



Research Data in the Humanities, Arts and Social Sciences: resources, protocols and opportunities

Alexis Tindall, December 2018



Key messages:

- Good data management leads to many benefits for the individual researcher, including
 - improving the quality of their research
 - increasing the likelihood of citation, collaboration and re-use
- Good data management is necessary to adhere to policies of institutions and granting bodies, and is a part of ethical professional practice

Learning objectives:

- Improved familiarity with issues in data use and management, especially within HASS disciplines
- Understand and be able to discuss the roles and responsibilities of a HASS researcher in the data lifecycle
- Understand frameworks and resources that encourage better reuse, reproduction, and publishing of research data
- Increased familiarity with formal protocols and ethical considerations, including in relation to sensitive data, personal data and Indigenous data

Ingredients:

Pre-reading:

- Miriam Posner's "Humanities Data: A Necessary Contradiction". Blog post. Read [here](#)
- Christine L. Borgman. 'Research Data: Who will share what, with whom, when and why?' Fifth China – *North America library Conference* 2010, 8-12 September 2010, Beijing. Read [here](#)

Other resources:

- [Australian Code for the Responsible Conduct of Research, 2018](#)
- [Guidelines for Ethical Research in Australian Indigenous Studies](#)
- ARDC: ANDS Guides
 - <https://www.ands.org.au/working-with-data>
 - <https://www.ands.org.au/working-with-data/fairdata>
 - <https://www.ands.org.au/working-with-data/sensitive-data>
- [Data Management Plan Checklist, Qualitative Data Repository](#)
- ["The Quartz guide to bad data"](#).

Pre-workshop requirements:

(e.g. programs to be downloaded or registrations established)

- None

Method:

Series of steps to follow

What is research data?

*Research Data: Data are **facts, observations or experiences on which an argument, theory or test is based**. Data may be numerical, descriptive or visual. Data may be raw or analysed, experimental or observational. Data includes: laboratory notebooks; field notebooks; primary research data (including research data in hardcopy or in computer readable form); questionnaires; audiotapes; videotapes; models; photographs; films; test responses. Research collections may include slides; artefacts; specimens; samples. Provenance information about the data might also be included: the how, when, where it was collected and with what (for example, instrument). The software code used to generate, annotate or analyse the data may also be included.*

The University of Melbourne makes no functional distinction between physical research products, digital research data and records of research, which can include items such as correspondence, application documents, reports and consent forms.

University of Melbourne Policy on the Management of
Research Data and Records (emphasis added)

What is research data in HASS? Why is it different from some other disciplines?

Research data in the humanities, arts and social sciences can be almost anything. 'Observed material' including archives, records, social media content, works of art, sensor data, audio-visual materials, works of fiction, news media and many other materials form the basis for HASS research, as does data generated through research, including interviews, recordings, code and other things.

Our understandings of data are commonly drawn from sciences and other disciplines, but research data are "any recorded information that are used to answer a research question"¹.

Some challenges and differences for this community?

- Accessibility of data, legacy formats
Many humanities scholars draw on data that may not originally exist in digital form. This can include archives, news media, objects and works of art. Even old forms of data, such as historic census records and previous humanities projects can be digitally inaccessible. Limiting one's research to data that can be discovered and accessed digitally can skew the outcomes of research.

¹ Heather Coates, IUPUI University Library [Data + Humanities](#) (accessed October 2018)

- Diversity of assets, non-traditional forms, 'datafication of un-data'
Data driven research uses computational analysis to expand and accelerate research outcomes. Attempting data driven research in fields that traditionally draw on analogue resources adds a degree of research complexity and can require a researcher to justify and defend their research. New forms of analysis are rarely immediately or universally accepted by traditional research communities.
- Sensitivity
Many HASS researchers study human activity in various forms. Motivation to share and expose data for re-use may be thwarted by ethical considerations around private information. Discovery of useful data to augment or enhance research may face similar challenges.
- Assertions of objectivity, researcher's role in the generation, collection and interpretation of data
This can be a sticking point when we attempt to define data driven HASS research. Humanities scholars have a deep and well-considered understanding of the near impossibility of research objectivity, which can lead to skepticism towards HASS data.

That last point can be a significant inhibitor in HASS data-driven research. It is well explained here:

When you call something data, you imply that it exists in discrete, fungible units; that it is computationally tractable; that its meaningful qualities can be enumerated in a finite list; that someone else performing the same operations on the same data will come up with the same results. This is not how humanists think of the material they work with ...

Humanists tend to believe that the scholar's own subject position is inextricably linked to the scholarship she produces.

Miriam Posner, [Humanities data: a necessary contradiction](#)

In response to this, humanities and arts scholars often find themselves considering their research practice, as well as their research. This can be a healthy scenario, and provides even greater motivation to be familiar with, and adhere to, community accepted standards and protocols related to data management.

Benefits of good data management for HASS researchers (and everyone else!)²

- To make research easier
- To stop yourself drowning in irrelevant stuff
- In case you need the data later
- To comply with law, regulations or funder expectations
- To respond to critique
- To share data for the benefit of your community

² List inspired by similar shared by Peter Neish, University of Melbourne, presented as part of Researcher@Library Week 2018.

- Recognition of your work
- To enable reproducibility

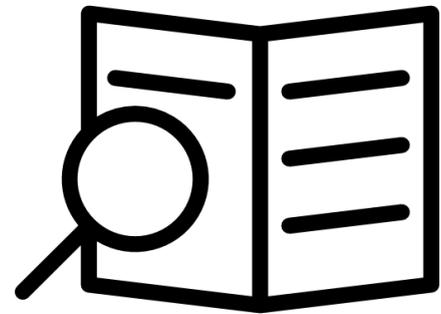
Research data management training often accentuates the community benefit, or simple ‘worthiness’ of the researcher with good data management skills. As data driven research becomes more common, researcher skills expand, and greater quantities of data can be created, shared and used, the primary benefits of good data management can be for the researcher themselves. Accentuating the first three points (highlighted in red) can help researchers understand that individual benefit.

What is good data management?

With researchers working in diverse institutional and professional or disciplinary environments, it is impossible to develop generic data management advice that would suit all researchers.

Good data management should address the following elements, which are a summary of several resources, including the Qualitative Data Repository’s [Data Management Planning Checklist](#).

- Make use of your librarians and/or research office!
- Practicalities: who, how, what formats, how much?
- How will you organise your data?
- What processes?
- Security/integrity?
- Sensitivity?
- Documentation of your data
- Storage and preservation
- Sharing and accessibility

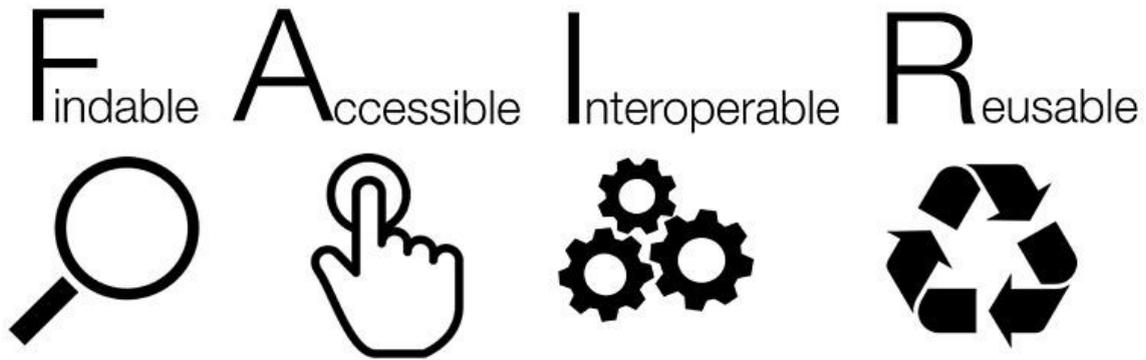


Protocols, guidelines and standards: necessary *and* helpful

Data driven HASS research demands adherence to a range of guidelines, protocols and standards that are often perceived as a burden or an obligation, when re-framing these protocols as resources can provide valuable guidance to a researcher considering their data management choices. Several key protocols and guidelines relevant to HASS research are explored below.

[FAIR Data](#)

FAIR Data are a set of principles that were designed by diverse stakeholders at a workshop in 2015, with the objective of enabling data re-use. They have since been recognised by major decision-makers and funders as a model of thinking about data that will maximise use, which means better value from research investments and increased opportunities for collaboration.



Key points about the FAIR Data Principles:

- focus on making research data re-useable extends to machines as well as humans, making data more discoverable and harvestable for big data analysis
- they are technologically agnostic, and deliberately general, written without any particular discipline in mind
- they address both data and metadata, and implementation can require review of policies, infrastructure, governance, tools and software being used.

Given the concerns about sensitivity in HASS research data, it is important to assert that FAIR does not require data to be open.

FAIR Data are:

Findable:

- Assigned a unique, resolvable, persistent identifier (doi, handle etc)
- Rich metadata and machine-readable descriptions support searching
- Registered or indexed in a searchable resource (this can affect your choice of repository)

Accessible:

- Clearly defined access and security protocols
- Metadata is always accessible (even if data is not)
- May include making the data open (but not always)

Interoperable:

- Machine interpretable formats for data and metadata
- Using accepted vocabularies, ontologies, metadata schema
- Linked to other resources

Reuseable:

- Licence and provenance provided
- Community standards used

It is helpful to note that the application of FAIR principles exists on a continuum, it is not an absolute measure. Not all datasets will be able to be completely FAIR, but consideration of the principles and

making efforts to be more FAIR can improve your research data practice. The ARDC [FAIR Self-Assessment Tool](#) can help you learn more about FAIR data in relation to your data and research.

[Australian Code for the Responsible Conduct of Research, 2018](#)

This Code, developed between the Australian Research Council, Universities Australia and the National Health and Medical Research Council, articulates the broad principles that characterise an honest, ethical and conscientious research culture. The Code outlines responsibilities for institutions, as well as researchers.

In addition to demanding that researchers support a responsible culture of research in their institution and profession, the Code addresses data management under Recommendation 22:

Responsibilities of researchers

- R22 Retain clear, accurate, secure and complete records of all research including research data and primary materials. Where possible and appropriate, allow access and reference to these by interested parties.

Adherence to the Code is a condition of ARC/NHMRC funding, and breaches of the Code are considered research misconduct.

[Guidelines for Ethical Research in Australian Indigenous Studies](#)

These Guidelines, created by the Australian Institute of Aboriginal and Torres Strait Islander Studies, touch on intellectual property, traditional knowledge, cultural practices and address emerging issues of digitisation, data and information management practices.

The Guidelines advocate the highest standards of ethics and human rights, and ensure that research with and about Aboriginal and Torres Strait Islander peoples follows a process of meaningful engagement and reciprocity between the researcher and individuals or communities engaged in the research.

Recording, management, access and ongoing use of data in research are considered throughout these Guidelines, and they are formatted in a way that provides advice and guidance.

Data management issues can be explored through several of the fourteen principles in these guidelines. For example, principles around consultation, agreement and mutual understanding refer to consultation around 'intended research outputs, including data', while the group of principles that consider ongoing management of research results calls for ongoing access for Indigenous people to the data arising from research, issues of confidentiality and encourage the deposit of research results in libraries and archives.

Sample activities:

Activity 1: Tell your neighbour about your data

Objective of this learning activity: Start thinking about how to talk about and evaluate data. Quickly identify opportunities, challenges and risks. Encourage interaction among group.

Four minute activity.

Turn to the person next to you, and describe a dataset that you use or are familiar with in two minutes. After two minutes, swap places.

Instructor calls on several participants to tell the group about their neighbour's dataset.

This can teach participants to quickly summarise their dataset, including any strengths and challenges, and learn that others may perceive their data differently from themselves.

It could be re-run at the close of the workshop in different pairs, and partners could be asked to report back on their neighbour's dataset in the context of what they have learned about community principles and protocols.

Activity 2: How FAIR are you?

Objective of this learning activity: Introduce participants to the ARDC FAIR Self-Assessment Tool. Embed an understanding that FAIR assessment is on a continuum, isn't an absolute.

Ten minute activity.

Open the ARDC [Fair Self-Assessment Tool](#).

Use it to evaluate your own dataset. Alternatively the instructor can provide links to relevant datasets for quick evaluation, these can be drawn from Research Data Australia or institutional or disciplinary repositories.

Next steps:

Further research data management resources available through Tinker include:

Data Curation Framework

A model for the documentation of decisions and negotiations between researcher and data has been developed through consideration of three reference datasets, and consultation with researchers and data custodians. Read more [here](#).

10 HASS Data Things

An adaptation of the ARDC 23 Research Data Things, this document gives researchers the opportunity to take the next step in learning about research data, with discipline relevant examples and self-guided learning activities. Dip into the 10 HASS Data Things [here](#).

With the exception of logos or where otherwise indicated, this work is licensed under the Creative Commons 4.0 International Attribution Licence.

Recommended attribution: Alexis Tindall, [Tinker](#)

