Archiving research data at the University of Bath

Alex Ball
University of Bath

28 September 2017
University of Bath

- Established 1966

- Three Faculties and one School:
  - Faculty of Engineering and Design
  - Faculty of Humanities and Social Sciences
  - Faculty of Science
  - School of Management

- Roughly 17,300 students
  - of which about 1600 (9%) postgraduate researchers
Timeline

2011   Research 360
   ▶ Active research data storage provision
   ▶ Data management plan (DMP) templates
   ▶ Online guidance

2013   Research Data Service, Library
   ▶ University-wide training programme
   ▶ Research data helpdesk
   ▶ DMP checking service

2014   Adoption of Research Data Policy

2015   Launch of Research Data Archive
## A tale of two systems

<table>
<thead>
<tr>
<th>CRIS (Pure)</th>
<th>Research Data Archive (EPrints)</th>
</tr>
</thead>
<tbody>
<tr>
<td>✔ Used for managing REF returns</td>
<td>✗ Not hooked up to University data sources (currently)</td>
</tr>
<tr>
<td>✔ Used for generating CVs</td>
<td>✔ Handles rich dataset details</td>
</tr>
<tr>
<td>✔ Handles links with people, projects, papers, funders, etc.</td>
<td>✔ Highly configurable</td>
</tr>
<tr>
<td>✗ Handles only basic dataset details</td>
<td>✔ Archival quality backend storage</td>
</tr>
<tr>
<td>✗ Limited scope for configuration</td>
<td></td>
</tr>
</tbody>
</table>

---

---
Network of dataset records

- CRIS (Pure)
- External Data Archive
- Research Data Archive: Metadata + Data
- Research Data Archive: Metadata only
Archiving process (1)

**Researcher**
- Choose Bath RD Archive
- Deposit externally
- Register data in Pure
- Fix problems

**Administrator**
- Log as request
- Validate record
Initial validation

- Title
- Creators
- Dates
- Responsible archive (‘Publisher’)
- Project (thereby funder)
Archiving process (2)

**Researcher**
- Add data access statement to paper

**Administrator**
1. Validate record
2. Import record to RD Archive
3. Send for editing
   - Predict DOI
Using DOIs

We prefer our datasets to have DOIs:

- Recognized and understood by researchers
- Seen as ‘mark of quality’
- Makes datasets easier to cite
- Makes informal references easier to track
- Participation in wider infrastructure
  - DCI and OpenAIRE harvest metadata via DataCite
  - Sharing metadata with CrossRef, ORCID
Archiving process (3)

**Researcher**
- Add data access statement to paper
- Complete record
- Upload data
- Deposit record

**Administrator**
- Review record
Detailed record review

Abstract  Is it detailed enough?

Documentation  Is it detailed enough?

Software  Are all specialist packages adequately described?

Sensitivity  Do the access settings match the level of sensitivity?  
Is the level of anonymization adequate?

Licences  Have they been applied?
Archiving process (4)

- **External**
  - Paper published
  - Dataset published

- **Researcher**
  - Fix problems
  - Notify readiness

- **Administrator**
  - Review record
  - Add to publication queue
  - Make record live
  - Mint DOI
  - Update Pure record
<table>
<thead>
<tr>
<th>Metadata Element</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creator</td>
<td></td>
</tr>
<tr>
<td>Title</td>
<td></td>
</tr>
<tr>
<td>Publication Year</td>
<td></td>
</tr>
<tr>
<td>Publisher</td>
<td></td>
</tr>
<tr>
<td>Identifier</td>
<td></td>
</tr>
<tr>
<td>Resource Type</td>
<td></td>
</tr>
<tr>
<td>Subject</td>
<td></td>
</tr>
<tr>
<td>Description × 3</td>
<td></td>
</tr>
<tr>
<td>Contributor</td>
<td></td>
</tr>
<tr>
<td>Date × 1</td>
<td></td>
</tr>
<tr>
<td>Geo-location</td>
<td></td>
</tr>
<tr>
<td>Related ID × 5</td>
<td></td>
</tr>
<tr>
<td>Alternate ID</td>
<td></td>
</tr>
<tr>
<td>Format</td>
<td></td>
</tr>
<tr>
<td>Version</td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td></td>
</tr>
<tr>
<td>Rights</td>
<td></td>
</tr>
<tr>
<td>Language</td>
<td></td>
</tr>
<tr>
<td>Funding Ref</td>
<td></td>
</tr>
</tbody>
</table>
Sharing data

- Most of our datasets are open access.
- Datasets with non-digital data or a restricted file have a request button.
- We can handle embargoes.
- For handling access requests, we follow instructions from depositors:
  - for some, we have criteria for granting access;
  - for others, we refer the request onwards.
Advocacy and training

- We discuss the archiving process at departmental meetings.
- Several of our training courses deal with archiving:
  - Introduction to Research Data Management
  - Research Data Management: Planning
  - Research Data Management: Archiving and Sharing
- Our Web pages have detailed instructions for filling out the forms.
- We provide help by email and face-to-face.
Outstanding issues

Near term

› Allowing researchers to predict their own DOIs
› Tracking where researchers archive externally
› Discovering when linked papers are published
› Checking for data access statements

Long term

› Tracking impact of datasets
› Supporting discipline-specific metadata fully
Recommendations

We find these work well:

- Insisting on a good abstract → **Findable, Assessable**
- Insisting on a good methodology → **Reproducible**
- Encouraging use of licences → **Reusable**
- Minting DOIs → **Citable, Traceable**
- Providing full metadata to DataCite → **Findable, Accessible, Interoperable**
Thank you for your attention