

Two side-by-side images are positioned to the left of the title. The left image is a close-up of a textured surface, possibly a piece of wood or stone, showing a grid-like pattern of lines. The right image is a traditional Aboriginal rock painting, featuring a central figure with a sunburst-like head and a body composed of vertical lines, set against a reddish-brown background.

Ethics of Data Publication: Same same or different?

Brave new world of data publication & citation, interaction with other scholarly outputs, and quandaries for publishers and researchers.

Dr. Sarah Olesen

Who are ANDS?

- Funded by the Australian Government, National Collaborative Research Infrastructure Strategy (NCRIS)
- Consortia of Monash Uni, ANU, CSIRO
- Providing resources and collaborating with institutions to better manage, publish, discover, and reuse data
- Including Research Data Australia (researchdata.ands.org.au)
- ands.org.au

“Research data is presently a publicly-funded resource that passes into private hands without explicit .. remuneration to the public purse. The overwhelming volume of research across the disciplines is funded by government via research councils and institutions of higher education and by non-profit-making institutions set up for the public good. Organisations wish to maximise value in their investment and there is growing opinion from funders that access to data is part of that value.”

Sturges, P. et al. Research data sharing: developing a stakeholder-driven model for journal policies. *Journal of the Association for Information Science and Technology*

<http://eprints.nottingham.ac.uk/3185/>

'Data sharing by scientists: Practices and Perceptions'

(Tenopir et al., 2011, PLoS one: 6(6) <http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0021101>)

- 85% of researchers were interested in using others' data if easily accessible
- 74% believed that their data could be used in other ways (across other research fields)
- But...
- Only 36% report that their own data is easily accessible

Why the gap?

- Inadequate data management practices/policies and storage resources
- Time pressure
- Researchers do not feel well equipped or knowledgeable about how to publish their data
- Legal and ethical concerns

- Data management primer
- What is data publication?
- Why publish, share and cite data?
- Tricky questions and researcher concerns
 - Real & perceived barriers
 - Sensitive and personal data
- Over to you

What do we mean by data?

Quick primer: What counts as research data?

- | | |
|--|---|
| 1. Observations | 7. Specimens |
| 2. Recordings (e.g., sound, spatial, geol) | 8. Samples (e.g., medical) |
| 3. Measurements | 9. Software code |
| 4. Images | 10. |
| 5. Instrument/computer output | 53. 'what needs to be kept to validate the results of research' |
| 6. Objects (e.g., artefacts) | |

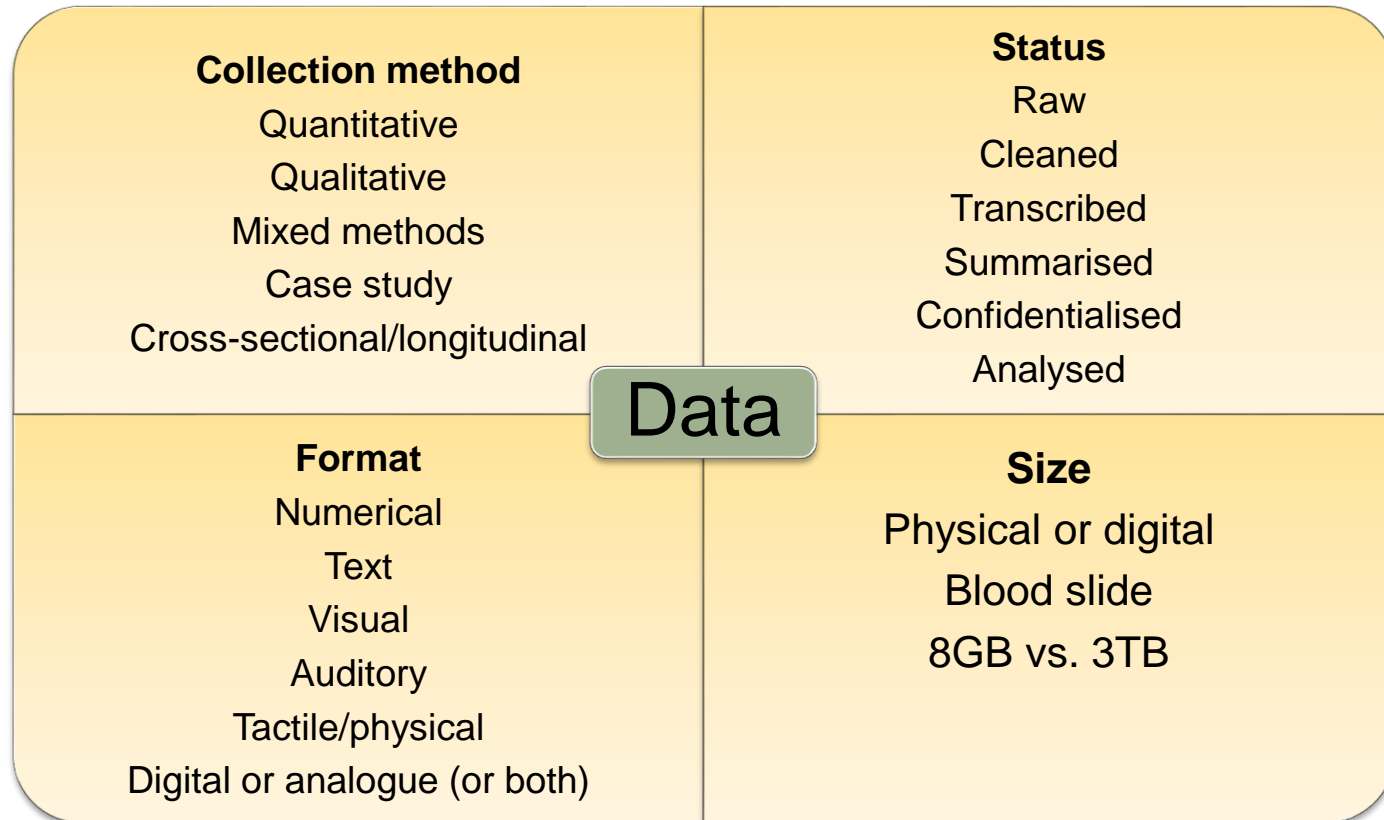
- Vary by discipline
- That contribute to knowledge (an argument, theory, test or hypothesis) and/or other research outputs
- The building blocks of research knowledge

http://www.dcc.ac.uk/sites/default/files/documents/publications/DCC_Howto_Discover_Requirements.pdf

http://royalsociety.org/uploadedFiles/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf

http://www.mopp.qut.edu.au/D/D_02_08.jsp; <http://www.orei.unimelb.edu.au/content/research-integrity-and-misconduct>

What do we mean by data?



- Elements
 - data organisation
 - documentation
 - backups
 - archiving for long-term preservation
 - data sharing or publishing
 - ensuring security of confidential data
 - data synchronisation
 - Written data management *plan*
- Henty et al. (2008): 82.3% of surveyed researchers at 3 of Australia's largest universities did not have a DMP

- Responsibility of researcher
- Becoming more (in)famous as major funders emphasise data management and reporting
 - E.g., NIH (US), Research Councils UK
 - E.g. Australian Research Council: From Feb 2014, application forms (Discovery & Linkage Grants) highlight importance of DMP and data sharing

"The ARC considers data management planning an important part of the responsible conduct of research and strongly encourages the depositing of data arising from a Project in an appropriate publicly accessible subject and/or institutional repository".

(<http://www.arc.gov.au/pdf/DP15/Funding%20Rules%20for%20the%20Discovery%20Program.pdf>)

“Outline plans for the management of data produced as a result of the proposed research, including but not limited to storage, access and re-use arrangements”

(ARC Discovery Projects, Instructions to Applicants - http://www.arc.gov.au/pdf/DP15/DP15_ITA.pdf)

See also: <http://ands.org.au/datamanagement/guide-to-filing-in-the-data-management-section-for-the-arc-march2014.pdf>

What does DM have to do with research ethics?

- Institutional Policies and Ethics Committees may require a DM plan
- Australian Code for the Responsible Conduct of Research states:
 - *“Policies are required that address the ownership of research materials and data, their storage, their retention beyond the end of the project, and appropriate access to them by the research community”.*
(<https://www.nhmrc.gov.au/guidelines/publications/r39>)
- Effective research data management is vital to (ethical) managing risk (e.g., data loss or corruption, inability to validate/replicate, privacy breaches)

**Because data publication and sharing is predicated
by good data management**

- Preparing and disseminating data to the scientific community
- Often alongside other scholarly outputs
- **Publication = making the data 'discoverable'**
- Sharing = distributing actual data to others

- At minimum, publishing data means making metadata and *information about how to access data* discoverable by the public
- Access to (i.e., sharing of) the data may then have no or various conditions/restrictions, depending on legal, ethical, funding or commercial restrictions

'Openness' of data publication

Public-Public (Open access)	Public-Private (Mediated open access)	Private-Private (Closed access)
Metadata is fully discoverable	Metadata is fully discoverable	Metadata is not publicly available
Data are accessible and immediately downloadable	Mediated access to data via data custodian	Data not discoverable or available to third parties
Preferred option for non-sensitive data from completed projects	Good option for sensitive or confidential data	Safest option for highly-sensitive data

More on mediated access = metadata are published, access to actual data is conditional and/or mediated by a custodian

Examples:

- Sensitive data (e.g., medical records, personal + political, religious, endangered/conserved species)
- License has more conditions than CC-BY (AusGOAL Creative Commons, Restricted)
- Researcher approval required
- Validation of party accessing the data or purpose of access
- Researcher or data owner may retain some control over access

Formal

(conventions around metadata and more)

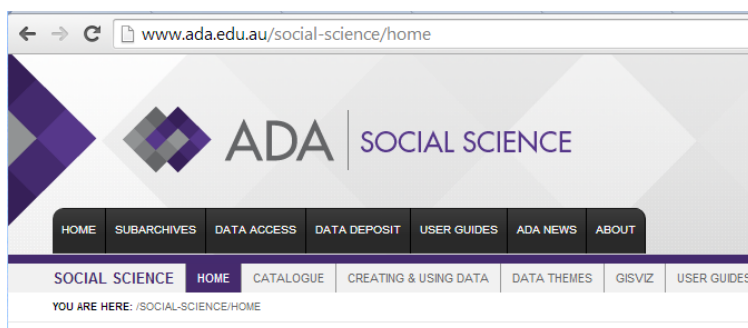
Informal

(few conventions)



Repositories

Peer networks



- Formality of storage does not correlate to 'openness' of data
- Examples of repositories by discipline: <http://databib.org/>

Let's regroup

- Data needs to be managed to be effectively published and shared
- Publishing does not necessarily = sharing/open access
- Sharing is the goal, but may be mediated
- You can publish metadata so the dataset can be discovered but still mediate access to actual data
- *Sharing often follows publication*

Incentives to publish and share data

- Technology advances \leftrightarrow Research is becoming more data intensive
- Also the ability to make data more readily and widely available
- And an expectation that it will be

- Benefits including individual, publisher, and institutional reputation (and thus, funding success...)

Incentives to publish and share

- 🥕 Provide academic credit
- 🥕 Data becomes *discoverable*
- 🥕 Collaborations
- 🥕 Reputation/Profile – future funding (e.g., where sharing lead to funding)
- 🥕 Citations (of data itself, and linked research papers)*
- 🥕 Tracking the reach, impact and reuse of all your research outputs
- 🥕 Ability to publish in leading journals
- 🥕 Ahead of the wave
- 🥕 Legacy
- 🥕 Security
- 🔪 Fulfilling requirements
- 🔪 Avoiding being left behind

* Piwowar et al. (2007) <http://www.plosone.org/article/info:doi%2F10.1371%2Fjournal.pone.0000308>

- Some publishers of scientific research now require or encourage authors make their data available
- PLoS (March 2014): *“make all data underlying the findings described in their manuscript fully available without restriction, with rare exception”* (<http://www.plosone.org/static/policies#sharing>)

JoRD (Journal Research Data Project; Joint Information Systems Committee [JISC])) (<http://eprints.nottingham.ac.uk/3185/>)

- Majority of journals had no/unknown formal policy
- Journals with high impact factors tended to have the strongest policies/mandates, as did certain disciplines
- Great variation in content
- From 'strong' to 'weak' to none (also Piwowar & Chapman, 2008 http://elpub.scix.net/data/works/att/001_elpub2008.content.pdf)

Possible exceptions to sharing as per policies

- Sensitive data (e.g., medical records) that are under legal and/or ethical restrictions
- Data from a third party
- These exceptions align with common ethical concerns about sharing

Great resource: <http://www.plos.org/plos-data-policy-faq/>

Why mandate data publication and sharing?

- Based on ethical research values
 - Validation / Replication
 - Reinterpretation
 - Reuse reduces cost, participant fatigue
 - Publically funded = public resource
 - Quality check

Also in this space and reflecting the same beliefs about value of data

- More traditional journal model
- To expose and share meta/data, promote reuse
- Data articles describe datasets without analyses (like and extended metadata)
- Receive a *journal* citation
- Peer reviewed
- Examples:
 - Geoscience Data Journal (Wiley and Royal Meteorological Society)
 - Scientific Data (Nature)
 - Journal of Open Archaeology Data (Ubiquity)

- Not always directed by publisher
- If guidance is given, its similar to funder recommendations
- In a formal, public data repository (institutional, discipline-specific, general)
 - See datatib.org for list
 - with metadata
 - May still be licensed / mediated in some way, e.g., “available on request” if necessary (see <http://www.plos.org/plos-data-policy-faq/>)
- Or sometimes as additional material to article (smaller datasets)

One of the big hooks for data publication is the more immediate returns offered by data citation

- Piwowar et al. (2007)* publications received more citations if study data was also available
- If data (or metadata at least) is in a repository it has a public record and **can be cited**
- Just like citing a journal article
- ‘Gold standard’ includes a persistent identifier, linked to other scholarly outputs

*<http://www.plosone.org/article/info:doi%2F10.1371%2Fjournal.pone.0000308>



[Back to Search](#)

[My Tools](#)

[Search History](#)

[Marked List](#)



[Save to EndNote online](#)

[Add to Marked List](#)

[Back to List](#) ◀ 1 of 1 ▶

Household, Income and Labour Dynamics in Australia (HILDA) Survey

By: [Wooden, Mark](#)

Household, Income and Labour Dynamics in Australia (HILDA) Survey

Source URL: <http://melbourneinstitute.com/hilda/>

Viewed Date: 11 Mar 2014

Published: 2000

Abstract

The Household, Income and Labour Dynamics in Australia (HILDA) survey is an ongoing nationally-representative panel survey that commenced in 2001 with a sample of around 8000 Australian households. Members of these households, as well as any individuals who subsequently join, have been followed over time on an annual basis. In 2011 a new cohort of just over 2100 responding households was added. The survey is funded by the Australian Government through the Department of Social Services, but responsibility for the design and administration of the survey, and for the production and dissemination of data, rests with the Melbourne Institute of Applied Economic and Social Research (at the University of Melbourne). The survey has been designed with the broad objective of providing researchers and policy-makers a tool for assessing a wide range of economic, demographic, and social policy issues, including: (i) the incidence and dynamics of poverty; (ii) asset allocation, wealth accumulation, and income changes throughout the life course; (iii) the correlates and impact of changes in physical and mental health; (iv) household formation and dissolution; (v) social capital; and (vi) subjective well-being. The unit-record data from the HILDA Survey are made available for use to researchers in academia, government and other research organisations (both in Australia and overseas). The survey has been designed with the broad objective of providing researchers and policy-makers a tool

Citation Network

296 Times Cited

0 Cited References

[Create Citation Alert](#)

(data from Web of Science™ Core Collection)

All Times Cited Counts

493 in All Databases

296 in Web of Science Core Collection

24 in BIOSIS Citation Index

0 in Chinese Science Citation Database

0 in Data Citation Index

0 in SciELO Citation Index

Most Recent Citation

Ambrey, Christopher L. [The causal effect of income on life satisfaction and the implications for valuing non-market goods](#). *ECONOMICS LETTERS*, MAY 2014.

[View All](#)

This is an example of a record for the HILDA dataset from
Thompson Reuters Data Citation Index (Web of Science)
- 296 citations

How are data cited?

- Like a journal article
- Best practice includes Persistent Identifier
 - Traceable ('resolvable')
- Digital Object Identifier (DOI)
 - 'digital identifier of an object' (scholarly)
 - International standard
 - Can be indexed (e.g., Thomson Reuters Data Citation Index)
 - Looks like:
 - [10.4225/13/50BBFCFE08A12](https://doi.org/10.4225/13/50BBFCFE08A12)
 - <http://dx.doi.org/10.4225/13/50BBFCFE08A12>
 - <http://doi.org/10.4225/13/50BBFCFE08A12>

How do you get a DOI?

- Some repositories generate ('mint') DOIs when data is deposited (e.g., Figshare)
- ANDS mints and manages DOIs (for institutional repositories) on behalf of DataCite
- DataCite is the major administrator for data DOIs, of which ANDS is an agent.

Example of data citation with DOI

From *Molecular Ecology* (2013) 22, 3098–3111

origin of *Betula tortuosa* Ledeb. *Reports from the Kevo Subarctic Research Station*, 10, 70–84.

Wang N, Thomson M, Bodles WJA *et al.* (2012) Data from: Genome sequence of dwarf birch (*Betula nana*) and cross-species RAD markers. *Dryad Digital Repository* <http://dx.doi.org/10.5061/dryad.v5gd2>.

Wilsey BJ, Saloniemi I (1999) Leaf fluctuating asymmetry in tree-line mountain birches. *Betula pubescens* ssp. *tortuosa*

From *PLOS ONE* 10.1371/journal.pone.0074510

37. Klemenhagen KC, O'Brien SP, Brody DL (2013) Data from: Repetitive concussive traumatic brain injury interacts with post-injury foot shock stress to worsen social and depression-like behavior in mice. *Dryad Digital Repository*. doi:10.5061/dryadv1t54.

38. Symington SH, Paul LK, Symington MF, Ono M, Brown WS (2010)

- Note the citation is treated completely on equal footing with citations to other sources
- The first reference is a dataset that's incorporated into the bibliography, with complete metadata (title, authorship, data center, etc.), and a persistent DOI link.

1. **Sensitive data** is data that can be used to identify an individual or object and place them at risk of discrimination or unwanted attention.
 - Familiar examples include: patient health records, personal information alongside religious, cultural or political affiliations, and location of endangered species or conservation efforts.
 - Sensitive by context

- Legal (Privacy Act) and Ethical (Human Research Ethics) may be involved
- Clear institutional guidance is often unavailable or not sought

Can it be done?

- Yes! if appropriate actions are taken
- Plan ahead. Best option is to request participants' consent to share their 'de-sensitised' data before it is collected
- What about **existing data**?
 - Now its about how data are treated/modified
 - And/or mediated before shared (remember 'public/private' distinction, and licensing)
- Once de-identified, the Privacy Act generally does not apply in most cases.
- Placing reasonable controls around access and reuse (i.e., via type of licensing) can further protect against re-identification

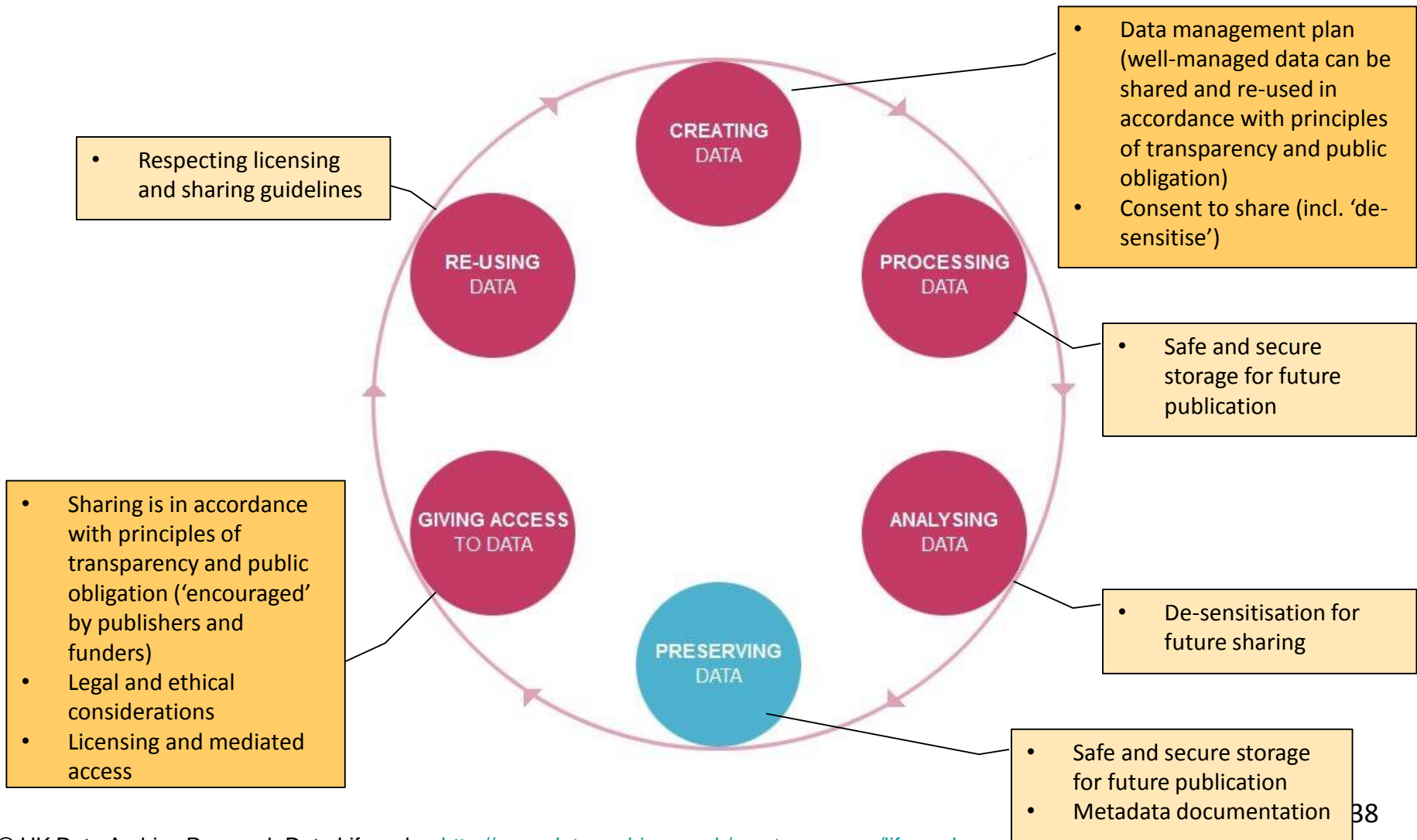
More resources on sensitive data

- www.nss.gov.au/nss/home.NSF/pages/NSS+Resources – data sharing
- <http://www.ausgoal.gov.au/> - licensing
- <http://www.oaic.gov.au/privacy/privacy-engaging-with-you/previous-privacy-consultations/de-identification-resources-may-2013/> - de-identification / Privacy Act
- <http://www.data-archive.ac.uk/create-manage>
- Big Data and Innovation, Setting the Record Straight: De-identification *Does Work* (Cavoukian & Castro, June 16, 2014, Information and Privacy Commissioner Ontario, Canada)
- <http://www.privacybydesign.ca/index.php/paper/big-data-innovation-setting-record-straight-de-identification-work/>

Stay tuned for **ANDS Guide** to be released later in 2014

- Decision tree
- Consent form examples
- Legal and Ethical expert consultation
- Clear direction

Ethics and data life cycle



- Publishing (to make data discoverable) and sharing where possible are in accordance with principles of research
- So, instead of thinking ‘is it ethical to share?’, some have begun to start from position that it may be unethical not to share
- Due to: Obligations to public funding and research subjects
- Repeating costs time, money, effort or research and participant communities
- Research must be validated/peer review

Take home ideas

- Carrots far outweigh sticks
- Advantages far outweigh disadvantages
- Plan ahead if possible (DM, ethical consent)
- ***Mediated access***
- Licensing
- Publish with Persistent Identifier
- It is all do-able!
- Reap your rewards!