What are health-y data?
Why might they be trickier to publish?

Sarah Olesen, ANDS
Publishing & Sharing Health-y Data Workshop, 23 Apr, 2015
Data = ‘records, files or other evidence, irrespective of their content or form (e.g. in print, digital, physical or other forms), that comprise research observations, findings or outcomes, including primary materials and analysed data’ (Monash University 2010)

Health data = records etc. that includes information about people’s physical, mental, cognitive status, health or wellbeing

Examples: epidemiological and social surveys, experimental, clinical trials, administrative health data, genetic and other samples, ....?
What makes it trickier?
What makes it trickier?

• Data comes from or is about people, it is often paired with personal information, or inextricably linked to person/s (e.g., in-depth interviews, genetic data...)

• This is **personal data** - ‘information or an opinion .., whether true or not, and whether recorded in a material form or not, about an individual whose identity is apparent, or can reasonably be ascertained ..’ (**Privacy Act 1988, s6**)  

• Often **sensitive data** – When personal information (i.e., identifiable) is accompanied by information about...

  ‘..must not use or disclose the information for another purpose (the secondary purpose) unless:
  (a) the individual has consented to the use or disclosure of the information..
  (b) ..’
What makes it trickier?

**Sensitive data** – ‘any data that contain information that can be used to identify an individual and introduce a risk of discrimination, harm, or unwanted attention.’

What makes it trickier?

- Human data is subject to greater scrutiny and security
- Additional steps to publication not relevant for many other kinds of data
- Often in unfamiliar territory
- Limited guidance to date
ANDS Guide

Publishing and Sharing Sensitive Data Decision Tree

1. Define
   - Collecting new data
     - See: Data planning
   - Previously collected by you
   - Collected by others
     - See: Data Citation

2. Data collection

3. Ownership
   - Do you have the right to publish?
     - See: Intellectual Property and data sharing
     - Yes
     - No
       - may be able to publish metadata only

4. Confidentialise
   - Can the data be made non-sensitive?
     - Yes
       - See: Confidentialise your data, and Formatting data for publication
     - No
       - may be able to publish metadata only

5. Research ethics
   - Was consent for data sharing given by (any human) research participants?
     - Yes
     - No
       - Some data may still be shared if confidentialised & sharing was not precluded during consent. Check: What's legal and Ethical considerations

6. Licensing
   - All published Australian data intended for reuse should have a licence. This explains to the re-user how the data should be used and attributed to the original data owner.
   - Go to: the Licensing section in the Guide, ANDS licensing page and AusGOAL

7. Making data discoverable
   - Write a data description, i.e., metadata, for publication in a data repository.
   - See: How to select a repository in the Guide

8. Publishing and sharing data
   - Publish metadata and data? Or metadata only with conditional access to the data itself?
     - At minimum, metadata can be made publicly available.
     - For confidentialised data, it is often appropriate to have public metadata and 'conditional access' (see in the Guide) to the data itself.
     - Also see: How to select a repository in the Guide
What’s your sticking point?
This work is licensed under a Creative Commons Attribution 4.0 International License

NCRIS
National Research Infrastructure for Australia
An Australian Government Initiative

ANDS is supported by the Australian Government through the National Collaborative Research Infrastructure Strategy (NCRIS).