Flinders UNIVERSITY

inspiring achievement
Finding and Using Research Data
You don’t have to collect it all yourself…

Liz WalkleyHall
Open Scholarship and Data Management Librarian
Finding and Using Research Data

• **Structure for today**
  • 2.00 pm - Introduction
  • 2.05 pm - Who are you? / What are you hoping to get out of today?
  • 2.20 pm - I will talk at you
  • 2.50 pm - Spend time looking at other data sources
  • 3.10 pm - Questions
  • 3.20 pm - Wrap up/close
What is research data?

Do you write "the data is..." or "the data are..."?
Is it singular or plural?

That is a deep philosophical question, actually.

It depends on whether you consider data to be facts (plural) or information (which is singular).

It's a fascinating grammatical conundrum.

"is"? "are"

information ➔ data ➔ facts

What if I only have one data point?

Then you have bigger problems than grammar.

"Piled Higher and Deeper" by Jorge Cham
www.phdcomics.com
Reused with permission.
Research data generally

• What does research data mean to you?
  – Spreadsheets?
  – Photographs?
  – Files from an instrument?
  – A pile of historical papers?
  – Audio files?
  – Results from a survey?
  – Can be just about anything…
Why would you use someone else’s?

• It’s convention in your discipline
  – Eg economics
• You have come across a paper and feel the data used could be useful for your question
• You might want to add another dimension to your research
  – Eg adding historical rainfall measures to your data about animal behaviour
• You don’t want to repeat a collection of data that’s already there
Can you use someone else’s?

• Quite often – yes!
• Data can be made available in a number of ways
  – Openly under license
  – For sharing, subject to permission/collaboration
  – For use under formal agreement
Licensing – Open Data

Before using data – always check the license

Open data often uses Creative Commons licenses.

Content in the Dryad Digital Repository is offered "as is." By downloading files, you agree to the Dryad Terms of Service. To the extent possible under law, the authors have waived all copyright and related or neighboring rights to this data.
You may need to ask before getting access to data. Access may not be granted! The data owner may wish to work with you to ensure the data is interpreted correctly.
Before using data – you may need to sign an agreement and there could be an access charge.

The University has agreements with a number of data providers. More information later…
How can you find data?

Easy way? Just follow links from publications (when they are provided)…

Original Article

Focused Analysis of Exome Sequencing Data for Rare Germline Mutations in Familial and Sporadic Lung Cancer

Yanhong Liu, PhD\textsuperscript{a}, Farrah Kheradmand, MD\textsuperscript{a}, Caleb F. Davis, PhD\textsuperscript{a}, Michael E. Scheurer, PhD\textsuperscript{a}, David Wheeler, PhD\textsuperscript{a}, Spiridon Tsavachidis, MS\textsuperscript{a}, Georgina Armstrong, MPH\textsuperscript{a}, Claire Simpson, PhD\textsuperscript{b}, Diptasri Mandal, PhD\textsuperscript{c}, Elena Kupert, MS\textsuperscript{d}, Marshall Anderson, PhD\textsuperscript{e}, Ming You, MD, PhD\textsuperscript{f}, Donghai Xiong, PhD\textsuperscript{g}, Claudio Pikielny, PhD\textsuperscript{h}, Ann G. Schwartz, PhD\textsuperscript{a}, Joan Bailey-Wilson, PhD\textsuperscript{i}, Colette Gaba, MPH\textsuperscript{j}, Mariza De Andrade, PhD\textsuperscript{k}, Ping Yang, MD, PhD\textsuperscript{l}, Susan M. Pinney, PhD\textsuperscript{m}, The Genetic Epidemiology of Lung Cancer Consortium Christopher I. Amos, PhD\textsuperscript{n}, Margaret R. Spitz, MD\textsuperscript{o}.

\textsuperscript{a} Baylor College of Medicine, Houston, TX, USA
\textsuperscript{b} National Institutes of Health, Baltimore, MD, USA
\textsuperscript{c} Louisiana State University Health Sciences Center, New Orleans, LA, USA
\textsuperscript{d} University of Cincinnati College of Medicine, Cincinnati, OH, USA
\textsuperscript{e} Medical College of Wisconsin, Milwaukee, WI, USA
How can you find data? – Data Journals

FindIt@Flinders

Show bx Hot Articles

Results 1 - 30 of 61 for Search All
Sort by: Relevance

Show only
Available in the Library (1) | Full Text Online (45)

Journal of physical and chemical reference data (Online)
American Institute of Physics; United States. National Bureau of Standards.; American Chemical Society. 2000-
Online access

Biodiversity Data Journal (Online)
2013-
Online access

European Journal of political research. Political data yearbook.
What’s out there?

- All links at:

  http://flinders.libguides.com/openscholarship/data
What’s out there? — Scraping the surface

- **Discipline/Generalist Repositories**
  - Australian Data Archive
  - Dryad
  - Genbank
  - FigShare
  - ICPSR (Interuniversity Consortium for Political and Social Research *US*)
  - TERN (Australian Terrestrial Ecosystem Research Network) Data Discovery Portal
  - World Health Organisation Global Health Observatory
What’s out there? — Scaping the surface

• Open Australian Government data
  – ABS
  – Geoscience Australia
  – BOM
  – Data.gov.au
  – Data SA
  – GovPond

• Overseas sources
  – See dataportals.org
What’s out there? — Scraping the surface

- **Data catalogues**
  - Research Data Australia
  - CSIRO Data Access Portal

- **Catalogues of repositories**
  - re3data.org – Registry of Research Data Repositories
  - Open Access Directory – data repositories
What’s out there? — Scraping the surface

These datasets are available under agreements entered into by the University.

For info and access,
Contact the University’s Statistical Consultant,
Pawel.Skuza@flinders.edu.au

- Australian Confidentialised Unit Record Files (CURFs) from ABS
- New Zealand Confidentialised Unit Record Files
- Luxembourg Income Study
- Household, Income and Labour Dynamics in Australia (HILDA)
- Longitudinal Study of Australian Children
- Longitudinal Study of Indigenous Children
- Longitudinal Study of Humanitarian Migrants
How to use it?

It depends on the data.

Ask yourself:
• Is the source of the data reliable?
• Can you use the data within the licensing/agreement conditions?
• Is there enough information ABOUT the data available for you to be sure that it’s relevant to your work? (Context, related publications?)
• Do you understand the data well enough to interpret it or reuse it? (Do you know what the variables are? Is there a data dictionary?)
How to use it?

• Can you use the data the way you wish to? (Can it be combined with other data sources?)
• Is the data “clean” for your purposes – or how much time will you need to spend to get it ready for your use?

Get more specific advice from your supervisor.

If you do need to combine/clean spreadsheet-based data, may be worth investigating OpenRefine open source software.
Once you use it – cite it.

Data is a research output just like any other. Most data will provide information on how to cite. Some will provide export functionality to reference managers.
Questions?

Amanda Nixon,  
*Manager, eResearch@Flinders*

[mailto:amanda.nixon@flinders.edu.au](mailto:amanda.nixon@flinders.edu.au)

X 15217

Liz Walkley Hall  
*Open Scholarship and Data Management Librarian*

[mailto:Liz.walkleyhall@flinders.edu.au](mailto:Liz.walkleyhall@flinders.edu.au)

x13080