Digital Preservation Planning
Case Study

Getting Started in Digital Preservation

British Library Preservation Advisory Centre

13 May 2013

Ed Fay, Digital Library Manager

e.fay@lse.ac.uk | @digitalfay

LSE **Library Services**
British Library of Political
and Economic Science

Digital Preservation Planning

Or, how to decide what you don't need to do

Digital Preservation Planning

Two important aspects:

→ collection

characterisation, risk assessment, prioritisation

(workflows, testing, validation)

→ organisation

capacity planning, staff skills, sustainability

Don't panic!

Digital collections: scope the problem

- Collections audit (format diversity, volume/growth)
- Risk assessment (threats to our strategic objectives)
- Prioritisation (where to start...)

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	LSERO	PDF	4710	Server/Tape Backup	http://eprints.lse.ac.uk	
		Word/data	?	Server/Tape Backup	http://eprints.lse.ac.uk	
e-Theses	LSETO?					
Data Resource Archive		text/PDF/SB?	~5000?	Server	lib-6	stable (once it has bee
Paddy Ashdown diaries	M3348					
		Word/html/er	1628	CD (4)	Archives	Expecting more Ashd

User Generated Risk Register DRAMBORA interactive
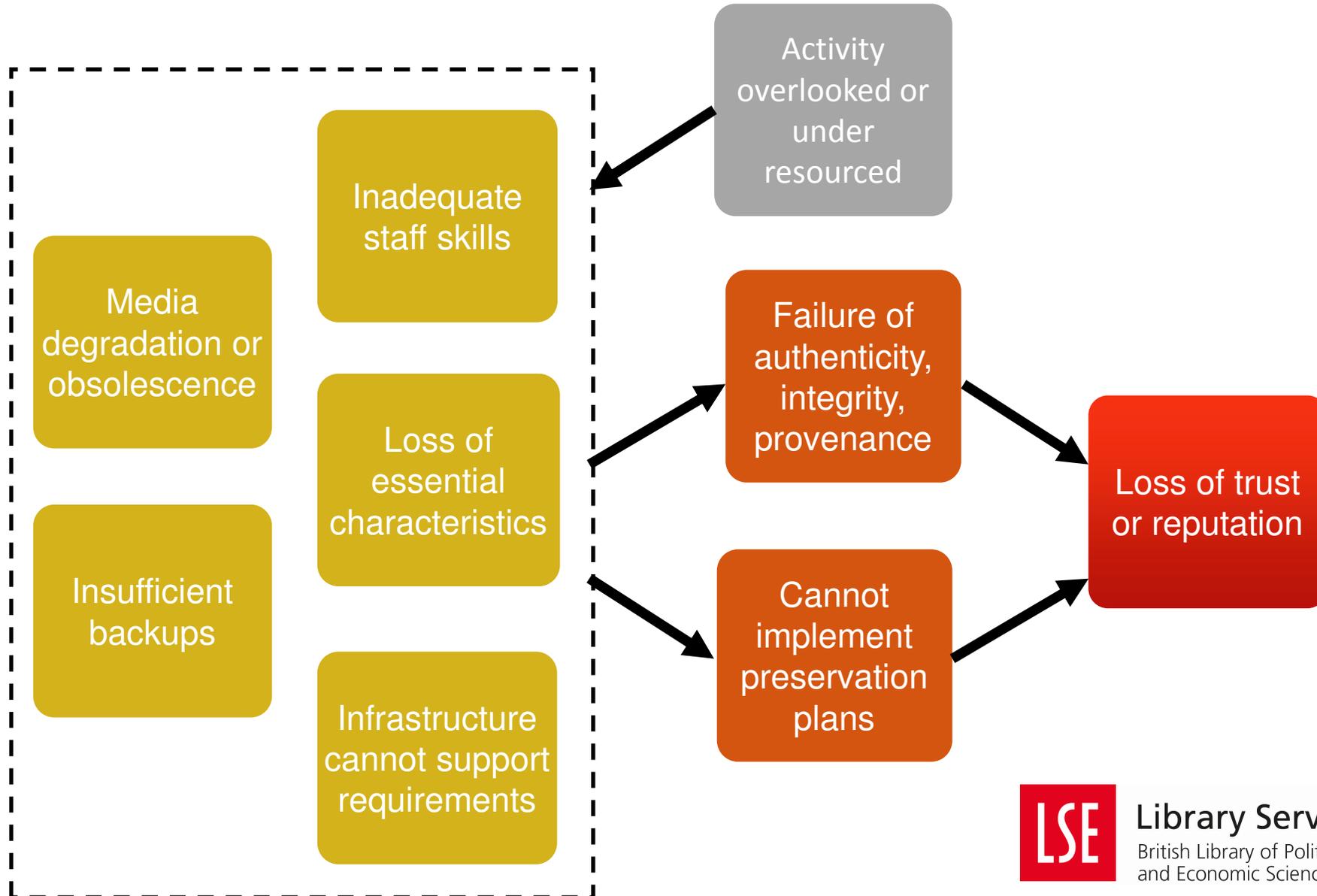Digital Collections

Risk Number 1

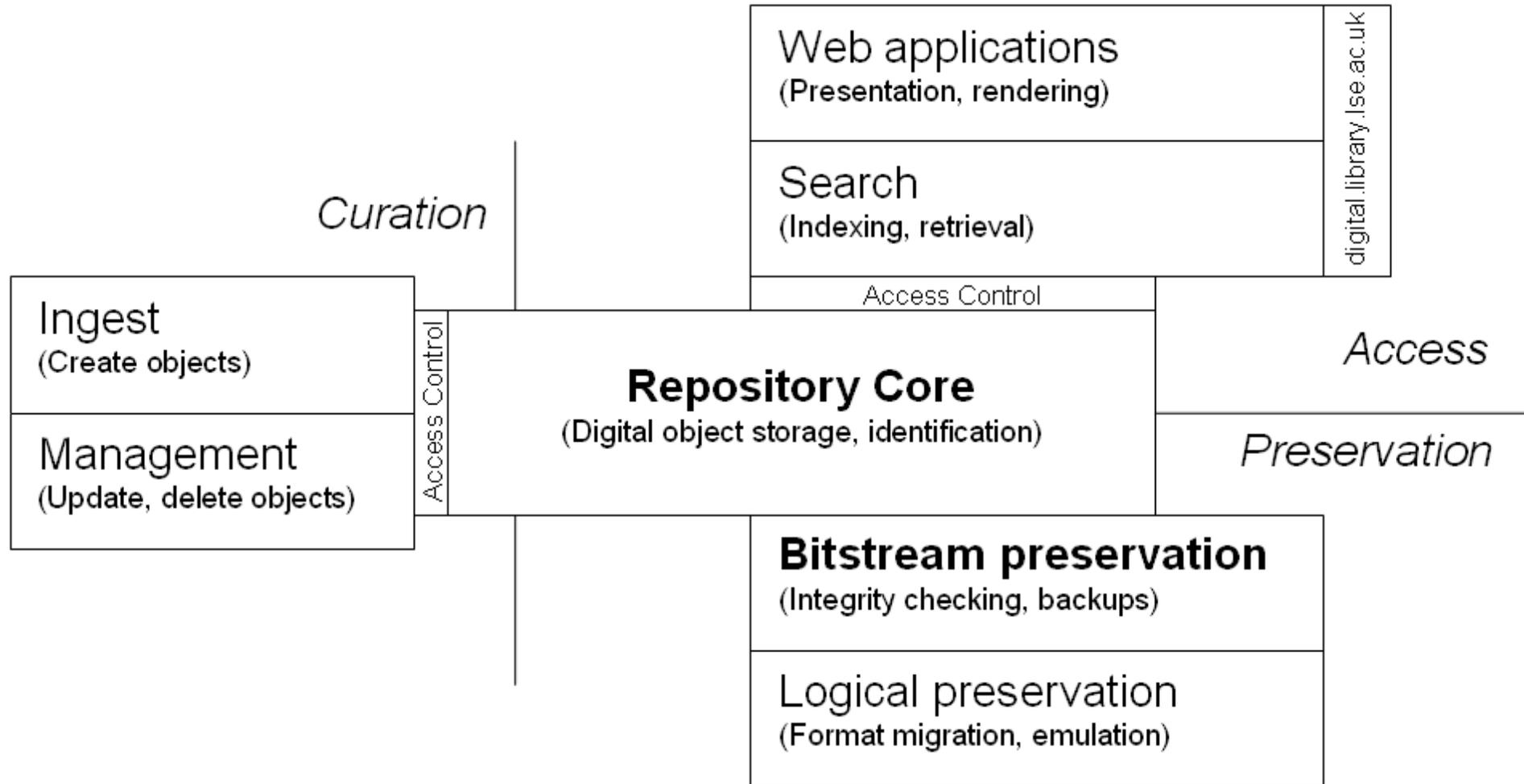| Risk Name:	Loss of trust or reputation
Risk Description:	One or more stakeholder communities have doubts about the repository's ability to achieve its objectives.
Risk Areas:	Personnel, Management & Admin Procedures Operations & Service Delivery
Vulnerability(ies):	* An irrecoverable loss of digital 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

LSE Library Services
British Library of Political and Economic Science

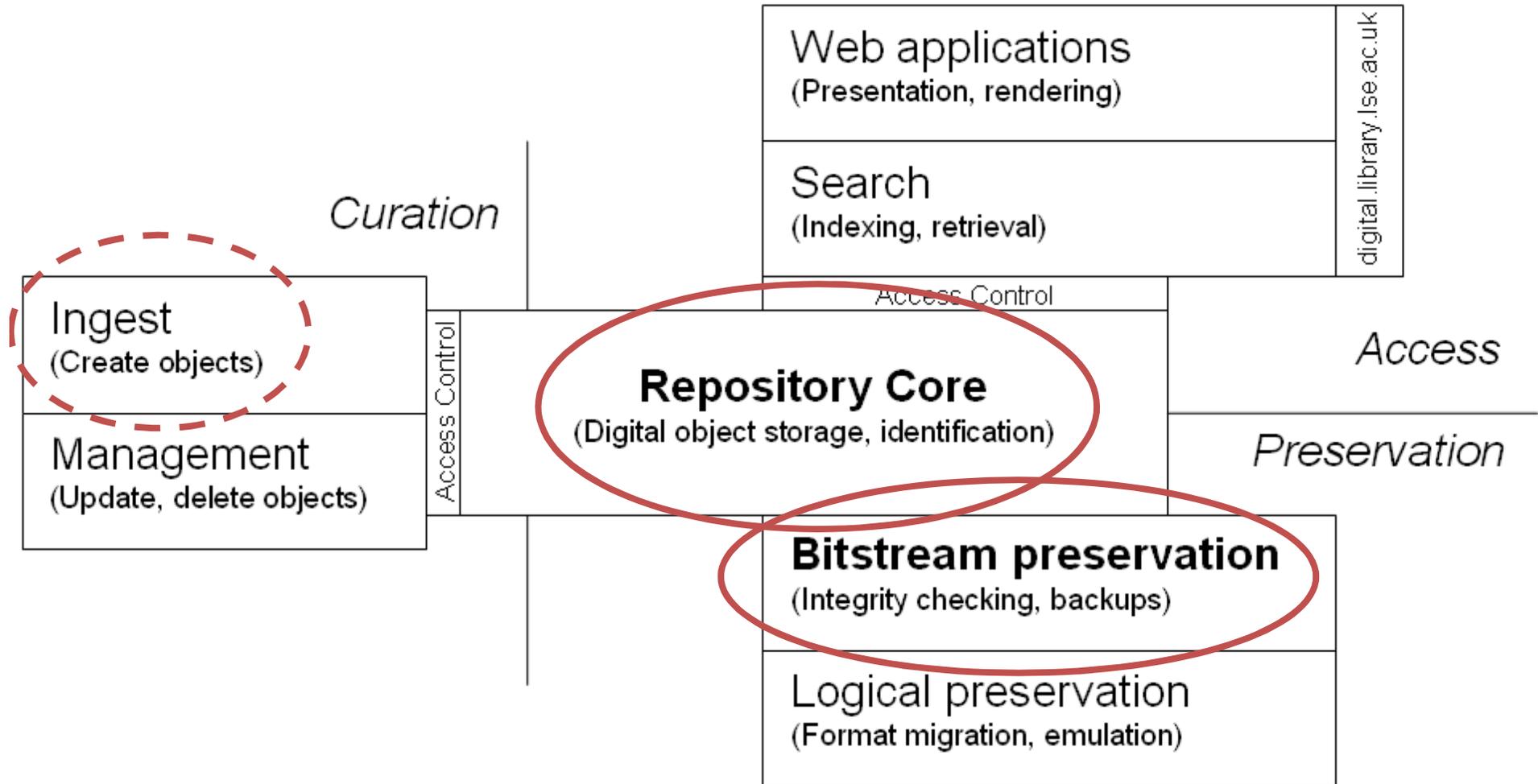# Digital collections: risk assessment

Prioritisation

Curation

Web applications
(Presentation, rendering)

Search
(Indexing, retrieval)

Access Control

Ingest
(Create objects)

Management
(Update, delete objects)

Access Control

Repository Core
(Digital object storage, identification)

Access

Preservation

Bitstream preservation
(Integrity checking, backups)

Logical preservation
(Format migration, emulation)

digital.library.lse.ac.uk

Phased implementation of
technical infrastructure, staff skills

Prioritisation

Phased implementation of
technical infrastructure, staff skills

Web applications
(Presentation, rendering)

Search
(Indexing, retrieval)

digital.library.lse.ac.uk

Curation

Access

Preservation

Ingest
(Create objects)

Management
(Update, delete objects)

Access Control

Access Control

Repository Core
(Digital object storage, identification)

Bitstream preservation
(Integrity checking, backups)

Logical preservation
(Format migration, emulation)

LSE Library Services
British Library of Political
and Economic Science

Three examples

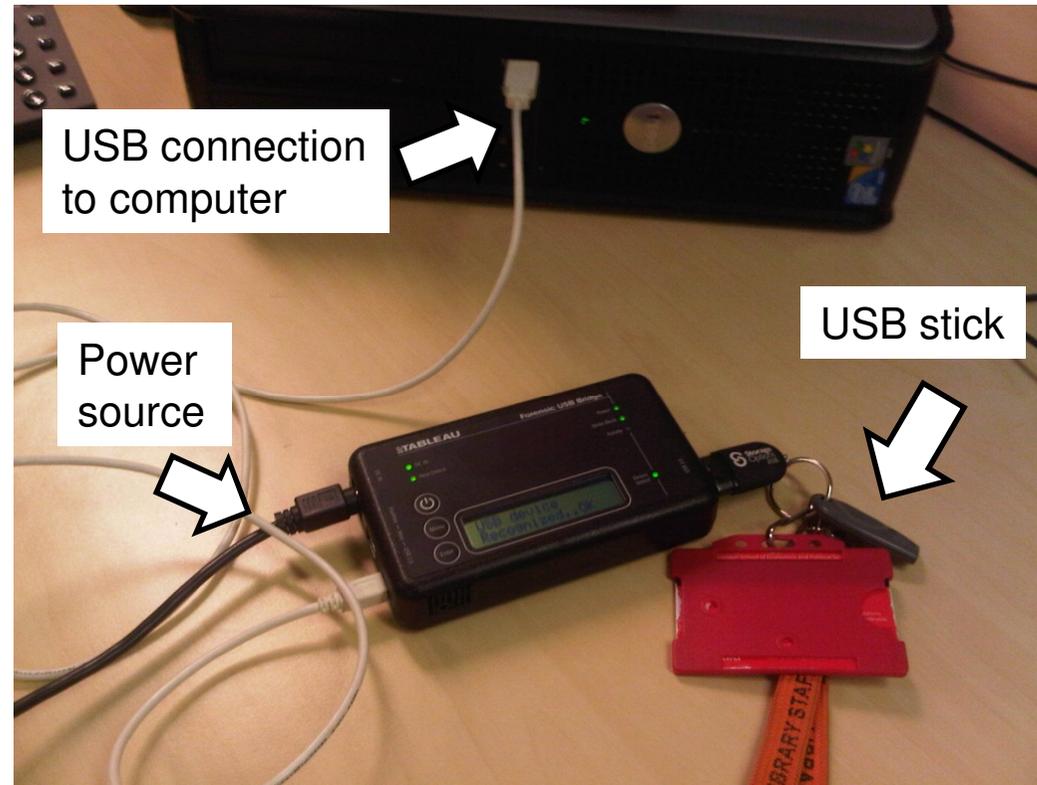| Collection	Format	Risk	Action
Legacy media	Floppy disks CD/DVDs	• Media obsolescence • Format obsolescence	• Forensic imaging • Backup disk images • Characterise formats
Digitisation	Mostly TIFF (c.11 total)	• Separation of digital files from metadata • (Format diversity)	• Full collection audit • Repository ingest
Public lectures	HD video Web video Audio	• Separation of HD/Web versions • Large file sizes	• Normalise formats (compress HD video) • Repository ingest

Example: legacy media (archives)

Risk: media obsolescence   Action: forensic imaging

USB connection
to computer

Power
source

USB stick

60 collections (hybrid and digital archives), total size 70GB

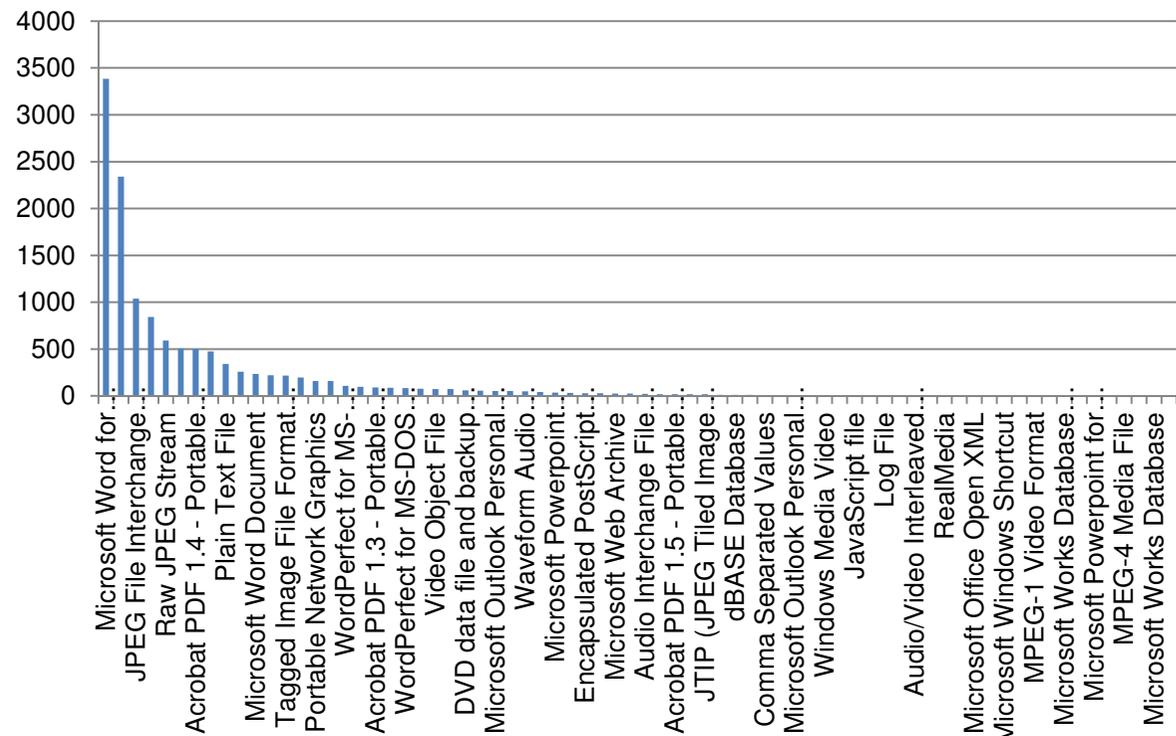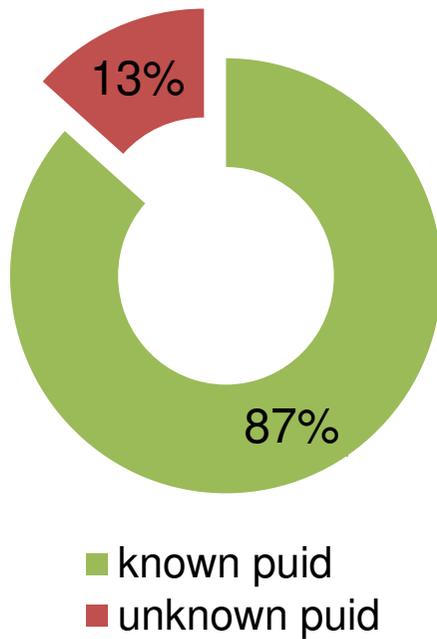14,829 files, average 247 files per collection

LSE Library Services
British Library of Political
and Economic Science

Example: legacy media (archives)

Risk: format obsolescence

Action: characterise

60 collections (hybrid and digital archives), total size 70GB

14,829 files, average 247 files per collection

Example: legacy media (archives)

Collection audit

→ find old media

→ assess: risk of 1) media obsolescence (known); 2) format obsolescence (unknown)

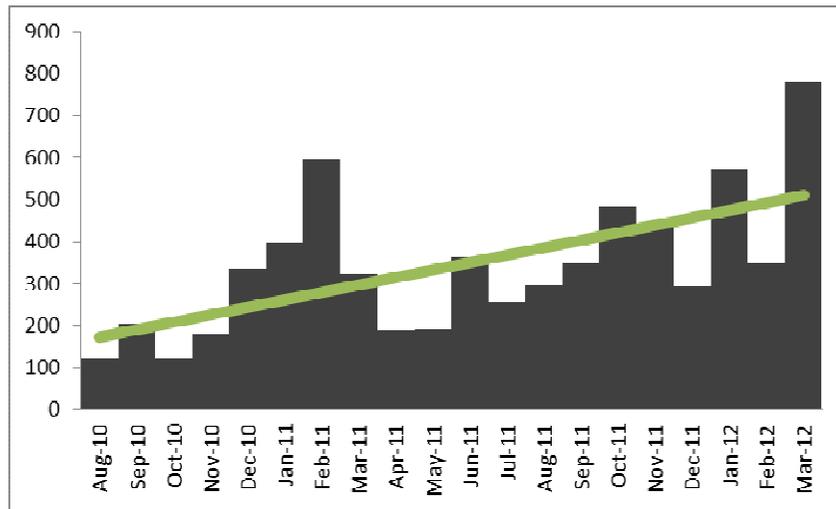→ action: forensic imaging + backups

→ result: bitstreams secure

Next step

→ assess: risk of format obsolescence (unknown)

→ action: characterise

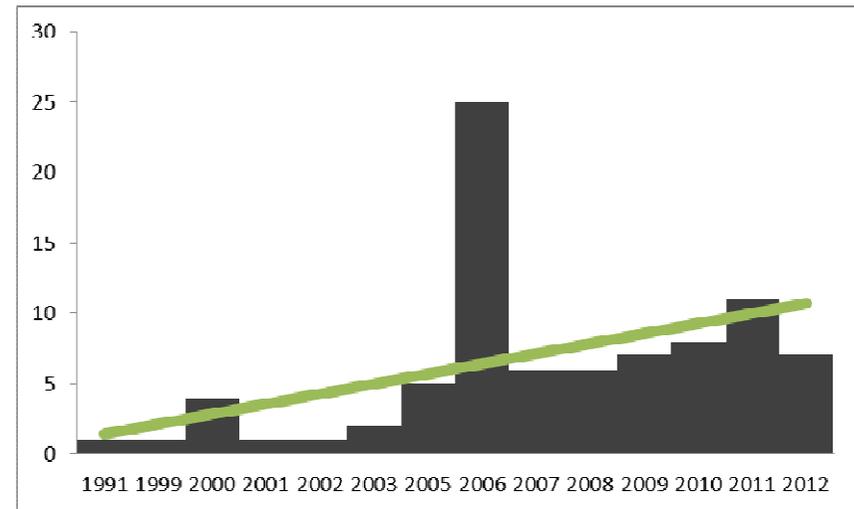→ result: low risk (known), revisit later

Planning: organisation perspective

- Increasing volume and diversity
  - Capacity planning vs format planning

Institutional repository: additions per month
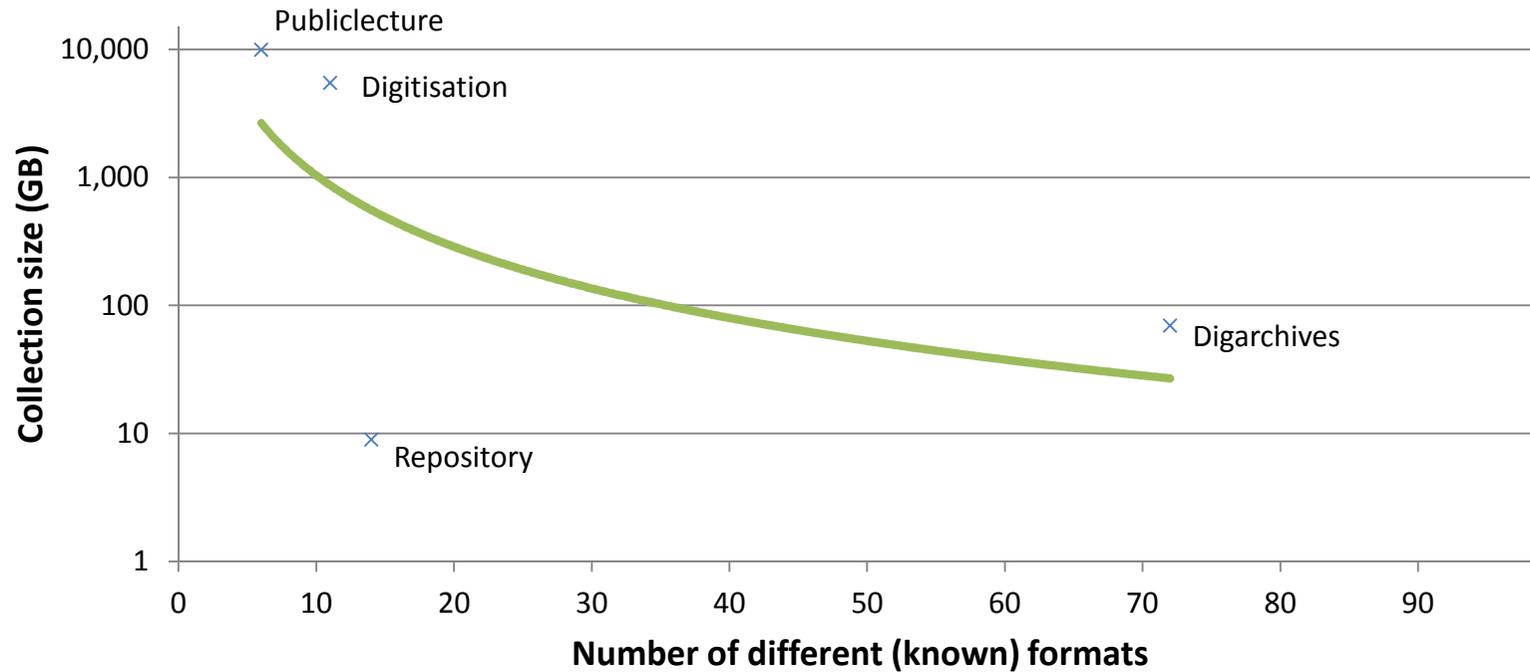

Archives: new hybrid or digital additions per year

LSE Library Services
British Library of Political and Economic Science

Planning: organisation perspective

- Increasing volume and diversity
  - Capacity planning vs format planning

Planning: roles and responsibilities

Senior Management
- Strategy
- Resources

Digital Library Team
- Innovation (dev/UX)
- Policy

Academic Services
- User/depositor liaison
- Collection development
- Info skills training

Archive Services
- User/depositor liaison
- Archival description

Collection Services
- Ingest/preservation
- Resource discovery
- Infrastructure (IT)

Planning: roles and responsibilities

Senior Management
- Strategy
- Resources

Digital Library Team
- Innovation (dev/UX)
- Policy

New posts

New skills

Academic Services
- User/depositor liaison
- Collection development
- Info skills training

Archive Services
- User/depositor liaison
- Archival description

Collection Services
- Ingest/preservation
- Resource discovery
- Infrastructure (IT)

New systems

~~Why it always goes wrong~~. Plan to adapt

There are known knowns; there are [digital collections] we know that we know.

There are known unknowns; that is to say there are [digital collections] that, we now know we don't know.

But there are also unknown unknowns – there are [digital collections] we do not know we don't know.

LSE Library Services
British Library of Political
and Economic Science

Final thoughts

- Planning for digital preservation is about <u>continuous improvement</u>, not monolithic 'solutions'.

- Plans must be based on <u>evidence</u>.

- Take small steps, use outcomes to <u>make the case</u> for the next step.

Doing nothing is the biggest risk.

LSE Library Services
British Library of Political and Economic Science

?

Ed Fay, Digital Library Manager

e.fay@lse.ac.uk | @digitalfay

LSE | **Library Services**
British Library of Political
and Economic Science