

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



Library Services

British Library of Political
and Economic Science

Digital Preservation Planning



Library Services

British Library of Political
and Economic Science

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



Library Services

British Library of Political
and Economic Science

Digital Preservation Planning

Two important aspects:

→ collection

characterisation, risk assessment, prioritisation
(workflows, testing, validation)

→ organisation

capacity planning, staff skills, sustainability



Library Services
British Library of Political
and Economic Science



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	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

User Generated Risk Register
Digital Collections

DRAMBORA interactive

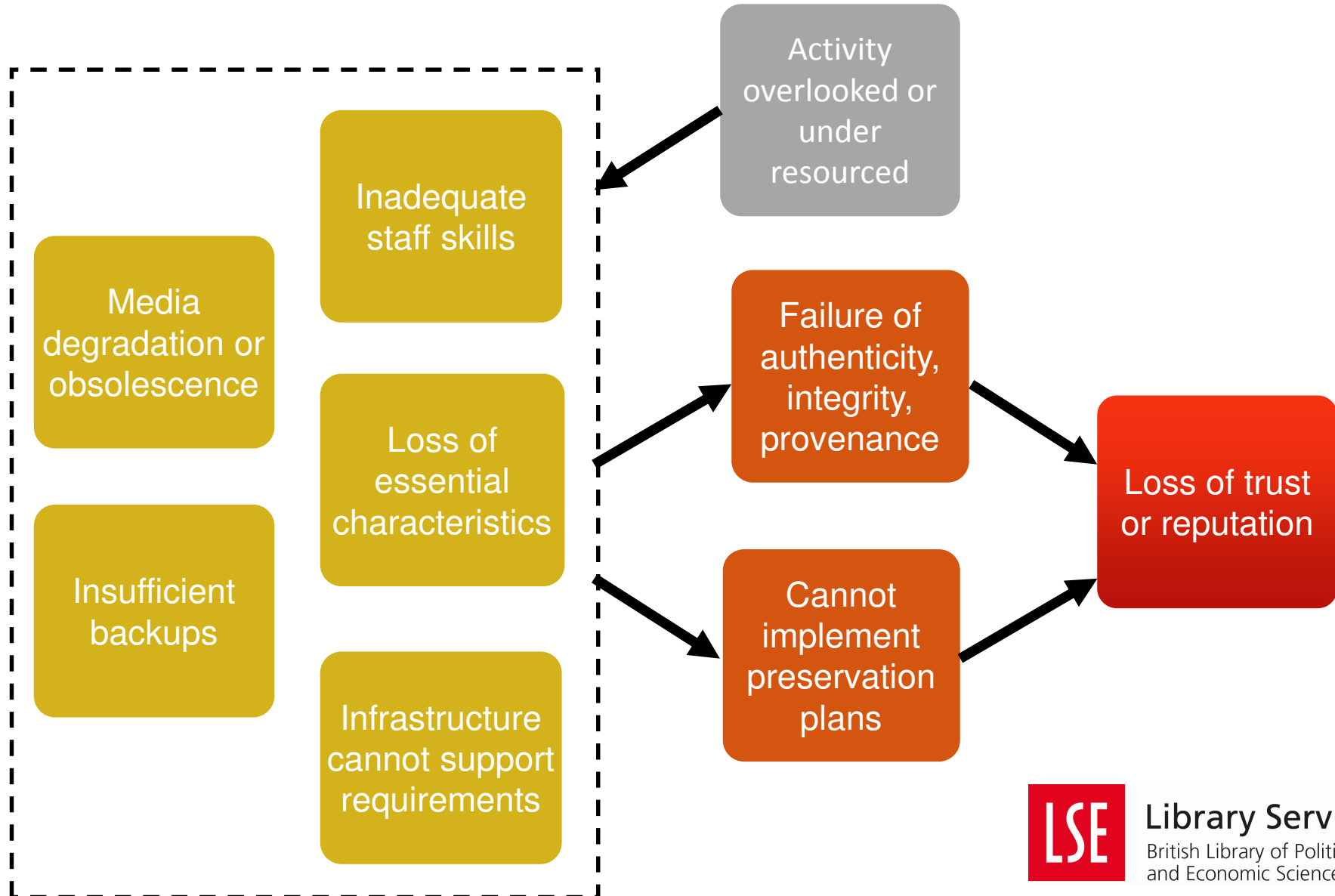
Risk Number 1

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

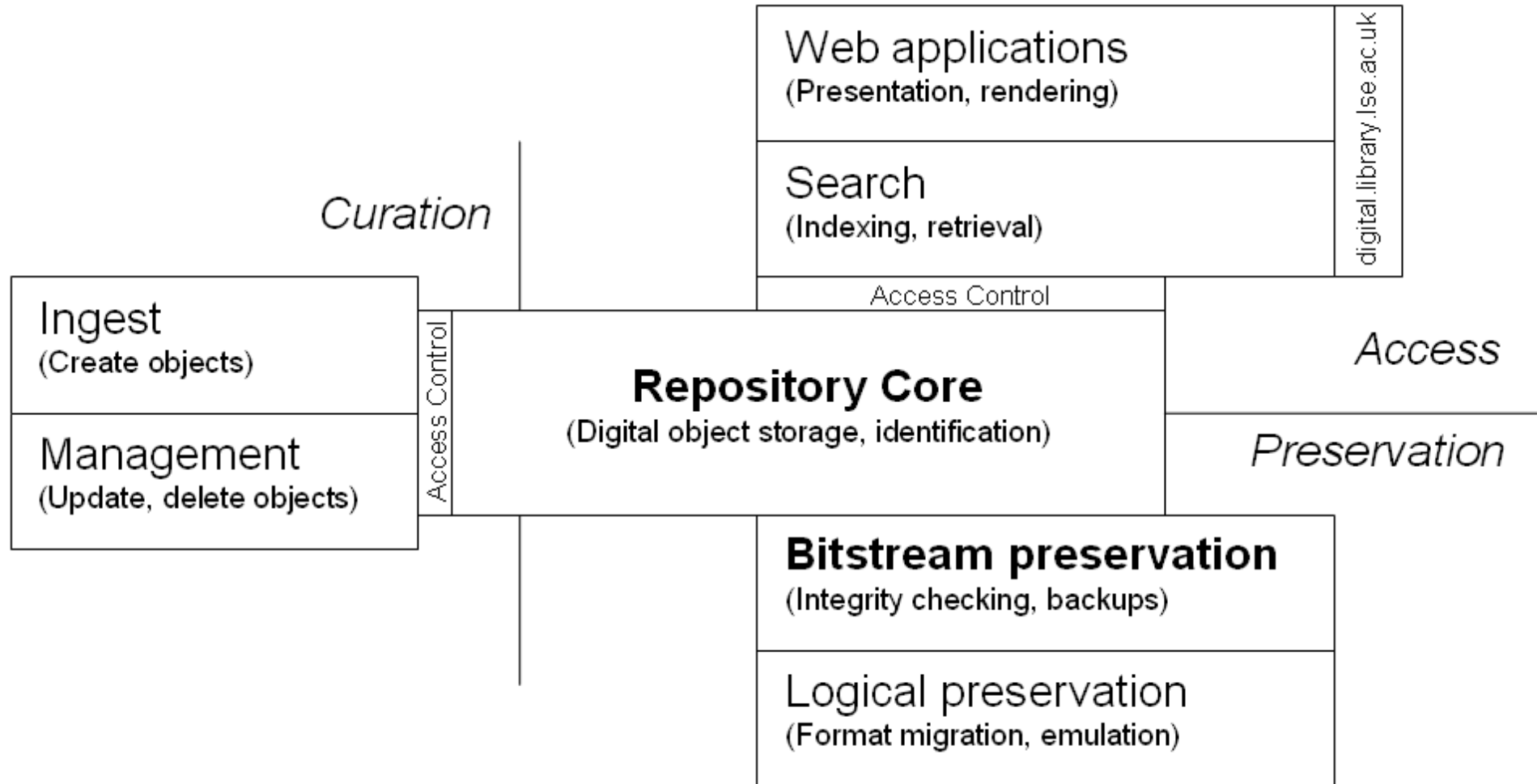


Library Services
British Library of Political
and Economic Science

Digital collections: risk assessment



Prioritisation

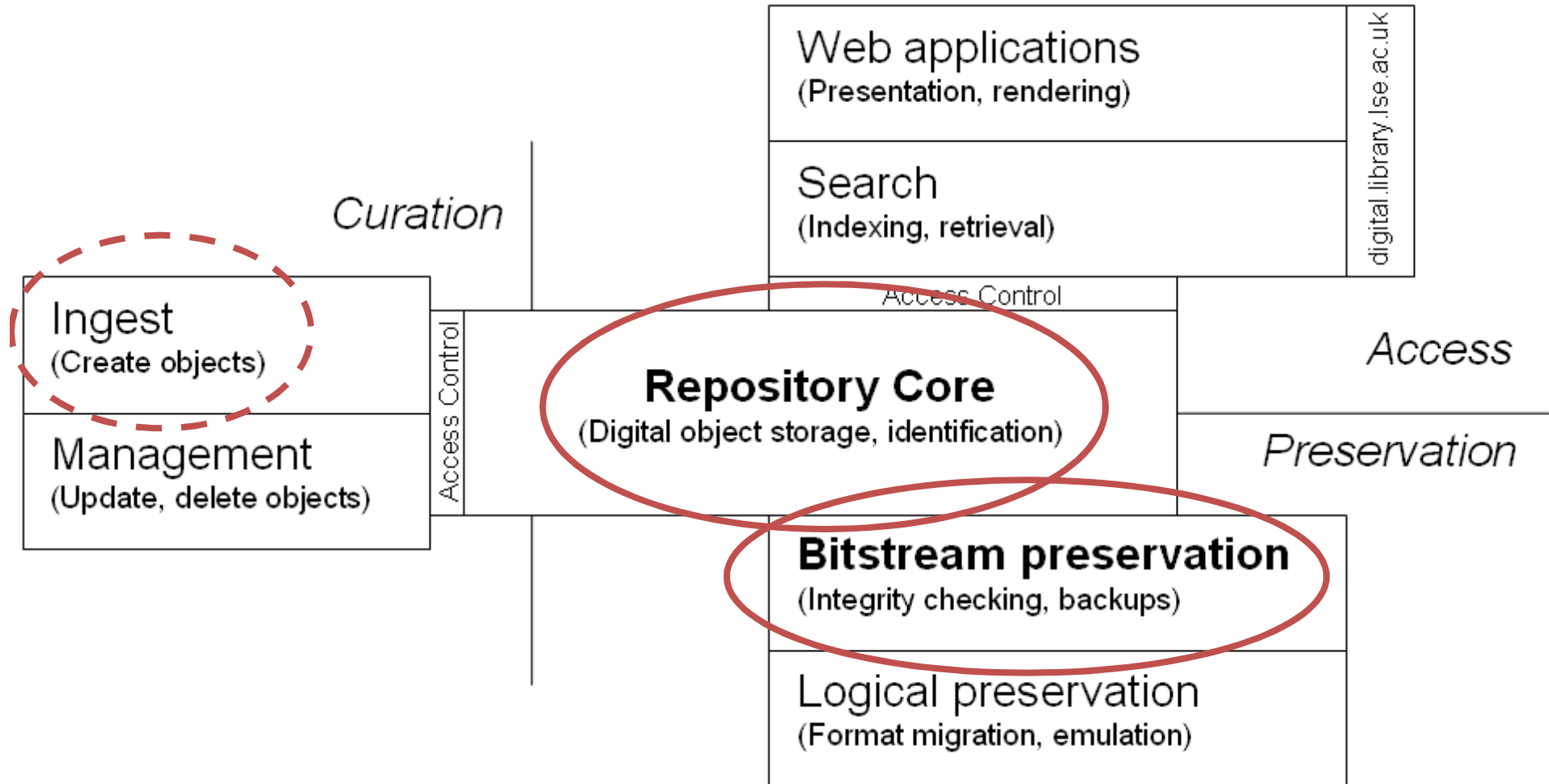


Phased implementation of
technical infrastructure, staff skills



Library Services
British Library of Political
and Economic Science

Prioritisation



Phased implementation of
technical infrastructure, staff skills



Library Services
British Library of Political
and Economic Science

Three examples

Collection	Format	Risk	Action
Legacy media	Floppy disks CD/DVDs	<ul style="list-style-type: none">• Media obsolescence• Format obsolescence	<ul style="list-style-type: none">• Forensic imaging• Backup disk images• Characterise formats
Digitisation	Mostly TIFF (c.11 total)	<ul style="list-style-type: none">• Separation of digital files from metadata• (Format diversity)	<ul style="list-style-type: none">• Full collection audit• Repository ingest
Public lectures	HD video Web video Audio	<ul style="list-style-type: none">• Separation of HD/Web versions• Large file sizes	<ul style="list-style-type: none">• Normalise formats (compress HD video)• Repository ingest



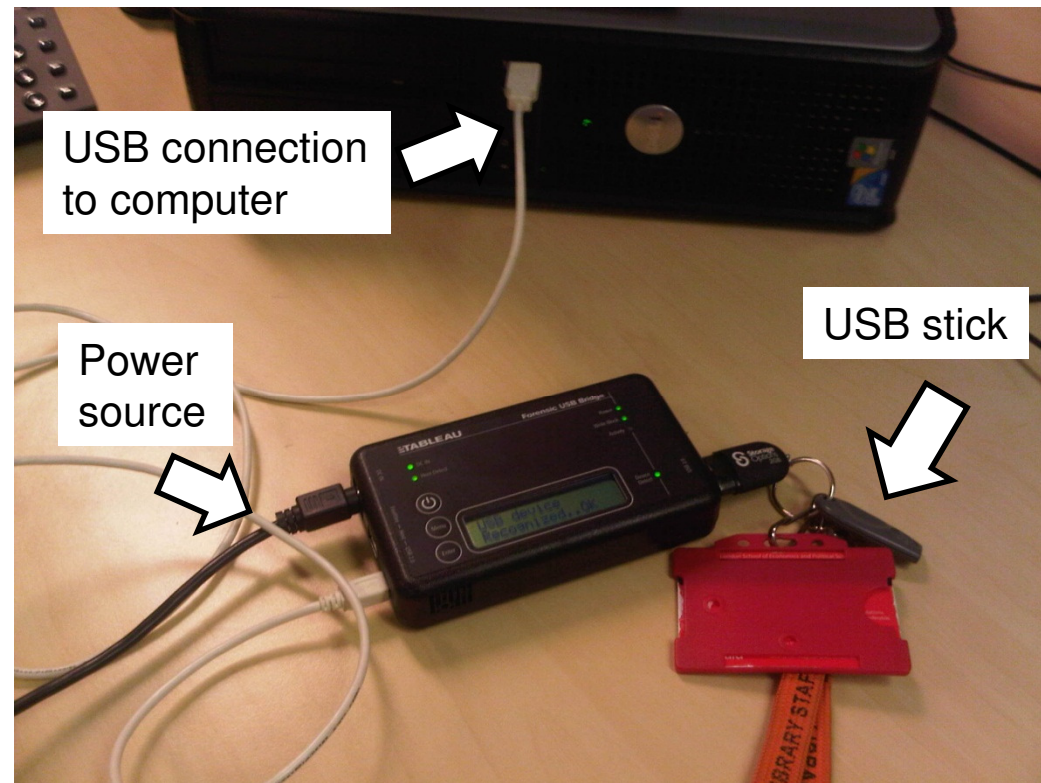
Library Services

British Library of Political
and Economic Science

Example: legacy media (archives)

Risk: media obsolescence

Action: forensic imaging



60 collections (hybrid and digital archives), total size 70GB
14,829 files, average 247 files per collection

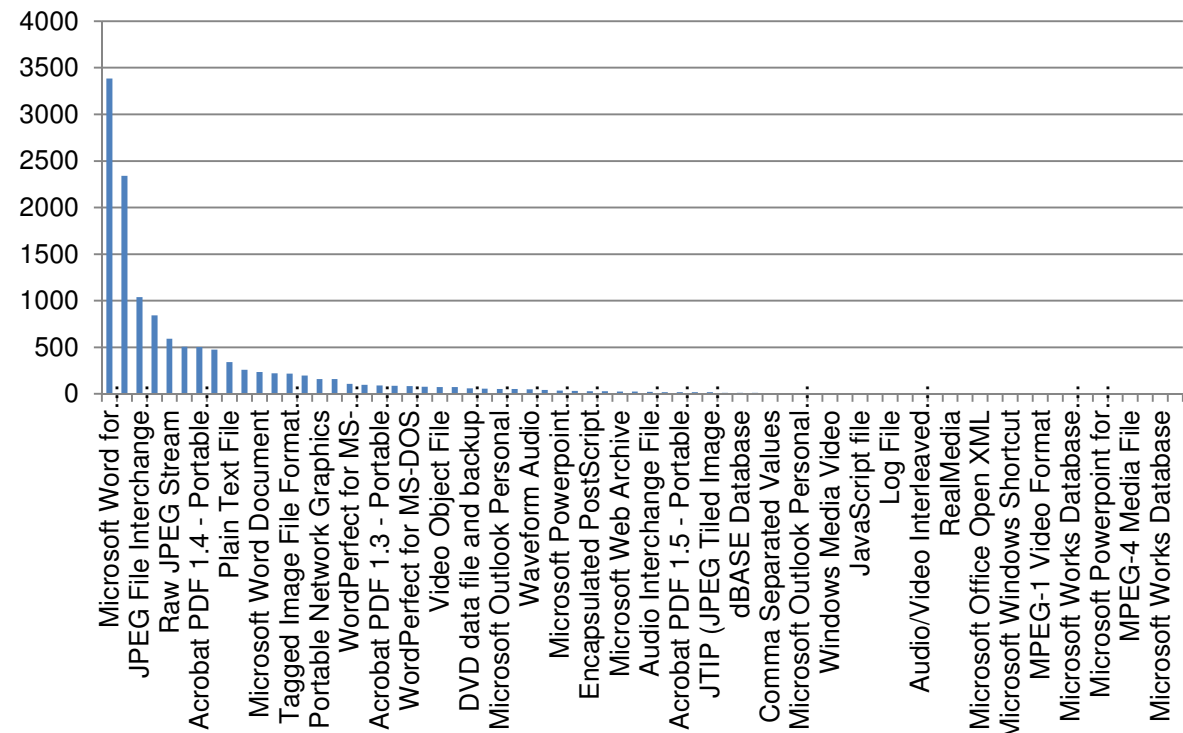
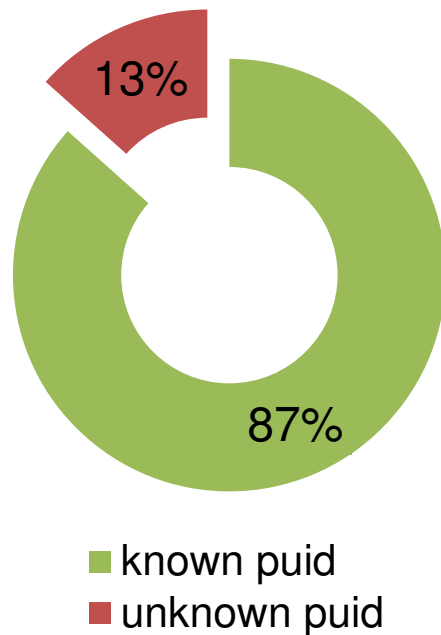


Library Services
British Library of Political
and Economic Science

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



Library Services

British Library of Political
and Economic Science

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

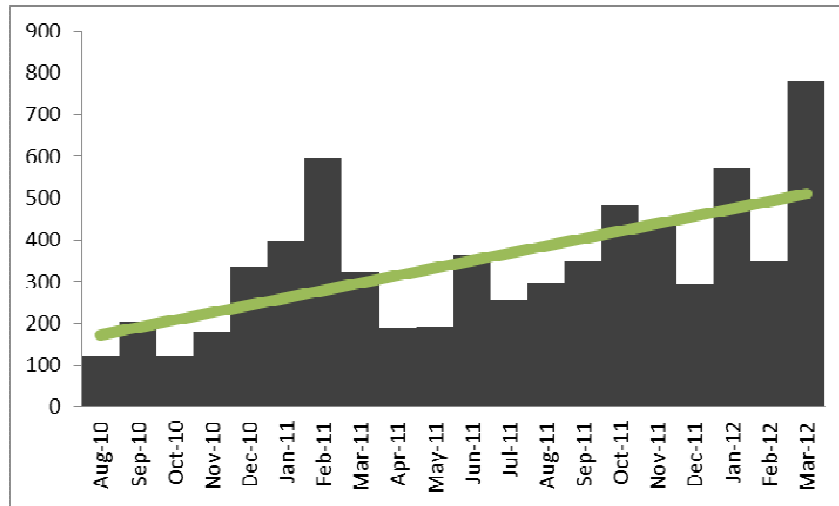


Library Services

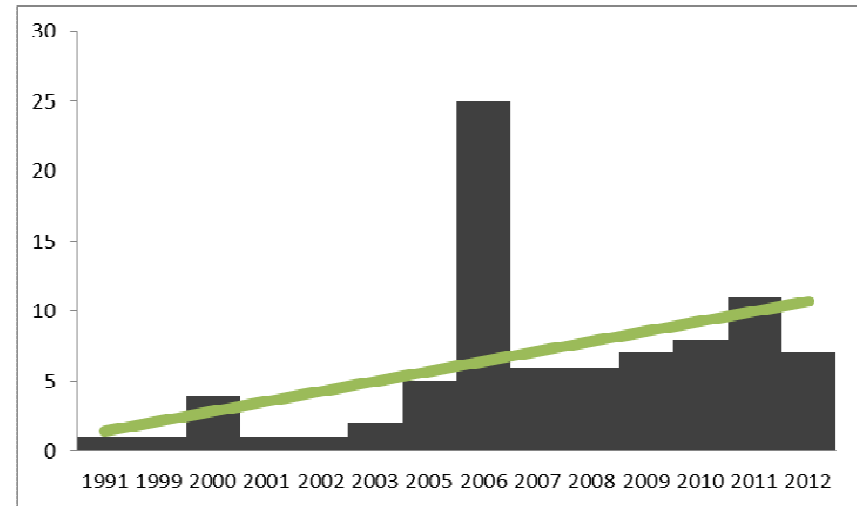
British Library of Political
and Economic Science

Planning: organisation perspective

- Increasing volume and diversity
 - Capacity planning vs format planning



Institutional repository: additions per month



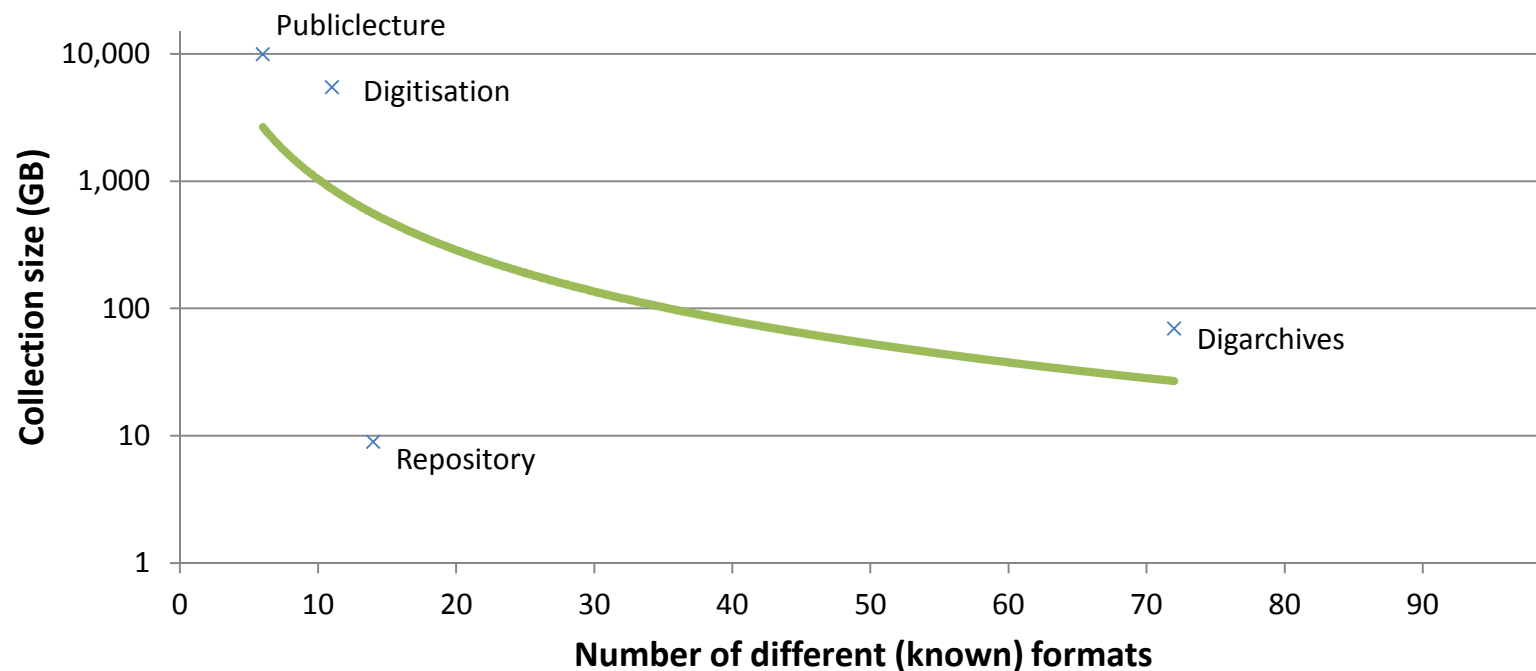
Archives: new hybrid or digital additions per year



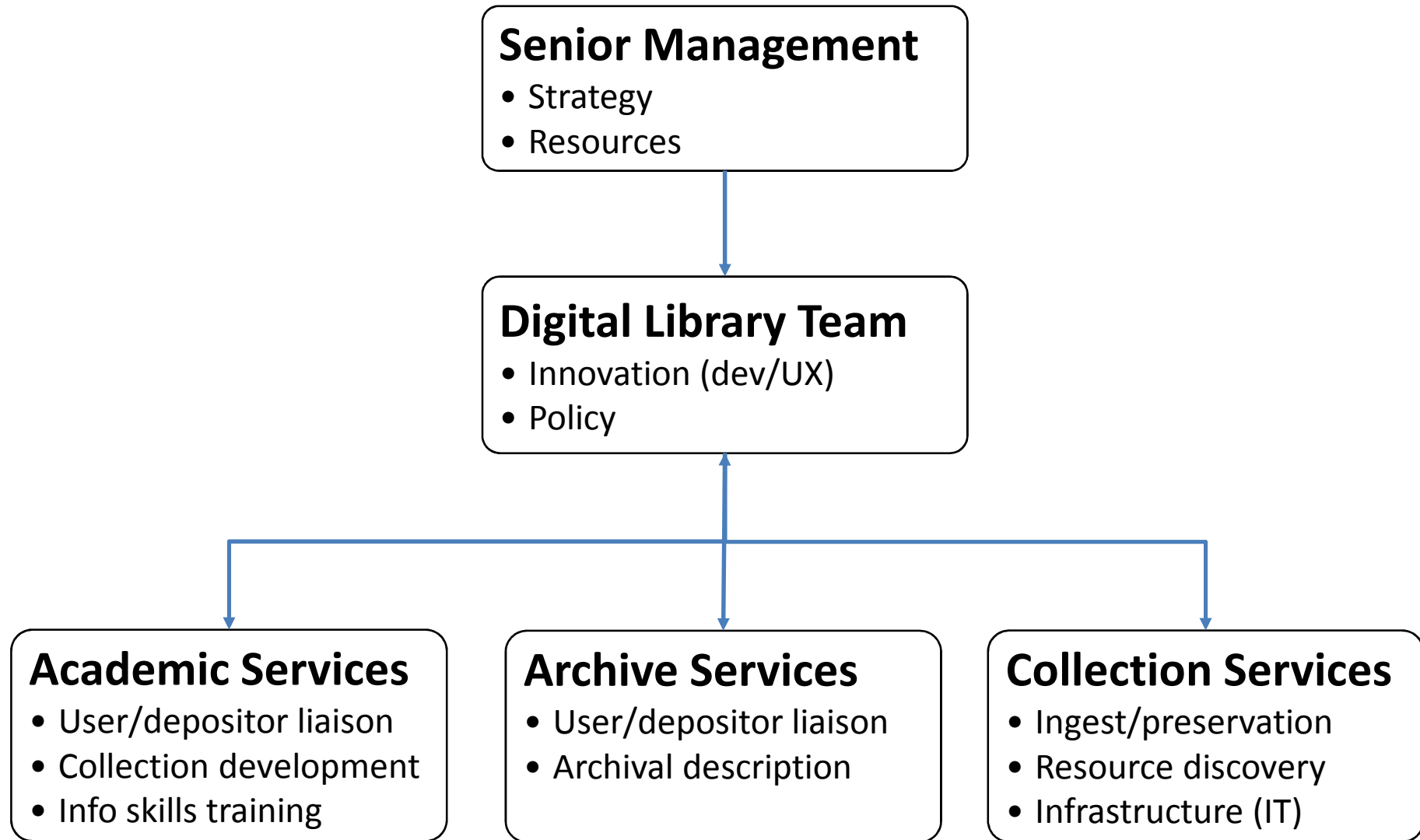
Library Services
British Library of Political
and Economic Science

Planning: organisation perspective

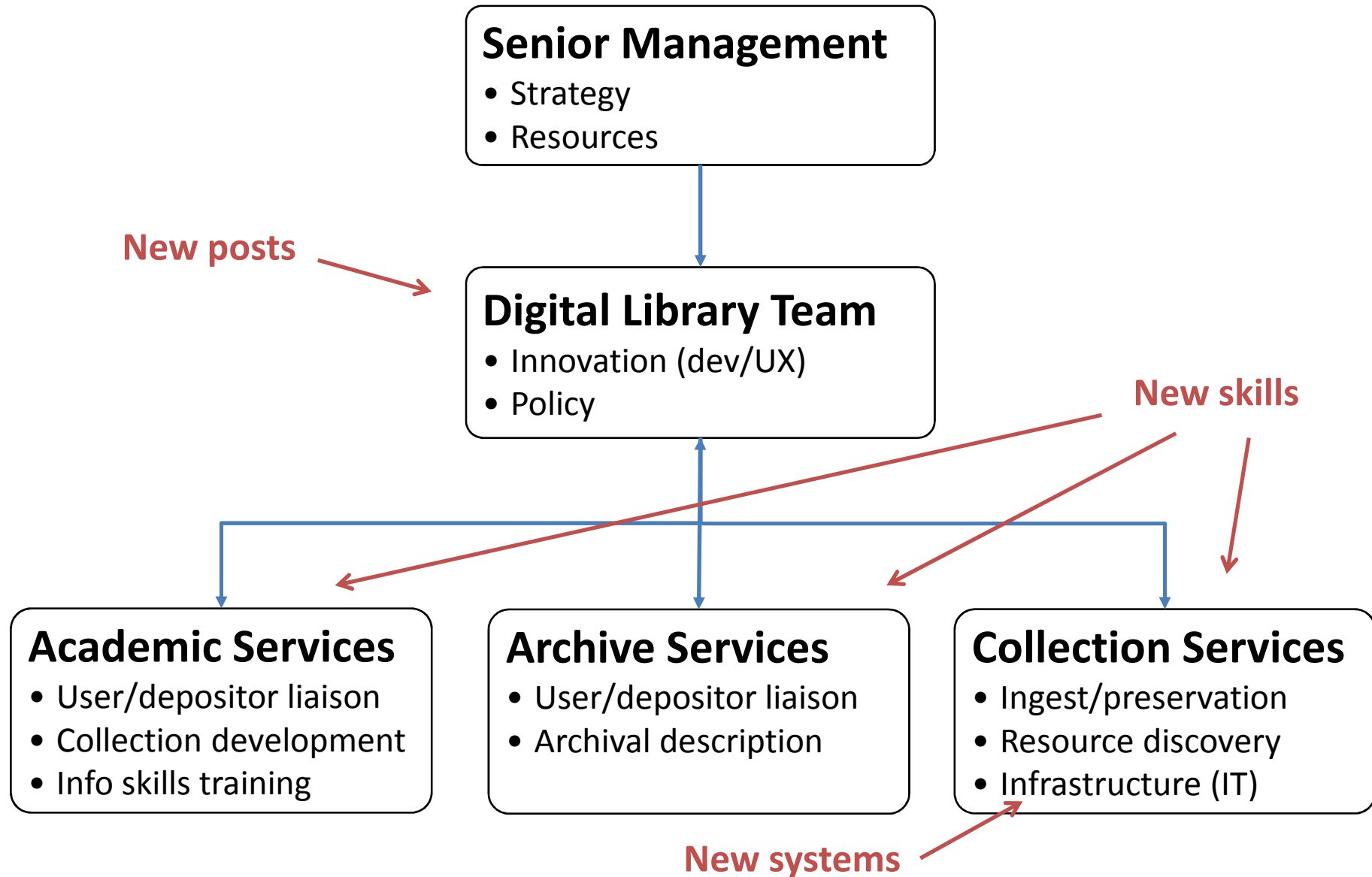
- Increasing volume and diversity
 - Capacity planning vs format planning



Planning: roles and responsibilities



Planning: roles and responsibilities



~~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.

United States Secretary of Defence, Donald Rumsfeld
12 February 2002



Library Services
British Library of Political
and Economic Science

Final thoughts

- Planning for digital preservation is about continuous improvement, not monolithic ‘solutions’.
- Plans must be based on evidence.
- Take small steps, use outcomes to make the case for the next step.

Doing nothing is the biggest risk.



Library Services
British Library of Political
and Economic Science



Ed Fay, Digital Library Manager
e.fay@lse.ac.uk | @digitalfay



Library Services

British Library of Political
and Economic Science