Open Data:
Data preservation
Data citation

DPC 'Open Data' Day
5th July 2013, York

http://archaeologydataservice.ac.uk
G8 Open Data Charter unveiled

Picture: FACUNDO ARRIZABALAGA/EPA
We’re all in the same boat!

Picture: AFP.GETTY
The Five Principles

1. Open Data by Default
2. Quality and Quantity
3. Usable by All
4. Releasing Data for improved Governance
5. Releasing Data for Innovation
ADS: Who we are and what we do (in a nutshell)

• Accredited digital archive for archaeological data
• Established in 1997, hosted by the University of York
• Cross-sectoral work from professors to the public
Open Data by Default

• (Almost) all ADS hosted data is ‘Open’, i.e. openly accessible
• Licenced to be reused (excluding commercial use)
• Embargoes; need to have a good reason
• Freedom of Information and the Data Protection Act
• RCUK expectations
Quality and Quantity

• Supporting research, learning and teaching with free, high quality and dependable digital resources
• ...and many of them
• The more you have the more potential they have
Useable by All

- Non-proprietary formats
- Accessible online to download
- Free
- Instructions for use
- Advice and guidance
Releasing Data for Improved Governance

- OASIS and an online archive of grey literature
- Work with national bodies, local planning authorities, museum services
Releasing Data for Innovation

Correlating radiocarbon dating and collagen degradation in dinosaur bones: Prof Matthew Collins

---

**Figure 23: Research efficiency benefits, by data centre**

<table>
<thead>
<tr>
<th>Benefit</th>
<th>ADS</th>
<th>BADC</th>
<th>CDS</th>
<th>ESDS</th>
<th>NGDC</th>
</tr>
</thead>
<tbody>
<tr>
<td>It has reduced the time required for data acquisition / processing</td>
<td>79%</td>
<td>68%</td>
<td>76%</td>
<td>80%</td>
<td>92%</td>
</tr>
<tr>
<td>N=</td>
<td>67</td>
<td>618</td>
<td>174</td>
<td>262</td>
<td>36</td>
</tr>
<tr>
<td>It has improved the efficiency of research</td>
<td>79%</td>
<td>62%</td>
<td>75%</td>
<td>67%</td>
<td>89%</td>
</tr>
<tr>
<td>N=</td>
<td>67</td>
<td>622</td>
<td>179</td>
<td>261</td>
<td>36</td>
</tr>
<tr>
<td>It has reduced the financial cost of data acquisition / processing</td>
<td>65%</td>
<td>62%</td>
<td>61%</td>
<td>73%</td>
<td>78%</td>
</tr>
<tr>
<td>N=</td>
<td>66</td>
<td>612</td>
<td>175</td>
<td>262</td>
<td>36</td>
</tr>
<tr>
<td>It has reduced duplication of effort (i.e. unnecessary recreation of data)</td>
<td>57%</td>
<td>57%</td>
<td>68%</td>
<td>62%</td>
<td>81%</td>
</tr>
<tr>
<td>N=</td>
<td>65</td>
<td>609</td>
<td>176</td>
<td>258</td>
<td>36</td>
</tr>
<tr>
<td>It has enabled me to undertake a greater quantity of research</td>
<td>52%</td>
<td>42%</td>
<td>50%</td>
<td>54%</td>
<td>77%</td>
</tr>
<tr>
<td>N=</td>
<td>63</td>
<td>614</td>
<td>176</td>
<td>257</td>
<td>35</td>
</tr>
</tbody>
</table>

Source: Technopolis survey of data centre users, January 2010

http://archaeologydataservice.ac.uk
Practicalities: Data Preservation

• How do we ensure the longevity of this ‘open Data’
• Not very ‘Open’ if it’s not there in 10 years (if it needs to be)
• Not very ‘Open’ is you can’t reuse it
• Metadata requirements and file formats
• Active, long-term management, planning for the future
Practicalities: Data Citation

- How do we know this is useful?
- Digital Object Identifiers (DOIs)
- DataCite and the British Library
Are we going in the same direction?

Picture: AFP/GETTY
Are we doing this because we have to?

- … or because it’s a good thing.
- The IMPACT project to develop and refine a range of methods to measure the costs, benefits, economic impacts and value of the ADS
Is it worth it?

Increase in returns on investment over 30 years: £1 investment provides up to £8.30 return