Costing Digital Preservation of research data – University of Birmingham

Patricia Herterich - @pherterich
Overview

- UoB’s Open Data Programme
- Costing for Digital Preservation
  - Business case 1
  - Current data services usage
  - Current cost recovery
  - Business case 2
UoB Open Data Programme

- Started in July 2016
- Membership from Planning Office, IT services, Library Services and Research Support Services
- Designed for 3 phases
- Several projects to cover all aspects of RDM
# UoB Open Data Programme

<table>
<thead>
<tr>
<th>Project</th>
<th>Phase 1</th>
<th>Phase 2</th>
<th>Phase 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Basic requirements</strong></td>
<td><strong>Implementing and embedding</strong></td>
<td><strong>July 2018 – July 2019</strong></td>
</tr>
<tr>
<td>Institutional repository</td>
<td>Options appraisal</td>
<td>Setup</td>
<td>Transition to BAU</td>
</tr>
<tr>
<td>Storage</td>
<td>Options appraisal &amp; Business case</td>
<td>Monitor storage on campus</td>
<td>Digital Preservation!</td>
</tr>
<tr>
<td>DMPOnline pilot</td>
<td>Pilot customisations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Advocacy &amp; Training</td>
<td>Initial Advocacy</td>
<td>Increased amount of training sessions</td>
<td>Integrated training Skills gap analysis</td>
</tr>
<tr>
<td>Policy</td>
<td>Rewrite institutional policy</td>
<td>Align policy and IR service</td>
<td></td>
</tr>
</tbody>
</table>
Costing for Digital Preservation – business case 1

- January 2017
- Setting up a data repository integrated with storage from Arkivum, no preservation system
- Costed for storage: 100 TB pay as you go for 5 years to start with - £181,200
- Business case to university central budget
How much data do we need to preserve?

- Survey by IT services (2017 data)
How much data do we need to preserve?

- Survey by IT services (2017 data)

![Research Data Volumes (Publications Data) chart](chart)

- Number of Respondents
- Data Sizes:
  - <0.5: 130
  - 0.5 - 1: 43
  - 1 - 5: 31
  - 5 - 10: 52
  - 10 - 100: 2
  - > 100: 1
  - I do not know: 72

_n = 331_
Current eData repository usage

- Repository of last resort
- 151 records
- Currently using 0.133 TB of storage
- Mix of proprietary files (.mat) and open files
## Cost recovery

<table>
<thead>
<tr>
<th>Tape Storage TBs</th>
<th>Price per TB for up to 10 Years with data held in two geographically separated data centres</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - 20 TBs</td>
<td>This is funded by the University from the Indirect Costs associated with research grants.</td>
</tr>
<tr>
<td>&gt; 20 TBs</td>
<td>£40 per TB. (charged in full to the project as a consumable)</td>
</tr>
</tbody>
</table>

[https://intranet.birmingham.ac.uk/it/teams/infrastructure/research/bear/research-data-service/RDS/research-bids-guidance.aspx](https://intranet.birmingham.ac.uk/it/teams/infrastructure/research/bear/research-data-service/RDS/research-bids-guidance.aspx)
Costs from grant applications?

IT survey 2019, n = 194

Why do you not include data storage costs?

- Use own storage
- Free IT storage sufficient
- Use of a subject repository
- Had not considered storage costs
- Other
Costing for Digital Preservation – preparation for business case 2

- Digital preservation working group in Library Services bringing together special collections and repository content
- Digital preservation policy
- Vendor presentations for preservation systems scheduled for June 2019
- Job description approved – waiting for funding to be able to recruit
Costing for Digital Preservation – preparation for business case 2

- Business case for a system based on
  - 42 GB publications data and metadata
  - 11 GB other outputs
  - 75 GB theses
  - 133 GB research data
  - 2.2 TB special collections material
Thank you!

@pherterich
p.s.herterich@bham.ac.uk