10 medical and health research data Things

Use **Repurpose** **Adapt** **Change**

10 medical and health research data Things is an opportunity to explore issues surrounding management of research data, specifically for people working with medical, clinical and health data.

This program was developed from the 23 (research data) Things program and the extensive ANDS resources and materials related to research data management and re-use.

Australian National Data Service

Contents

Summary

Effective research data management of medical, health and clinical data is increasingly recognised as a critical part of the research process. It enables:

- Trust in data you obtain for reuse from other sources
- Reproducibility of research through increasing veracity of data
- Increased quality of your research
- Strengthening of researchers' reputation through increased citations and reach of all research outputs
- Increased connectivity between all research outputs, and researchers
- More efficient use of scarce research funds
- Data description for sharing and collaboration
- Reduced risk of loss or corruption of data

How can I work through these 10 Things?

- All 10 Things have 2 or 3 Activities. You can pick'n'mix from the Activities to suit your interests.
- You can do as much or as little of the Things and Activities as you want to do, or need to know.
- Some of the Activities are intended as an introduction to a topic, and some delve a little deeper. Choose what interests you and suits your experience.
- You can work through Activities on your own at your own pace, or in a group.

What can I do after I work through the Activities?

- Check out the ANDS hub for medical and health data
- Consider exploring more research data issues through 23 (research data) Things

Ideas to reuse and repurpose these activities

This material is licenced with a CC-BY licence, meaning that you can use, repurpose, adapt, or change it to suit your needs.

Short workshops

- Pick'n'mix Things and Activities to suit your group and their needs
- Only include those Things and Activities which are of interest to your group
- Change the 'Consider' question/prompt at the end of each Activity to 'Discuss'
- Edit activities to discipline specific examples and links

Please note: this is a snapshot in time: research data as it was in 2016 - you may need to check resources and update resources and links to include more recent initiatives and policy changes.

Thing 1: Getting started with research data

Research data comes in many shapes and sizes and its management changes over time. Kick off your research data journey by exploring different types and forms of research data and how they fit into the research lifecycle.

Activity 1

What “research data” are we talking about?

1. Read the Defining Research Data section from University of Oregon library - note that for data to be reusable the data collection often needs to include algorithms, scripts, software. It's not 'just data'.
2. As we have just seen, research data can come in many forms. Some of these are human readable, and some are machine readable. Explore a couple of these types of formats commonly used for medical, clinical and health data:
    a. images
    b. clinical trial data
    c. questionnaires
    d. interviews
    e. genetic sequences

Consider: make a list of the forms of data you have used or seen in your work. What would people need to know about these data if they wanted to re-use these data?

Activity 2

Data in the research lifecycle.

Data often have a longer lifespan than the research project that creates them. Follow-up projects may analyse or add to the data, and data may be reused by other researchers.

A data lifecycle shows the different phases a dataset goes through as the research project moves from "having a brilliant idea" to "making groundbreaking discoveries" to "telling the world about it"

1. Take a look at either:
    a. UK Data Archive Research Data Lifecycle (if you are new to this concept)
    b. DCC Curation Lifecycle Model (if you are familiar with this concept)
2. Have a look at the NHMRC Statement on Data Sharing (2 pages) and note the lifecycle diagram for data sharing

Consider: have you been through all of the steps outlined in this lifecycle? If not, which ones are new to you?

Activity 3

How data differs across disciplines

1. Choose one of the three specialised data repositories below, or find another data repository of interest - particularly one in a discipline you are unfamiliar with, and spend some time browsing around your chosen repository to get a feel for the data available.

   1. RCSB Protein Data Bank
   2. Australian Data Archive (this archive contains Social Science, Historical, Indigenous, Longitudinal, Qualitative, Crime & Justice and International data)
   3. USGS Water Data

2. Think about how the data here differs from data you are familiar with. Consider for example, format, size and access method.

Consider how cross disciplinary research could be affected by discipline data conventions, and also one way cross disciplinary data access can be facilitated.

Thing 2: Issues in research data management

Research data is critical to solving the big questions of our time. So what are some of the issues we face in managing research data?

Activity 1

Considerations in data management

Research data is for everyone. Governments and Universities all around Australia and the world are now encouraging researchers to better manage their data so others can use it.

Research data might be critical to solving the big questions of our time, but so much data are being lost or poorly managed.

1. Take just a minute and browse over some ways Queensland Government Data is being used by businesses, families, travellers, farmers.
2. This 4.40mins cartoon put together by the New York University Health Sciences Library, is about what happens when a researcher hasn't managed their data (at all…). What could possibly go wrong?!?
3. As you watch the cartoon jot down the data management mistakes which interest or appal you.
4. Now, scan through the dot points in the *Consider the following….* section of the University of the Sunshine Coast's LibGuide which provides advice for researchers on how to manage their data.

Consider how just ONE of the data disasters depicted in the cartoon could have been avoided.

Activity 2

How do you manage “Big Data”?

"Big Data" is a term we're hearing with increasing frequency. Data management for Big Data brings much complexity - citing dynamic data, software, high volume computing, storage costs, transfer of petabytes of data, preservation, provenance, and more.

1. Genomics is an area where dramatically increasing amounts of data are being created each year. Watch this video *Genomics and the human health sector* (2:34mins) about how genomics data can lead to accurate, timely and effective solutions in healthcare.
2. Read this short article about Australia's leading role in integrating genomics into healthcare.

"Genomics is a "four-headed beast"; considering the computational demands across the lifecycle of a dataset—acquisition, storage, distribution, and analysis—genomics is either on par with or the most demanding of the Big Data domains." from Stephens ZD, Lee SY, Faghri F, Campbell RH, Zhai C, Efron MJ, et al. (2015) Big Data: Astronomical or Genomical? PLoS Biol 13(7): e1002195. doi:10.1371/journal.pbio.1002195

3. Read this post and presentation titled: *"Big Data: The 5Vs Everyone Must Know".* This article uses 5V's: volume, variety, velocity, veracity and value as a concept for how big data can be managed more successfully.

Consider whether the concept of 5Vs is useful to support better management and reuse of genomics "Big Data".

Activity 3

From analog to digital with eLab Notebooks

Laboratory Notebooks are used by researchers to formally record their lab based research activities. As research has become increasingly digital and collaborative the utility of traditional hard copy Lab Notebooks has been challenged. Not surprisingly then, eLab Notebooks (ELN) have emerged as an alternative.

Effective data management for constantly updated data, such as that within ELNs, is a real challenge for projects who wish to publish their data during the project.

1. Read this short definition of ELNs.
2. Then read this article: *International team of scientists open sources search for malaria cure* about how an international team of scientists and citizen scientists are using open source ELNs to speed up a cure for malaria. You can access their open ELNs. Click on on Matthew Todd's ELN to see what it's in it.

Consider a data management issue and possible solutions, where data is generated, stored and shared via an open ELN.

Thing 3: Data sharing and discovery

Data may be shared in many ways. Here we explore places, and ways, that data can be shared and is currently being shared.

Activity 1

Exploring repositories

Repositories enable discovery of data by publishing data descriptions ("metadata") about the data they hold - like a library catalogue describes individual materials held in a library. Most repositories provide access to the data itself, but not always. Data portals or aggregators draw together research data records from a number of repositories, e.g. Research Data Australia (RDA) aggregates records from over 100 Australian research repositories.

1. Click on this description of a dataset from the National Survey of Midlife Development in the United States (MIDUS): a collaborative, interdisciplinary investigation of patterns, predictors, and consequences of midlife development in the areas of physical health, psychological well-being, and social responsibility.
2. Have a close look at the record to see the ways this record is discoverable and accessible. Explore how a secondary user can access the data and what formats they can download it in, and see how it is connected to hundreds of research publications.

Discipline specific repositories

1. Start by going to re3data.org
2. Click on Browse > Browse by subject > click on Medicine in the second ring from the middle
3. Explore the range of repositories listed under 'Medicine'. Can you find one relevant to your research?

Activity 2

An introduction to 'open', 'shared', and 'closed' data

You may have noticed that not all data described is available for immediate access.

1. Watch this 2.5 minute video from the Open Data Institute titled *Open/Closed/Shared: the world of data.* Note that ‘shared’ data can also be called **mediated or controlled access** - this is often the preferred way for medical data to be published.
2. Now open this page on Open Data to see a more in-depth view of why data is sometimes open, shared or closed.
3. Have a look at one of these examples of open data from medical or health research:
    - State of the Tropics - Society - Health - Under 5 mortality
    - Hawaii Aging with HIV Cardiovascular Study, 2009-2014 (ICPSR 36389)
    - PALS (Pregnancy and Lifestyle Study)

Not all data is suitable to be openly shared in its original form, such as identifiable patient medical records. Strategies for sharing sensitive data will be explored in Thing 4: Sharing sensitive data.

Activity 3

Data sharing practices

1. Take a look at this infographic from Wiley titled *Research Data Sharing Insights* [PDF, 2.08MB] It provides a succinct overview of current data sharing practice and perceptions.

2. Note that 48% of ‘heath scientists’ say they are sharing their work. Has this been your experience? Also note where the work is shared. What implications does this have for your work / the people you work with?
3. Data sharing can also be undertaken by Industry - see what a collection of pharmaceutical companies are doing as an example. What are your thoughts on this?

Thing 4: Sharing sensitive data

Sharing sensitive data requires careful consideration, but it can be done. Find out how.

Activity 1

Sensitive data *can* be shared!

Sensitive data can be **Human data** (e.g. health and personal data, secret or sacred practices); or **Ecological data** (may place vulnerable species at risk).

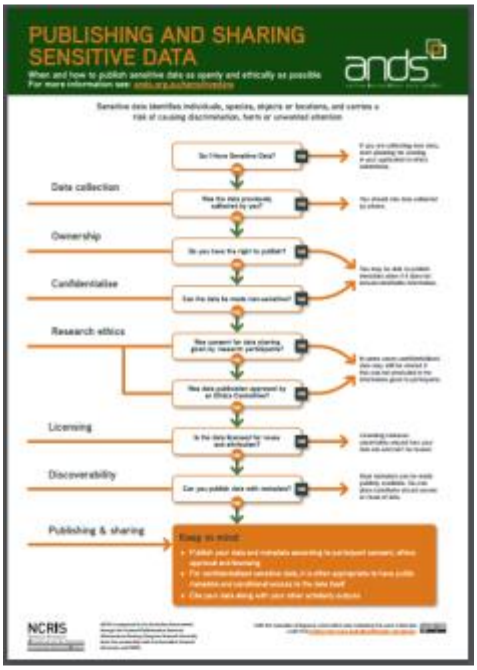
Given the nature of this type of data, you might expect that it can’t be shared and reused. But in many cases, it can be.

1. **Explore** one of these examples of published sensitive data:

    1. Remember we met the Pregnancy and Lifestyle study (PALS) dataset in Thing 3 Activity 2 as an example of open data? It shows how sensitive data can be safely de-identified and openly shared. Click on Pregnancy and Lifestyle study (PALS) and then “Go to Data Provider” to see the actual data.
    2. This 1 page story tells how sensitive data from the *Australian Longitudinal Study of Women’s Health data* has been successfully published for almost 20 years. Note the data is available through conditional access, as introduced in Thing 3: Data sharing and discovery Activity 2.

2. **How do you share and publish sensitive data?**

    1. Browse through the ANDS sensitive data webpage.
    2. Click on the Sensitive Data Decision Tree image to get an overview of issues and solutions.



3. **If you have time:** follow a couple of the links on the sensitive data page which are of particular interest to you.

Consider: Imagine you are either a researcher or a participant in a health survey:

- Participant: what questions might you first ask the researcher about intended sharing and reuse of the survey data?
- Researcher: What responses would you need to prepare to anticipate participants questions about publishing “their data for all the world to see”?

Activity 2

De-identification of data

De-identification is a process that balances the risks of producing safe data with maintaining useful data. When it is done well the risk of disclosing information referring to individuals should be negligible.

1. Explore this guide to anonymisation of medical data
2. Discover some tools and resources for information about de-identification of data.

Consider: are there any tools or resources you have come across that could help a researcher de-identify or anonymise their data?

Activity 3

Consent for data sharing

Informed consent is required from human participants before obtaining and publishing data. The best time to obtain consent is before the data are collected. Participants should, at a minimum, be informed about procedures for maintaining privacy and the conditions under which the data will be shared.

Explore one, or more, of the following consent forms that ask for permission to share research data:

- UK Data Archive sample consent form
- Global Alliance for Genomics and Health consent tools (halfway down page. Open and focus on Section C)
- Health Science Alliance Biobank Consent

Consider: would you be willing to sign that consent form?

If you have time, check out the Personal Genome Project's statement that they cannot guarantee privacy for published genomic data and therefore ask for "open consent" where participants acknowledge they may become identified. (...and if you're really keen check out this article examining the issue in more depth!)

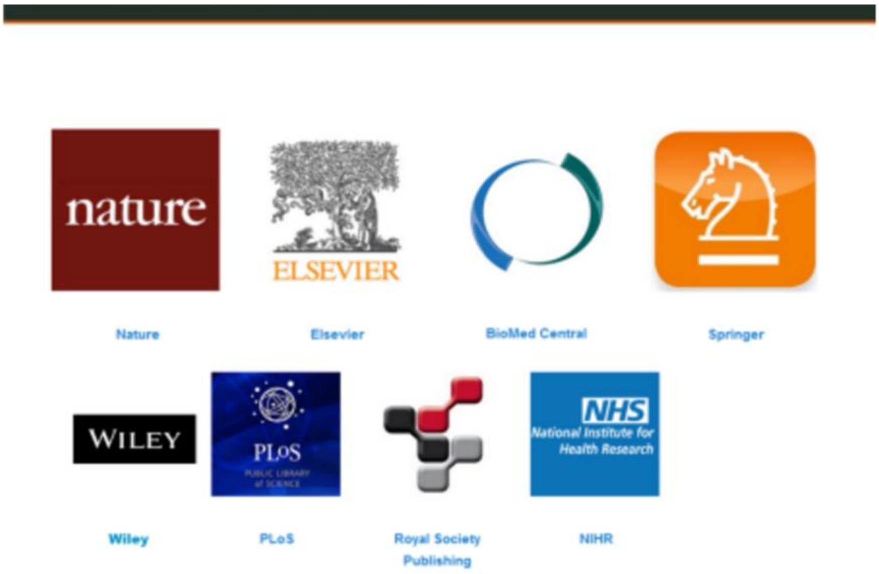
Thing 5: What are publishers & funders saying about data?

Data sharing policies are becoming increasingly common in Australia and internationally. Learn why research funders and journal publishers are particularly influential when it comes to encouraging data availability.

Activity 1

Learn about new journal data policies

More and more journal publishers are asking authors to make the data underpinning a journal article available. It's all about ensuring that the research being described in the article is based on solid, reproducible science. Thinking back to Thing 3: Data Sharing and discovery Activity 2, remember that *available* can be "open" or "shared" through mediated/controlled access.



1. Choose one of the links below to explore some of these policies.
    1. PLOS Medicine data policy, which also specifies preferred locations for the data.
    2. The British Medical Journal data policy (click on 'Data sharing').
    3. Annals of Internal Medicine data policy

2. Read this blog post Promoting research data sharing at Springer Nature about the 4 levels of research data sharing policies. Spend a bit of time clicking the links near the bottom of the post: FAQs, policies in full, list of trusted data repositories, preparing Data Availability Statements.

Consider how easy, or hard, it was for you to understand what is required upon submission to one of these journals in regard to research data.

Activity 2

Data sharing policies of major medical funders

The Australian National Health and Medical Research Council (NHMRC) funded more than $896 million in 2015 for health and medical research including 1030 new grants to universities, medical research institutions and hospitals across Australia.

1. Quickly review the NHMRC Statement on Data Sharing (2 pages) and lifecycle diagram for data sharing we saw in Thing 1.
2. International collaborations are increasingly common in our ever-connected world. Researchers in Australia are involved in projects funded by overseas bodies. Choose one of these major international funders of research and have a look at their data sharing policy:
    - Wellcome Trust in the UK
    - Bill and Melinda Gates Foundation in the US (also look at the FAQs for more information)
    - National Institutes of Health (NIH) in the US

Now, imagine it is 2020… **consider** what you think Australian research funders will be requiring of researchers who are seeking project funding.

If you have time, explore the statement supporting rapid data release of genomics data to the scientific community from the International Cancer Genome Consortium

Thing 6: Identifiers for data and people

What are DOIs and ORCIDs? These unique identifiers support data citation, metrics for data and related research objects, disambiguation of people, accurate attribution and impact metrics.

Activity 1

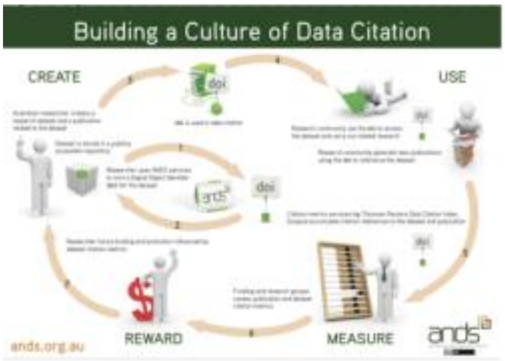
DOI’s are unique (just like you)

Digital Object Identifiers (DOIs) are unique identifiers that provide persistent access to published articles, datasets, software versions and a range of other research inputs and outputs. There are over 120 million Digital Object Identifiers (DOIs) in use, and last year DOIs were “resolved” (clicked on) over 5 billion times!

Each DOI is unique but a typical DOI looks like this:
http://doi.org/10.4225/08/50F62E0D359D5

DOIs can be used to collect citation metrics about the use of a dataset or article.

1. Start by watching this short 4.5min video Persistent identifiers and data citation explained from Research Data Netherlands. It gives you a succinct, clear explanation of how DOIs underpin data citation.
2. Have a look at the poster *Building a culture of data citation* (also shown below) - follow the arrows to see how DOIs are attached to data sets.



3. Let’s go to a data record which shows how DOIs are used. Click on this DOI to ‘resolve’ the DOI and take us to the record: http://dx.doi.org/10.4225/22/55BAE9DBD9670 (Population Health data collection for the City of Greater Bendigo)

The same record has been syndicated to Research Data Australia. Click on the DOI at the bottom of the page, under ‘Identifiers’. No matter where the DOI appears it always resolves back to its original dataset record to avoid duplication. i.e. many records, one copy.

If you have time: Want to know more about DOIs? Flick through the ANDS DOI Guide page

Consider: should DOIs be routinely applied to all research outputs? Remember that DOIs carry an expectation of persistence (maintenance costs etc.) but can be used to collect metrics as well as link articles and data (evidence of impact).

Activity 2

Getting to know ORCID

What about identifiers for people? Think about the many forms a person's name may take or common names. Is the author JK Rowling the same person as Joanne Rowling and Jo Rowling? More than 38,000 Americans have the name James Smith!

Universities, funders and publishers worldwide now use ORCID to differentiate between people with the same name by assigning individuals with a unique identifier.

1. Let's start by going to ORCID. In the search box at the very top of the page, enter *John David Burton* to search the ORCID registry. Scan the list of results to find the entry for John David Burton. How many versions of his name do you see?

2. Now enter *Toby Burrows* into the search box. Open his ORCID record to see a wonderful example of a rich ORCID record. Note he has combined his ResearcherID and his Scopus Author ID with his ORCID. Scroll through his list of works and look closely at *Source* to see the wide range of sources of his publications.

You can now choose from 3 activities that will get you in touch with ORCID.

Option 1. Don't have an ORCID record but would like one?

Use this time to create your ORCID profile and make it as complete as possible.

1. Visit ORCID and follow steps 1 and 2.
2. When you're done, add your ORCID iD to your email signature, LinkedIn profile and blog
3. Send your new ORCID iD to a colleague and ask for some feedback on your profile

Option 2. Already have an ORCID?

When was the last time you logged in to update or enhance your profile? You may be surprised at the additional functionality now available.

1. Read Alice Meadow's blog post Six Things to do now you have an ORCID iD
2. Now go to your ORCID profile and update it to be as current and complete as possible
3. When you're done, add your ORCID to your email signature, LinkedIn profile and blog
4. Consider using the new QR code feature for your ORCID iD in new and unchartered ways

Option 3. Don't want an ORCID?

Get up to date with the latest features, functionality and news on the ORCID blog and explore the Australian ORCID Consortium (most Australian universities are members).

Consider how ORCID can be used to enhance your online profile.

Thing 7: Data citation for access & attribution

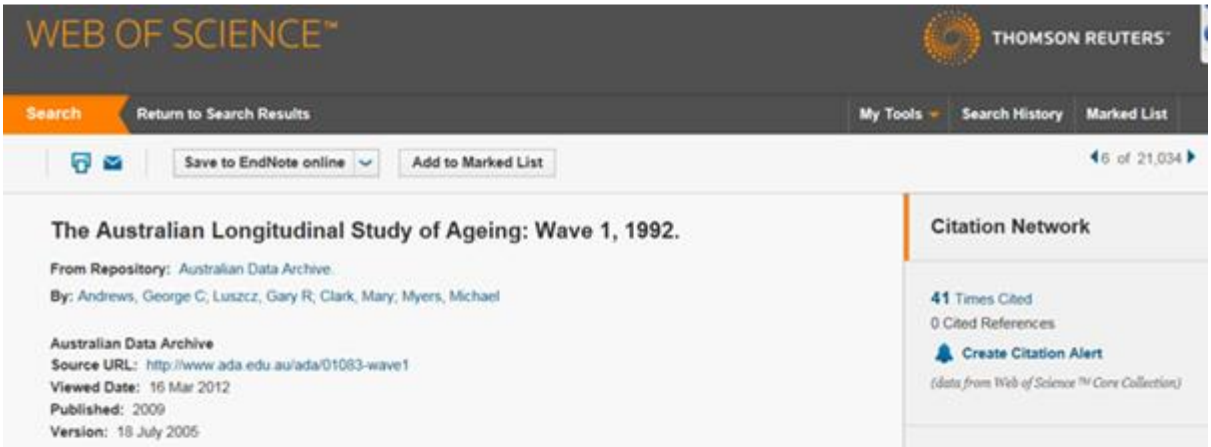
Citation analysis and citation metrics are important to the academic community. Find out where data fits in the citation picture.

Activity 1

Getting more out of your citation

Data citation continues the tradition of acknowledging other people's work and ideas. Along with books, journals and other scholarly works, it is now possible to formally cite research datasets and even the software that was used to create or analyse the data.

1. Have a look at this dataset from the Australian Longitudinal Study of Aging. Data Citations are available from the Thomson Reuters Data Citation Index - note the number of times this dataset has been cited



The Australian Longitudinal Study of Ageing: Wave 1, 1992.

From Repository: Australian Data Archive

By: Andrews, George C; Luszcz, Gary R; Clark, Mary; Myers, Michael

Australian Data Archive
Source URL: http://www.ada.edu.au/ada/01083-wave1
Viewed Date: 16 Mar 2012
Published: 2009
Version: 18 July 2005

41 Times Cited
0 Cited References
(data from Web of Science™ Core Collection)

2. Scan through the ANDS introduction to data citation
3. Now look at the Hutchinson Drought Index data record in Research Data Australia.
   a. This research data makes cross disciplinary connections between episodes of drought and correlated increases in rural mental health issues.
   b. The beauty of this record is that it shows the entirety of the research outputs - publications, software, related datasets and more - all of which are citable.
   c. Click on the 'Cite' button to see the similarities between the formats for citation of data and other scholarly publications. Did you notice that, as yet, there are no citation metrics to this record?

Consider: Data citation is a relatively new concept in the scholarly landscape and as yet, is not routinely done by researchers, or expected by most journals. What could be done to encourage routine citation of research data and software associated with research outputs?

Activity 2

Data Citation Principles

The Force11 Joint Declaration of Data Citation Principles are based on the premise that data citation, like the citation of other evidence and sources, is good research practice and is part of the scholarly ecosystem supporting data reuse.

Since they were published in 2014, the Principles have been endorsed by numerous individuals and more than 100 data centres, publishers and societies.

1. Start by reading the Force11 Principles
2. Then browse the list of people and organisations that have endorsed the Principles

Consider: Given such support and clear direction, why do you think data citation has not been uniformly adopted, so far, across all disciplines?

Thing 8: Licensing data for reuse

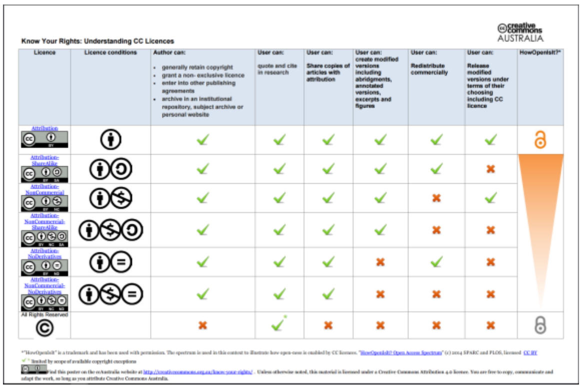
Understand the importance of data licensing, learn about Creative Commons and see how licensing data can assist in creating links with business and industry.

Activity 1

The cans and cannots of licensing

Consider this scenario: You've found a dataset you are interested in. You've downloaded it. Excellent! But do you know what you can and cannot do with the data? The answer lies in data licensing. Licensing is critical to enabling data to be reused and cited.

1. Start by reading this brief introduction to licensing research data.
2. Now have a closer look at the poster from creativecommons.org.au Click on the descriptions for more information. Notice they have used CC BY as the licensing information at the bottom of the poster so you know what you can do with the poster itself.



Know Your Rights: Understanding CC Licences

creative commons AUSTRALIA

Licence	Licence conditions	Author can: • generally retain copyright • grant a non- exclusive licence • enter into other publishing agreements • archive in an institutional repository, subject archive or personal website	User can: quote and cite in research	User can: Share copies of articles with attribution	User can: create modified versions including abridgments, annotated versions, excerpts and figures	User can: Redistribute commercially	User can: Release modified versions under terms of their choosing including CC licence	HowOpenIsIt?*
Attribution BY	BY	✓	✓	✓	✓	✓	✓	🔓
Attribution-ShareAlike BY SA	BY SA	✓	✓	✓	✓	✓	✗	
Attribution-NonCommercial BY NC	BY NC	✓	✓	✓	✓	✗	✓	
Attribution-NonCommercial-ShareAlike BY NC SA	BY NC SA	✓	✓	✓	✓	✗	✗	
Attribution-NoDerivatives BY ND	BY ND	✓	✓	✓	✗	✓	✗	
Attribution-NonCommercial-NoDerivatives BY NC ND	BY NC ND	✓	✓	✓	✗	✗	✗	
All Rights Reserved ©		✗	✓*	✗	✗	✗	✗	🔒

*"HowOpenIsIt" is a trademark and has been used with permission. The spectrum is used in this context to illustrate how open-ness is enabled by CC licences. "HowOpenIsIt? Open Access Spectrum" (c) 2014 SPARC and PLOS, licensed CC BY

✓* limited by scope of available copyright exceptions

Find this poster on the ccAustralia website at http://creativecommons.org.au/know-your-rights/ . Unless otherwise noted, this material is licensed under a Creative Commons Attribution 4.0 licence. You are free to copy, communicate and adapt the work, so long as you attribute Creative Commons Australia.

If you have time, flick through the (21) slides from a presentation ANDS gave in June 2016 about licensing data.

Activity 2

Licensing for data reuse

Enabling reuse of data can speed up research and innovation. Licensing is critical to enabling data reuse.

1. Start by watching this 4.30mins video in which Dr Kevin Cullen from the University of New South Wales explains their approach to licensing which aims to strengthen the University's relationship with business and industry.
2. Now read the Australian Government Public Data Policy Statement (2 pages) that was released in December 2015. Note in particular, the last dot point.
3. If you have questions, ANDS has a list of research data licensing FAQs

Discover: Does your institution have a policy or guidelines around data licensing?

Thing 9: Describing data: metadata and controlled vocabularies

Metadata elements are the lifeblood for finding and reusing research data. Data is only as valuable as the metadata which describes and connects it. In addition to selecting a metadata standard or schema, whenever possible you should also use a controlled vocabulary. A controlled vocabulary provides a consistent way to describe data.

Activity 1

Metadata: your new best friend

Metadata is structured information about a resource that describes characteristics such as content, quality, format, location and contact information. Creating metadata to describe research data is very similar to the process for descriptive cataloguing of library resources.

Metadata schema are sets of metadata elements (or fields) for describing a particular type of information resource. Numerous metadata schema exist for describing research data across different disciplines.

1. Start by watching this short (2:29 min) video about medical metadata

2. Read the short ANDS Introduction to Metadata to understand what metadata is and why is it the lifeblood of research data sharing!

3. Let's revisit at least one of the good quality metadata records for medical data we met in previous Things. Why do you think these records are considered 'good quality'? *Hint*: consider both the type and quality of information provided. What metadata included in this record help discovery and reuse of the data?

- The Australian Longitudinal Study of Ageing: Wave 1, 1992
- National Survey of Midlife Development in the United States (MIDUS)
- Population Health data collection for the City of Greater Bendigo

If you have time, explore the UK Digital Curation Center's Directory of Disciplinary Metadata. You might find a schema that is applicable to your research!

Consider why, if metadata is the lifeblood of data discoverability and reuse, is it often neglected or not richly done when data is published.

Activity 2

Control your language, please!

Controlled vocabularies significantly improve data discovery. They make data more shareable with researchers in the same discipline because everyone is ‘talking the same language’ when searching for specific data e.g. medical conditions, plants, animals etc. There are hundreds of controlled vocabularies used in medical practice, many of which can be utilised in research.

1. Explore one of the following:
    - ICD-10 for classification of diseases
    - RxNorm for clinical drugs (US based, so has US spelling and brand names)
    - SNOMED-CT for clinical terminology, which has an Australian version, SNOMED-CT-AU
    - MeSH for medical subject headings (particularly useful for making your research more findable in literature searches)
2. Have a flick through these (15) slides about health vocabularies

Consider: How would the use of a controlled vocabulary be helpful within your field?

Thing 10: Planning to publish

Some research institutions and research funders now require researchers to submit a Data Management Plan (DMP) for new projects. What should a DMP cover? Could you write or help with one?

Activity 1

Essentials of a Data Management Plan

We have explored several important data management concepts during these 10 Things. The best research practice is to consider these at the start of a project. By planning ahead the research team can improve research efficiency, guard against data loss, enhance data security, and ensure research data integrity and replicability.

A Data Management Plan (DMP) documents how data will be managed, stored and shared during and after a research project. Some research funders and human research ethics committees are now requesting that researchers submit a DMP as part of their project proposal.

1. Start by scanning this short introduction to Data Management Plans
2. Now browse through some public DMPs from either the CDL or DataOne, and open up one or two of the DMPs to see the type of information they capture:
    - California Digital Library
    - DataOne

You will have noticed that DMPs can be very short, or extremely long and complex.

Consider what are the 2 or 3 pieces of information essential to include in a medically related DMP and why these in particular?

Activity 2

Preparing a Data Management Plan

We have explored several important data management concepts during these 10 Things. Many Data Management Plan (DMP) templates are now freely available for reuse.

1. Choose one DMP template or guide from the Australian, International or Discipline examples at the bottom of the ANDS DMP page
2. Spend 5-10 minutes starting to complete the template, based on a research project you have been involved with in the past.

Consider the strengths and weaknesses of your chosen template.

What’s next?

Reflect on the changes you could, and perhaps should, make in research data management practices which will enable the ethical and efficient publication of health, medical and clinical data for reuse by the research community.

Consider

- Learning more about research data management by browsing the 23 (research data) Things program which includes data management issues not covered here.
- Making connections to other people who ‘know data’ in your institution e.g. Librarians, Repository managers, IT, Researchers
- Reviewing your technical skills through either:
  - Online courses such as those mentioned in Thing 21: Tools of the (dirty) data trade
  - ResBaz or one of the Carpentry courses