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
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...)

### User Generated Risk Register

<table>
<thead>
<tr>
<th>Collection</th>
<th>Identifier</th>
<th>Format</th>
<th>Size (MB)</th>
<th>Media</th>
<th>Location</th>
<th>Accession/Access</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exam papers</td>
<td></td>
<td>PDF</td>
<td>1126</td>
<td>Server/Tape Backup</td>
<td><a href="https://library-2.ise.ac.uk/protection-exams/">https://library-2.ise.ac.uk/protection-exams/</a></td>
<td></td>
</tr>
<tr>
<td>Research publications</td>
<td>LSERO</td>
<td>PDF</td>
<td>4712</td>
<td>Server/Tape Backup</td>
<td><a href="http://reprints.lse.ac.uk">http://reprints.lse.ac.uk</a></td>
<td></td>
</tr>
<tr>
<td>e-Theses</td>
<td>LSE07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Resource Archive</td>
<td>text/PDF/SB</td>
<td>60000</td>
<td>Server</td>
<td>ib-6</td>
<td>stable (once it has been</td>
<td></td>
</tr>
<tr>
<td>Paddy Ashdown diaries</td>
<td>K03E6</td>
<td>Word/Html</td>
<td>1638</td>
<td>Archives</td>
<td>Expecting more Arch</td>
<td></td>
</tr>
</tbody>
</table>

**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 irreversible 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**

- Insufficient backups
- Media degradation or obsolescence
- Inadequate staff skills
- Loss of essential characteristics
- Infrastructure cannot support requirements

Activity overlooked or under resourced

- Failure of authenticity, integrity, provenance
- Cannot implement preservation plans

Loss of trust or reputation
Prioritisation

Phased implementation of technical infrastructure, staff skills
Phased implementation of technical infrastructure, staff skills
# Three examples

<table>
<thead>
<tr>
<th>Collection</th>
<th>Format</th>
<th>Risk</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Legacy media</td>
<td>Floppy disks</td>
<td>• Media obsolescence</td>
<td>• Forensic imaging</td>
</tr>
<tr>
<td></td>
<td>CD/DVDs</td>
<td>• Format obsolescence</td>
<td>• Backup disk images</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Characterise formats</td>
</tr>
<tr>
<td>Digitisation</td>
<td>Mostly TIFF</td>
<td>• Separation of digital files from metadata</td>
<td>• Full collection audit</td>
</tr>
<tr>
<td></td>
<td>(c.11 total)</td>
<td>• (Format diversity)</td>
<td>• Repository ingest</td>
</tr>
<tr>
<td>Public lectures</td>
<td>HD video</td>
<td>• Separation of HD/Web versions</td>
<td>• Normalise formats</td>
</tr>
<tr>
<td></td>
<td>Web video</td>
<td></td>
<td>(compress HD video)</td>
</tr>
<tr>
<td></td>
<td>Audio</td>
<td>• Large file sizes</td>
<td>• Repository ingest</td>
</tr>
</tbody>
</table>
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
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
  o Capacity planning vs format planning

Institutional repository: additions per month

Archives: new hybrid or digital additions per year
Planning: organisation perspective

- Increasing volume and diversity
  - Capacity planning vs format planning

![Graph showing collection size vs number of different formats.]

- Publiclecture: 10,000 GB
- Digitisation: 1,000 GB
- Repository: 100 GB
- Digarchives: 10 GB

Collection size (GB) vs Number of different (known) formats

Library Services
British Library of Political and Economic Science
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

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 posts
New skills
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.

United States Secretary of Defence, Donald Rumsfeld
12 February 2002
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.