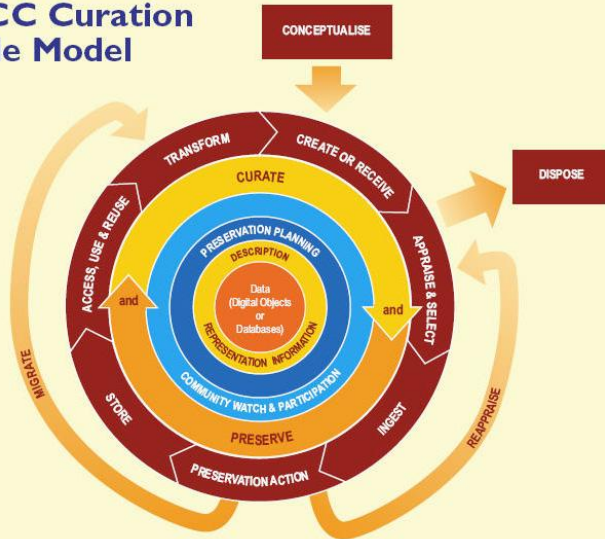
The **problems...**
...you've heard about

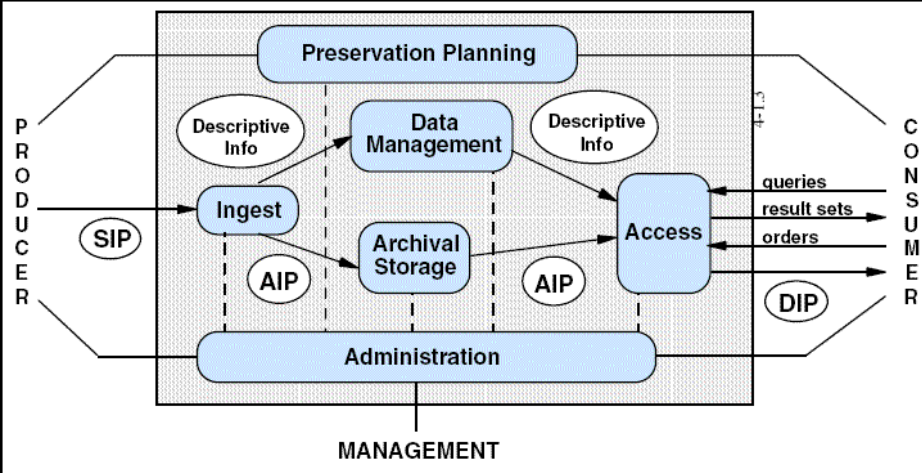
The **solutions...**
...are complex

Digital Preservation is HARD!

Digital Preservation is HARD!

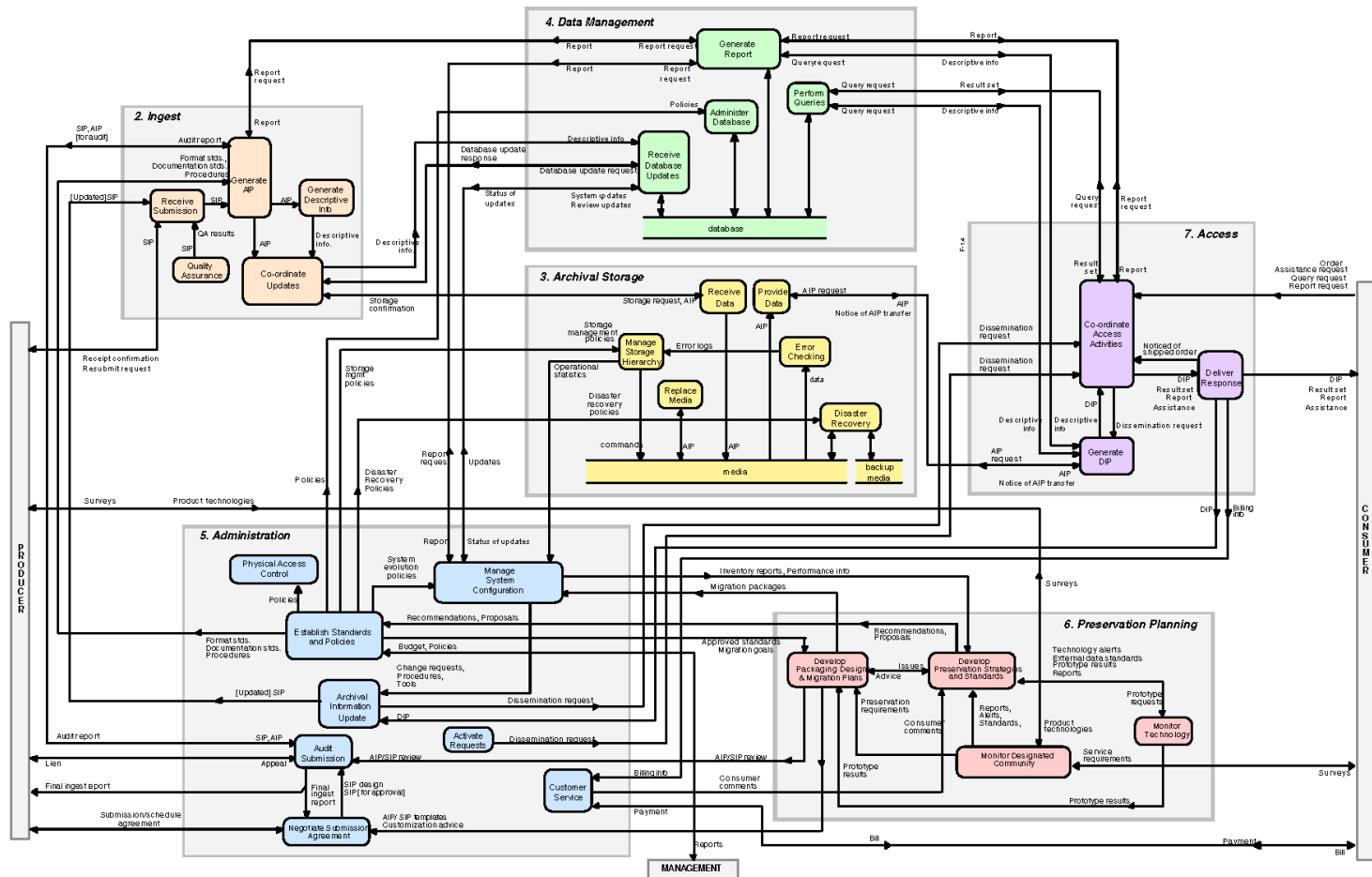
The DCC Curation Lifecycle Model





OAIS can help

But
it
can
also
scare
people!



Digital Preservation is HARD!

Shorter summary of DP: **know what you have and value, assess risk**, take action to avoid risk, repeat.

Problem: people don't do it

Steve Hitchcock

JISC KeepIt Project Manager

<http://twitter.com/#!/jisckeepit/status/25530206525591552>



Library

British Library of Political
and Economic Science

Small steps now...

- ...save big(ger) problems later
- **Learn by doing** – you (should) always have originals
- Focus on:
 - “**ingest**” = capture, identify
 - “**bit-preservation**” = redundancy, backups
- Use this as the basis for more **thorough risk assessment**
- Use that as basis to **make the case for investment**
- Then think about the long-term (policy and tech)

That's it!

Ed Fay
London School of Economics

EMAIL: e.fay@lse.ac.uk

TWITTER: [@digitalfay](https://twitter.com/digitalfay)



British Library of Political
and Economic Science