**SHARING YOUR DATA**

**by Simon Moss**

|  |
| --- |
| **Introduction** |

 Researchers often store their data electronically. For example, one researcher might construct an Excel spreadsheet that stores health information about 100 participants. Each row might correspond to one participant. Each column might represent one characteristic, such as the blood glucose levels or body fat of each person. Other researchers might construct spreadsheets that store information about birds, crops, rivers, and so forth.

 In the past, researchers usually stored these data on their own computer, inaccessible to anyone else. Nowadays, however, researchers are increasingly likely to store these data in databases that are accessible to other individuals. This document clarifies the benefits that researchers enjoy whenever they share their data. In addition, this document clarifies how researchers can share their data as well as access the data that other individuals share.

|  |
| --- |
| **Why should you share your data?** |

 Researchers often share their data for several reasons. First, if you share your data, your publications are more likely to be cited. That is, as research shows, if authors share their data, other researchers will often utilize these data and thus cite the paper. These citations significantly enhance the reputation and career prospects of researchers.

 Second, the data you share can also be cited. That is, the data file is assigned a doi or digital object identifier. Therefore, if other researchers utilize these data, they cite the data file, further enhancing your reputation.

 Third, many funding bodies, such as the ARC and NHMRC, discuss the importance of data sharing. They may, therefore, be more likely to fund research in which the investigators have constructed a plan to share data.

|  |
| --- |
| **How can you share your data and access other data?** |

 To share data, you need to identify a suitable repository on a relevant website. Which website is relevant depends on your discipline or field. This table presents some of the websites and the relevant disciplines.

|  |
| --- |
| **Websites in which you can share**  |
| **Google these websites**  | **Discipline** | **Other notes** |
| ada.edu.au | All Australian data |  |
| caNanoLab | Biomedical nanotechnology |  |
| cancer imaging archive | Medical images of cancer |  |
| ClinicalTrials.gov | Results of clinical studies |  |
| crystallography.net | Crystal structures of compounds and minerals |  |
| daac.gsfc.nasa.gov | Atmospheric composition, water & energy cycles, climate variability |  |
| datadryad.org | Mainly biology and ecology |  |
| **dataverse.org** | **Most disciplines** |  |
| DGVa | Genomic structural variations |  |
| Dryad | Most disciplines |  |
| ebi.ac.uk/ena | DNA sequences |  |
| EarthChem | Geochemical, geochronological, and petrological data |  |
| Electron Microscopy Data Bank | Electron microscopy density maps |  |
| **figshare.com** | **Most disciplines** |   |
| gbif.org | Biodiversity data |  |
| GenBank | DNA sequences |  |
| Gene Expression Omnibus | Genomics data |  |
| Harvard Dataverse | Most disciplines |  |
| Integrated Taxonomic Information System | Taxonomies of plants, animals, fungi, and microbes |  |
| LTER Network Data Portal | Ecological research |  |
| Marine Geosciences Data System | Marine geoscience data |  |
| Mouse Genome Informatics | Genetic, genomic, and biological data of mice |  |
| NCBI Taxonomy | Taxonomy of organisms |  |
| neuromorpho.org | Digitally reconstructed neurons |  |
| **NoMaD Repository** | **Most disciplines** |  |
| openICPSR | Social and behavioral sciences |  |
| osf.io  | Most disciplines | Also called the Open Science Framework |
| pangaea.de | Earth and environmental sciences |  |
| PCDDB | Protein circular dichroism spectra |  |
| PubChem | Chemical molecules and activities of these molecules |  |
| Research Data Australia | Data from Australian institutions |  |
| vectorbase.org | Allergies and infectious diseases | Genomic, phenotypic, and population-centric data on invertebrate vectors of human pathogens. |
| World Data Centre for Climate at DRKZ | Climate simulation data |  |
| wwpdb.org | Protein data bank |  |

**Example 1: How to use dataverse**

To use dataverse, you first need to establish an account. To achieve this goal

* Visit <http://dataverse.harvard.edu>
* Click “Sign Up”, near the top right hand corner.
* Enter your details, such as your name and institution, and finally click “Create Account”.

To upload your data file

* Click “Add a dataset”—an option that appears near the top left.
* Assign a title to your data, often the name of your study
* Complete any personal details that were not populated automatically
* Construct a brief description of your data in the relevant space. You might summarize the population as well as the key fields or measures, for example.
* Specify the “Subject”
* In the “Related Publication” box, enter the citation or an identifying code, such as the doi
* Click “Select Files to Add”—an option that appears towards the end of this page—and then upload your data file.
* Then click “Save Dataset” at the end of this page and finally “Publish” towards the top.

**Example 2: How to use figshare**

 In contrast, suppose you wanted to share an Excel data file in figshare.com. To complete this task, you would simply

* Sign up to generate an account
* Press “upload”
* Then Browse your computer to identify the data file.

 Likewise, if you wanted to search the data files that other people have uploaded, you could simply type relevant keywords in the “search on figshare” box. For the other websites, the procedure is usually similar. Finally, the next table presents some other useful websites

|  |
| --- |
| **Other useful websites**  |
| **Website** | **Purpose** | **Other notes** |
| www.ands.org.au | The homepage of the Australian National Data Service. This website summarizes everything you need to know about data sharing |  |
| re3data.org | Enables you to search other repositories of data files on any topic |  |
| researchdata.ands.org.au | Enables you to search all repositories of data files in Australia |   |

|  |
| --- |
| **The ethics of data sharing** |

 Sometimes, data sharing is ethical and indeed appropriate. In other circumstances, data sharing is unethical and inappropriate. Several acts and guidelines—such as the Privacy Act 1988, the Australian Code for the Responsible Conduct of Research, and the National Statement on Ethical Conduct in Human Research—offer insights into data sharing. In short

* in general, the data you share should be de-identified. That is, you should delete any information that identifies specific people or organizations.
* in the plain language statement, you should inform participants the data might be shared and utilized by other researchers—but only after the data is de-identified.
* if the data cannot be de-identified, you could indicate that only specific individuals can access and utilize the data. That is, some databases permit mediated access, in which only relevant people or organizations can access the data.

In the near future, besides sharing data, researchers will also be expected to convey the specific procedures they utilized to collect and to analyze data. For example, even their R or SPSS codes will need to be shared.

|  |
| --- |
| **Licences** |

 When you share data, you are usually granted the option to choose a license. A license specifies how the data can be used. For example, you can specify whether

* users of these data must acknowledge the source, called attribution
* users can use these data for commercial purposes
* users can transform or extend the data in some way
* users can share or distribute the data

 This table specifies the types of licenses you can choose. Furthermore, this table explains the terms of each license and presents the corresponding symbol.

|  |
| --- |
| **License types**  |
| **Type** | **Explanation** | **Symbol** |
| CC-BY: Attribution | * Sharing: Users can redistribute the data file freely
* Adapt: Users can redistribute a modified version of these data file—such as with additional data—even for commercial purposes
* Acknowledge: Users must acknowledge the source of these data
 |  ` |
|  CC-BY-SA: Attribution Shared Alike | * Same as CC-BY except, when users share the data, they cannot change the license agreements
 |   |
| CC-BY-NDAttribution-No derivatives  | * Same as CC-BY except users cannot redistribute a modified version of these data file for any purpose
 |   |
| CC-BY-NCAttribution-Non-commercial | * Same as CC-BY except users cannot distribute a modified version of these data for commercial purposes
 |  |
| CC-BY-NC-SAAttribution-Non-commercialShared Alike | * Same as CC-BY-NC except, when users share the data, they cannot change the license agreements
 |  |
| CC-BY-NC-NDAttribution-Non-commercial- No derivatives | * Same as CC-BY-NC except users cannot redistribute a modified version of these data file for any purpose
 |  |

|  |
| --- |
| **Other resources** |

For more information, visit [www.ands.org.au](http://www.ands.org.au). In particular, this website covers other topics such as

* How should you cite data that you use in your papers [www.ands.org.au/\_\_data/assets/pdf\_file/0005/724334/Data-citation.pdf](http://www.ands.org.au/__data/assets/pdf_file/0005/724334/Data-citation.pdf)
* How to publish sensitive data ethically [www.ands.org.au/\_\_data/assets/pdf\_file/0010/489187/Sensitive-Data-Guide-2018.pdf](http://www.ands.org.au/__data/assets/pdf_file/0010/489187/Sensitive-Data-Guide-2018.pdf)