Sustaining the momentum, moving the DataVault project to a service

Claire Knowles (University of Edinburgh) @cgknowles

CC-BY
DataVault within the Research Data Services System portfolio
<table>
<thead>
<tr>
<th>Funders</th>
<th>RDMA</th>
<th>DMS</th>
<th>Grant Widening</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alzheimer's Society</td>
<td>N</td>
<td>N</td>
<td>N/A</td>
<td>No mention of RDM or data sharing in any of the applicant guidance. Open Access for publications is required</td>
</tr>
<tr>
<td>Anatomical Society</td>
<td>N</td>
<td>N</td>
<td>N/A</td>
<td>No mention of RDM or data sharing in any of the applicant guidance.</td>
</tr>
<tr>
<td>Arthritis Research UK</td>
<td>N</td>
<td>N</td>
<td>N/A</td>
<td>No mention of RDM or data sharing in any of the applicant guidance.</td>
</tr>
<tr>
<td>AXA Research Fund</td>
<td>N</td>
<td>N</td>
<td>N/A</td>
<td>The AXA Research Fund supports the basic principle of Open Access to research data. Nevertheless, considering the fast evolution of current debates and the diversity of applicants, all its funded researchers are expected to follow best practices: * By retaining files of all the research data they have produced and used during the course of their work, and that they be prepared to share these data with other researchers and public (with a clear data use agreement) * By seeking to establish and practice good research data management in accordance with the current best practices in their respective fields and to share their data with a wider audience. Researchers are expected to ensure that data are maintained for a period of 10 years after the completion of the research project in suitable accessible formats using established methods.</td>
</tr>
<tr>
<td>BBSRC</td>
<td>Y</td>
<td>Y</td>
<td>10 years</td>
<td>Researchers are expected to ensure that data are maintained for a period of 10 years after the completion of the research project in suitable accessible formats using established methods. Using data management plans to ensure that data are accessible to the wider research community. No grant applicant guidance available.</td>
</tr>
<tr>
<td>BRET Trust</td>
<td>N</td>
<td>N</td>
<td>N/A</td>
<td>No grant applicant guidance available</td>
</tr>
<tr>
<td>British Academy</td>
<td>N</td>
<td>N</td>
<td>N/A</td>
<td>No mention of RDM or data sharing in any of the applicant guidance.</td>
</tr>
<tr>
<td>British Council</td>
<td>N</td>
<td>N</td>
<td>N/A</td>
<td>No grant applicant guidance available</td>
</tr>
<tr>
<td>British Heart Foundation</td>
<td>N</td>
<td>N</td>
<td>N/A</td>
<td>No mention of RDM or data sharing in any of the applicant guidance.</td>
</tr>
<tr>
<td>British Society for Neuroendocrinology</td>
<td>N</td>
<td>N</td>
<td>N/A</td>
<td>No mention of RDM or data sharing in any of the applicant guidance.</td>
</tr>
<tr>
<td>Cancer Research UK</td>
<td>Y</td>
<td>Y</td>
<td>5 years</td>
<td>Once the funding for a project has ceased researchers should preserve all data resulting from that grant to ensure that data can be used for followup or new studies. We expect researchers to use data management plans to ensure that data are accessible to the wider research community. No grant applicant guidance available.</td>
</tr>
<tr>
<td>Carnegie Trust</td>
<td>N</td>
<td>N</td>
<td>N/A</td>
<td>No grant applicant guidance available</td>
</tr>
<tr>
<td>Carnegie Trust for the Universities of Scotland</td>
<td>N</td>
<td>N</td>
<td>N/A</td>
<td>26. The Trust does not currently impose any data-sharing or data-storage requirements on recipients of its Grants, but may reconsider its policy in this respect. No grant applicant guidance available.</td>
</tr>
<tr>
<td>Chest, Heart and Stroke Scotland</td>
<td>N</td>
<td>N</td>
<td>N/A</td>
<td>No grant applicant guidance available</td>
</tr>
<tr>
<td>Chief Scientist Office</td>
<td>N</td>
<td>Y</td>
<td>5 years</td>
<td>16. Preserving and Sharing Research Data</td>
</tr>
<tr>
<td>CHILDREN with CANCER UK</td>
<td>N</td>
<td>N</td>
<td>N/A</td>
<td>16.1 CSO, in common with other public research funders, strongly encourages the sharing of data from research it supports. Where the data may be of interest to researchers, they should be preserved in a way that encourages other analysts to use them. The best method for ensuring this is to deposit the data with full supporting documentation in a public repository. 13.8.</td>
</tr>
<tr>
<td>Daphne Jackson Trust</td>
<td>N</td>
<td>N</td>
<td>N/A</td>
<td>16.2 CSO recognises that the original investigator has a right to a limited period of exclusive use of the data, that secondary analyses may be most fruitfully conducted in close collaboration with the original investigator. 16.3 Whether or not the data are likely to be used for secondary analysis, the Chief Investigator must ensure that the raw data or results are stored for a minimum period of 5 years, as indicated in the notice of funding. No mention of RDM or data sharing in any of the applicant guidance.</td>
</tr>
<tr>
<td>No mention of RDM or data sharing in any of the applicant guidance.</td>
<td>No grant applicant guidance available.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Sustaining Research Data

- Know what data is in the Vault
- One Vault has 1 or more deposits
- Manifest of files is stored in the database and the deposit
- Deposit is in the LOC BagIt standard
- Checksum for each: deposit, chunk, file
- Geographical resilience, 3 copies
  - 1 in Edinburgh (UoE)
  - 1 outer Edinburgh (UoE)
  - 1 in the cloud, within the UK
Easy to use

- Integration with PURE
- Notifications of deposit and retrieval
- Metadata extraction
- Sharing of permissions
- Competitively priced
- Integration with single sign-on (LDAP)
DataVault

Archive your data securely, with automatic encryption

Transfer your data into long term storage and free up your working data store.

Why?
- You have data that needs to be protected from accidental change or deletion
- Your funder requires you to keep data safe for a set time period
- You are not able to share your data openly, but want to keep a copy that can be requested

Current Vaults

<table>
<thead>
<tr>
<th>Vault name</th>
<th>Deposit</th>
<th>Owner name</th>
<th>Date Created</th>
<th>Review Date</th>
<th>Deposit/Retrieval Completeness</th>
</tr>
</thead>
<tbody>
<tr>
<td>name here</td>
<td>Deposit</td>
<td>name here</td>
<td>00/00/0000</td>
<td>00/00/0000</td>
<td>Deposit</td>
</tr>
<tr>
<td>name here</td>
<td>Closing to new deposits in less than 3 months</td>
<td>name here</td>
<td>00/00/0000</td>
<td>00/00/0000</td>
<td>Retrieval</td>
</tr>
<tr>
<td>name here</td>
<td>Closed</td>
<td>name here</td>
<td>00/00/0000</td>
<td>00/00/0000</td>
<td>Retrieval with caution</td>
</tr>
</tbody>
</table>

Create and deposit in a new vault

1. Select Dataset  2. Create Vault  3. Deposit

You must record the details of your dataset in PURE. If you haven’t yet created a Pure dataset record containing the details of the data you’re about to deposit in the vault, please do so.

Please select the Pure record describing the data that will be contained in this vault from the list below:

Each record will have one of the following statuses:

- Deposit
- Retrieval
- Complete
- Incomplete
- Closing to new deposits in less than 3 months
- Closed with caution
Create new vault

You must record the details of your dataset in PURE. If you haven’t yet created a Pure dataset record containing the details of the data you’re about to deposit in the vault, please do so.

Please select the Pure record describing the data that will be contained in this vault from the list below:

- Biographies of individuals mentioned in the papers of Alexander Carmic

The ‘owner’ (PI) will be billed for any deposit, against the Grant ID they provided in relation to the vault, unless they have made other arrangements with Information Services Research Services. Rates are advertised on the Research Services charges page.

You will be aware that the data you place in the Vault belongs to the University of Edinburgh and you will only be able to retrieve the data as long as you are employed by the University of Edinburgh.

Vault Name
 Enter a descriptive name for the Vault e.g. the project or experiment.

Description

Retention Policy Read more about retention policies
 Please choose a retention policy
Engagement with researchers

- Research Data Steering Group
- Project Board
- Requirements gathering
- Interim Service use
- Training and support
- User Acceptance Testing
- College Reps on the Project Board
- Roadshows and workshops
- Support and help on the university webpages

https://www.ed.ac.uk/information-services/research-support/research-data-service/sharing-preserving-data/datavault
Costing

- Aim to be cheaper than active storage
- Compared costs to cloud storage offerings
- Free storage for first 100GB per user
- Initial cost is £500 per terabyte for 10 years
- Full economic costing is underway
Data curation services are the most likely to be externalized. Infrastructure and other technical resources needed to support Curation services are expensive to implement and manage and require expertise that may not be readily available on campus. Additionally, Curation services tend to be scalable, and therefore well-suited for provision as a shared service.
General Data Protection Regulation (GDPR) became law on 25th May 2018
- Research data and GDPR
- Encryption of all deposits
- Encrypt all
- Key management
- Chunking of deposits
Implications of encryption

- Deposit
- Lost data

- 3 copies

Sketches illustrating the concepts of deposit and lost data, as well as the idea of 3 copies.
Communications and Engagement

- Training
- Workshops
- Liaison Librarians
- Research Data Support
- College IT Officers
- Blog posts
- Videos
Is this digital preservation?

No

But …

- Multiple copies
- Learnt from best practice
- Backlog of research data before active preservation
Thanks

Claire Knowles
@cgknowles
claire.knowles@ed.ac.uk

https://github.com/DataVault/datavault

https://www.ed.ac.uk/information-services/research-support/research-data-service/sharing-preserving-data/datavault

https://figshare.com/articles/Sustaining_the_momentum_moving_the_DataVault_project_to_a_service/6453122