Where to Submit Data

CONUL Information Sheet

This Information Sheet is aimed at researchers who want to share / submit their research data to a Repository. It has been compiled by the CONUL Research Group.[i]

1) **Reasons for Sharing Data**
- **Impact & longevity**: Your data may be cited by others. Open publications and data receive more citations, over longer periods
- **Compliance**: Funders, publishers and institutions may require that you share your data
- **Transparency & quality**: Your findings can be replicated and compared with other studies
- **Collaboration**: creates opportunities for follow on research and collaboration
- **Re-use**: Your data can be used in novel ways. Data sharing facilitates re-use of your data for future / follow-on research and discovery as data collection can be funded / collected once, and used many times for a variety of purposes
- **Efficiency**: Data sharing is good research practice!

There may be reasons for not sharing your data e.g. privacy and confidentiality issues, commercial value of the data. Horizon 2020 has coined the phrase "As open as possible, as closed as necessary."

If you are unable to publicly share your data, consider the possibility that you may wish to make your data available internally to future researchers to facilitate follow-on research, and/or to create a metadata record in your chosen archives or repository. A metadata record will describe your data and aid others in knowing about it. In order to ensure this can happen you will need to manage your data – see [refer to local Data Management Checklist]

2) Advantages of a Data Repository or Archive

A data repository allows researchers to upload and publish their data, thereby making the data available for other researchers to re-use. Similarly, a data archive allows users to deposit and publish data but will generally offer greater levels of curation to community standards, have specific guidelines on what data can be deposited and is more likely to offer long-term preservation as a service. Sometimes the terms *data repositories* and *data archives* are used interchangeably.

A data repository or archive will provide services such as:

- Persistent identifier such as a "digital object identifier" or DOI; the presence of a DOI facilitates discoverability and citeability
- Assistance with metadata provision e.g. through the use of a template
- Allow you to apply a licence to your data
- Aid compliance with the FAIR data principles (data that are Findable, Accessible, Interoperable, and Reusable) as data are published online with appropriate metadata and are assigned a persistent identifier, see Jones, Sarah, & Grootveld, Marjan. (2017, November). How FAIR are your data?. Zenodo. http://doi.org/10.5281/zenodo.1065991
- Accept a wide range of data types
- Long-term access and, in some cases, long-term preservation
- Offer useful search, navigation and visualisation functionality
- Reach a wider audience of potential users
- Manage requests for data on your behalf

3) When to Select a Data Repository
- Choose early so that you can familiarise yourself with the repository's requirements
- Requirements may include depositing in certain file formats, or using a specific metadata standard; and the inclusion of documentation to help describe your data. Understanding such requirements will enable you to design your data collection materials for easier metadata and documentation creation.

4) Initial Questions

- Has a data repository been specified by my funder? E.g.
    - NERC Data Centre: http://www.nerc.ac.uk/research/sites/data/ for research funded by the UK's Natural Environment Research Council

- Has a data repository been specified by my publisher? E.g.
    - SpringerNature via their recommended repositories: http://www.springernature.com/gp/authors/research-data-policy/repositories/12327124
    - PLOS Recommended Data Repositories http://journals.plos.org/plosone/s/data-availability#loc-recommended-repositories
    - *Scientific Data* Recommended Data Repositories https://www.nature.com/sdata/policies/repositories

- Is there a disciplinary-specific community-recognised data repository I can submit my data to, thereby helping to preserve my data according to recognised standards in my discipline? E.g.:
    - Irish Social Science Data Archive: www.issda.ie
    - Cancer Imaging Archive: http://www.cancerimagingarchive.net/
    - PubChem: https://pubchem.ncbi.nlm.nih.gov/
    - PANGAEA: https://www.pangaea.de/

5) **Re3data.org**

- This is the primary place to locate a data repository. Search by specific research discipline and then filter by access categories, data usage licenses, whether the repository gives the data a persistent identifier etc. Re3data uses a series of symbols to indicate these key services. E.g.:
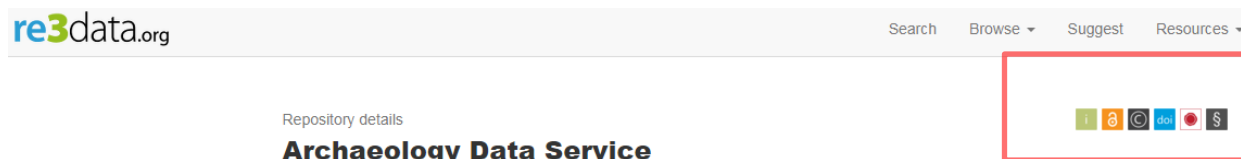
- To be registered in re3data.org a research data repository must:

    - be run by a legal entity, such as a sustainable institution (e.g. library, university)
    - clarify access conditions to the data and repository as well as the terms of use
    - have focus on research data

- https://www.re3data.org/suggest

6) Multidisciplinary Repositories

If there is no disciplinary-specific repository in your area select a general repository. These can handle a variety of different data types. Charges may apply but can be included in a funding application. Key general repositories are listed in the table below. This list is for information purposes only and is not exhaustive:

Repository	Link to Re3data
Dryad Digital Repository http://datadryad.org/	View Re3data.org entry
figshare https://figshare.com/	View Re3data.org entry
Dataverse https://dataverse.org/	View Re3data.org entry
Open Science Framework https://osf.io/	View Re3data.org entry
Zenodo https://zenodo.org/	View Re3data.org entry

7) **How to Select a Data Repository**

Ask:

- Is it reputable? Is it listed in Re3data thereby meeting their conditions of inclusion?
- Is it appropriate to my discipline?
- Will it take the data you want to deposit?
- Is there a size limit?
- Does it provide a DOI / persistent identifier?
- Does it provide guidance on how the data should be cited?
- Does it provide access control, where necessary, for your research data?
- Does it ensure long-term preservation / curation?
- Does it provide expert help with e.g. metadata provision, curation?
- Is there a charge?

Other questions may pertain depending on your requirements. For more information see the UK's Digital Curation Centre's checklist:
http://www.dcc.ac.uk/resources/how-guides-checklists/where-keep-research-data/where-keep-research-data

8) **Code**

GitHub is the main platform for hosting and reviewing code: https://github.com/

GitHub offers a number of advantages such as assigning DOIs (which facilitates discoverability and citeability) and allowing integration from Zenodo and FigShare repositories to enable the citing of your GitHub repository in academic literature.

9) **Local Services [include list specific to Institution]**

XXX Library can provide assistance in managing research data, including:

- Reviewing data management plans (such as those required by various funding organisations)
- Organising research data according to best practice
- Etc.

10) Further Information and Assistance

Further assistance and details of research data management services are available from:

XXX

Guides:

XXX

Date last updated:

---

[i] CONUL is the Irish Consortium of National and University Libraries (www.conul.ie) and is the representative body of research libraries in Ireland and Northern Ireland. The CONUL Research Group is one of the groups within CONUL, with a remit of exploring and promoting best practice and providing guidance and expertise in a wide range of activities, including research data management, open access, scholarly communications, research impact, digital repositories, digitisation and digital preservation, common infrastructures, and digital scholarship.