10 Eco Data Things
Introduction

10 Eco Data Things is an opportunity to explore issues surrounding management of research data, specifically for people working with ecological 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 reuse.

How can I work through these 10 Eco Data Things?

- All Things have a choice of 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.
- Set aside some time each week to do one Thing or do a ‘Crash course’ and work through them all in a few hours.

You don't have to do them all

- You can do as much or as little of the Things as you want to do, or need to know.
- Choose from the different activities, doing one or all of them.

In a group

- 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’.

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

Please note: this is a snapshot in time - research data as it was in 2018. You may need to check resources and update resources and links to include more recent initiatives and policy changes.

Have fun!
Table of Contents

Introduction ........................................................................................................................................2
Table of Contents ...............................................................................................................................3
Thing 1: Getting started with research data ....................................................................................4
Thing 2: Issues in research data management ..................................................................................5
Thing 3: What are publishers & funders saying about data? ..............................................................7
Thing 4: Data management plans ....................................................................................................8
Thing 5: Describing data: metadata and controlled vocabularies ....................................................10
Thing 6: Data sharing practices .........................................................................................................13
Thing 7: Data discovery and deposit ................................................................................................16
Thing 8: DOIs, data citation and metrics ..........................................................................................17
Thing 9: Licensing data for reuse .....................................................................................................20
Thing 10: Spatial data ........................................................................................................................22

About ecoEd

ecoEd is a new innovative training and skills development program for university lecturers, researchers and industry professionals. The program is developed to enhance the translation of Australia’s eResearch infrastructure to the Ecoscience community by educating and upskilling the next generation of environmental scientists.

ecoEd provides access to information resources that can be used in courses that focus on topics such as ecology, biogeography, environmental management and spatial analysis. These resources explain the theoretical underpinnings of ecological models, and provide real-world examples to show how these concepts can be analysed in online tools.

More info: www.ecoed.org.au

ecoEd is sponsored by:
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 is research data?

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 and collation/preparation often needs to include algorithms, scripts, software. It's not 'just data'.
2. Then open up the dataset details in these repositories:
   - Atlas of Living Australia: species information about the endangered Corroboree Frog
   - LTERN: Connell Rainforest Plot Network, long-term data
   - AEKOS - TERN: SWATT leaf carbon isotope ratios
     Leaf stable carbon isotope response of terrestrial plants to aridity along a bioclimatic gradient.
   - Australia’s Environment explorer: gridded data
   - CSIRO: Marmion Benthic Survey 2007 data record

Consider: How could complexity and range of data formats affect access and re-use possibilities?

Activity 2: The data 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 ground breaking discoveries" to "telling the world about it".

Research Lifecycle Models provide a high-level overview of the stages and actions during the research lifecycle required for successful management of data. Lifecycle diagrams can be very complex, and the complexity is not always related to big data.

1. Take a look at either:
   a. DataOne data lifecycle
   b. UK Data Archive Research Data Lifecycle
   b. DCC Curation Lifecycle Model
**Consider:** What other things impact how data is managed through the lifecycle, e.g. funding policies which make sharing the data mandatory? Have you been through all of the steps outlined in this lifecycle? If not, which ones are new to you?

![Diagram of data lifecycle stages: Planning, Collecting, Processing and Analysing, Publishing and Sharing, Preserving, Re-Using Data.]

**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: Managing data for reuse**

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 re-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. Have a look at the [various case studies](#) that show how Queensland Government Data is used in apps that for example can identify areas with air or water pollution, list native plants that are suitable to grow in your area, gather weather information for predictive crop modelling, and more!
2. Check out Figure 1 in this paper to get a quick overview of ‘information entropy’.

3. This 4.40 minute 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!?

4. As you watch the cartoon jot down the data management mistakes which interest or appal you.

5. Now, scan through the dot points in the Consider the following.... section of the University of the Sunshine Coast's Library Guide 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 are hearing with increasing frequency. Data management for Big Data brings much complexity - citing dynamic data, software, high volume compute, storage costs, transfer of petabytes of data, preservation, provenance, and more.

1. 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: Your views on whether the concept of 5Vs is useful to support better management and reuse of "Big Data". If you don't think 5Vs is of value, is there another framework or concept model which could be useful for exploring data management for big data?
Thing 3: 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: 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.

Choose one of the links below to explore some journal publisher policies:

1. PLOS ONE Data policy
2. Springer Nature Data policy
3. Elsevier Data policy
4. Wiley Data Sharing and Citation policy

Consider: How easy, or hard, it was for you to understand what they are asking researchers to do in regard to research data?

Activity 2: Data journals and data papers

Explore this relatively new form of data publishing: the data journal. Data journals focus on describing data, rather than discussion and analysis of the data (as in traditional journals).

1. Data journals, such as Scientific Data, are publications whose primary purpose is to expose datasets. They enable the author to focus on the data itself, rather than producing an extensive analysis of the data which occurs in the traditional journal model. Read this short introduction: What are data journals?
2. Ecology is a journal that accepts ‘data papers’ in addition to other more traditional papers. Browse:
   a. The PanTHERIA data paper, describing a species-level data set compiled for analysis of life history, ecology, and geography of all known extant and recently extinct mammals.
   b. The BAAD data paper, describing the Biomass And Allometry Database for woody plants.
   c. The journal’s data paper instructions.

Consider: Why do you think authors might choose to describe their data in data journals or journals that accept data papers?
Activity 3: Research funders and data sharing

Funders of research are increasingly developing policies requiring data to be well managed and shared.

1. Start by reading the ARC Research Data Management policy.
2. Check out the ANDS guide to filling out the data management section in ARC grant applications.
3. International collaborations mean that research teams need to consider Australian as well as the data sharing policies of their overseas collaborators. Choose one of these major funders of research overseas and have a look at their data sharing policies:
   a. National Science Foundation in the US, and note the Data Management Plan requirement (see Thing 4: Data Management Plans).
   b. Wellcome Trust in the UK (1 page).

Consider: The ARC Research Data Management Policy says “Since 2007, the ARC has encouraged researchers to deposit data arising from research projects in publicly accessible repositories”. Why do you think this hasn’t happened across the all ARC funded research yet?

Thing 4: Data management plans

Some research institutions and research funders now require researchers to submit a Data Management Plan (DMP) for new projects. What is a DMP and what should it cover?

Activity 1: An introduction to Data Management Plans

A Data Management Plan (DMP) documents how data will be managed, stored and shared during and after a research project. It is an important part of the planning process for a project, and is very helpful when researchers reach later stages of the data lifecycle (see Thing 1, Activity 2).

1. Start by scanning this short introduction to Data Management Plans.
2. Check out Figure 1 (page 2) in the ANDS Research Data Management in Practice report to see an overview of where DMPs fit into the bigger picture of data management.
3. Now browse through some public DMPs using the following links and open up one or two of the DMPs to see the type of information they capture:
   a. DMP Tool
   b. DataOne
4. Many Universities have their own guidelines about data management plans, explore one of these or find out if your Institution has one
Consider: You will have noticed that DMPs can be very short, or extremely long and complex. What do you think are the 2 or 3 pieces of information essential to include in every DMP and why that is?

Activity 2: Data Management Plan Tools

DMP tools enable researchers to create, review, and share data management plans that meet institutional and funder requirements. To make DMPs more useful and effective, these tools are being re-imagined in new ways.

1. First have a look at the The DMPTool 90 sec video to see what the DMPTool offers researchers and institutional data managers.
2. Then read [this blog](#) about how the Digital Curation Centre UK (DCC) and University of California Curation Centre (UCCC) have joined forces to create a new tool, called DMPRoadmap.

3. Next browse the DMPRoadmap [development wiki](#) on GitHub.

**Consider:** The DMPRoadmap project is aiming to enable DMP’s to be active, dynamic, machine-readable and [FAIR](#). Can you think of any additional goals or further enhancements?

---

**Thing 5: Describing data: metadata and controlled vocabularies**

Metadata 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 datasets.

**Activity 1: Metadata: describing research data**

Metadata is your best data 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. The metadata standards most relevant to Ecoscience are EML, ISO 19115 and ISO 19139 (see Activity 2).

1. Start by reading the ANDS [Introduction to Metadata](#) to understand what metadata is and why is it the lifeblood of research data sharing!

2. Now look closely at one of these good quality metadata records for research data:
   b. Van Dijk et al. (2017) OzWALD Gridded Environmental Data for Australia. [doi: 10.4225/41/5934faaf27397b](#)

   Why do you think this record is considered ‘high quality’? 
   Hint: consider both the type and quality of information provided. What metadata included in this record help discovery and reuse of the data? Look back at the ANDS Introduction to metadata for ideas and think about what we’ve looked at in previous Things.

**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: Metadata schema and standards

A metadata standard is a schema that has been formally approved and published, with governance procedures in place to maintain and update the standard. Examples include ANZLIC and DDI (Document, Discover and Interoperate).

1. Have a brief look at one of the following metadata standards relevant to ecological data:
   a. [Darwin Core standard](http://www.dwc.org)
   b. [Ecological Metadata Language (EML)](https://www.earthobservations.org/eml)
   c. [ISO 19115](https://www.iso.org/obp/ui/#iso:std:iso:19115:ed-1) for geographic data

Numerous metadata standards exist and the standard chosen to describe resources such as research data should be appropriate to the project or discipline.

1. Start by reading this short guide to [Evaluating and Selecting a metadata standard](https://www.ukoln.ac.uk/information-october-2010/). Keep this Guide open.
2. Choose one disciplinary standard which sparks your interest from the UK Digital Curation Centre’s [Directory of Disciplinary Metadata](https://www.ukoln.ac.uk/dcm/). Use the Guide to critique your chosen standard.

**Consider:** Describe the particular standard you evaluated and perhaps add other reasons to recommend (or not) that standard.
**Activity 3: Control your language, please!**

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 - location, time, place name, subject. Read a short explanation of controlled vocabularies [here](#).

Controlled vocabularies significantly improve data discovery. It makes data more shareable with researchers in the same discipline because everyone is ‘talking the same language’ when searching for specific data e.g. plants, animals, medical conditions, places etc.

Some examples of vocabularies used in ecology are:

1. [Environmental Thesaurus](#) (EnvThes)
2. [Global Change Master Directory](#) (GCMD) Keywords
3. [Long Term Ecological Research](#) (LTER) – and see it visually represented [here](#)
We are going to see some controlled vocabularies in action in the Atlas of Living Australia (ALA).

1. Do a search in the ALA search engine. Type “whale” in the search box and click on search. Choose one of the records listed and click on the (red text) View record link.
2. Any metadata field where you see Supplied... tells you that the information supplied by the person who submitted the record (often a ‘citizen scientist’) has been changed to the controlled vocabulary being used in metadata fields e.g. Observer, Record date and Common name.
3. Have a scroll down the record and consider how many of the metadata fields probably have a controlled vocabulary in use (e.g. taxonomy, geospatial etc.).

Consider: How do you think we could encourage people to use controlled vocabularies in their data descriptions?

**Thing 6: Data sharing practices**

Sharing research data can be beneficial as such data are a valuable resource and in many cases have significant value beyond their original use, such as:

- It promotes the research and the researchers that created the data.
- It can lead to re-use, discovery, and increased citation.
- It can provide an important resource for education and training.
- It reduces the cost of duplicating or collecting similar data.
- It maximises transparency and accountability.
- It can lead to new collaborations between the data creator and user.

Here we are exploring some of the concepts around data sharing, and in Thing 7 we will look at where data can be described and shared.

**Activity 1: An introduction to ‘open’, ‘shared’ and ‘closed’ data**

You may have noticed that not all the data described in the repositories in Thing 3 was available for immediate access. This activity explains why different datasets may have different access conditions.

1. Watch this 2.5 minute video from the Open Data Institute titled Open/Closed/Shared: the world of data.
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. Note that the Australian and international conversation around data sharing is now also using the term FAIR (Findable, Accessible, Interoperable, Reusable). Have a look at the [ANDS introduction to FAIR](#).

4. If you have time, go to [Research Data Australia](#) and try searching for data that is 'open'. Hint: Look for the option to limit your search to data that is Publicly accessible online.

**Consider:** What are some reasons for data to be not publicly accessible or ‘open’?

### Activity 2: Data sharing practices

Repositories are one means by which research data may be shared but in order to get data into repositories, research teams must be willing to publish their data: there are huge differences between data sharing practices by country and by discipline.

1. Take a look at these infographics from Wiley titled [Research Data Sharing Insights](#) (2014), and [Global Data Sharing Trends](#) (2016). They provide a succinct overview of current data sharing practice and perceptions.

2. Now look closely at the sections titled 'Global Data Sharing Trends' and 'Data Sharing By Discipline'.

**Consider:** Why do you think there are differences between disciplines and countries? What changes over time can you see between the two surveys?

---

**Top 4 researcher motivations**

<table>
<thead>
<tr>
<th>Motivation</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increase the impact and visibility of my research</td>
<td>39%</td>
</tr>
<tr>
<td>Public benefit</td>
<td>35%</td>
</tr>
<tr>
<td>Transparency and re-use</td>
<td>31%</td>
</tr>
<tr>
<td>Journal requirement</td>
<td>29%</td>
</tr>
</tbody>
</table>

**Top 4 reasons why researchers are hesitant to share their data**

1. 50% - Intellectual property or confidentiality issues
2. 31% - Ethical concerns
3. 23% - I am concerned about misinterpretation or misuse of my research
4. 22% - I am concerned that my research will be scooped
Activity 3: Sharing sensitive data

Major, familiar, categories of sensitive data are **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** the [Sensitive species lists](https://www.ala.org.au/data/species) of the Atlas of Living Australia (ALA): these lists are used to obscure occurrence locations for the species listed to protect them. The location information is stored in ALA, but the general public can only see the associated information of the record and generalised coordinates of the location.
2. Read this article in [The Conversation](https://theconversation.com) about publishing sensitive data through secure data repositories to help conserve and manage the world’s most threatened species.
3. **How do you share and publish sensitive data?**
   a. Read this article about [Responsible Data Publishing](https://tern.org.au) from TERN.
   b. Read how [ALA handle sensitive data](https://www.ala.org.au) (towards the bottom of the page).
   d. The NSW Department of Environment, Climate Change and Water [Sensitive Species Policy](https://www.environment.nsw.gov.au) manages access to sensitive species locational data by categorisation of sensitive species into 3 groups according to risk status (pages 9-10).
   e. Scan the [ANDS sensitive data](https://ands.org.au) webpage.

**Consider**: What conditions can make ecological data sensitive?
Thing 7: Data discovery and deposit

Repositories and portals play an important role in making research data discoverable and accessible. Here we explore places that data can be shared and is currently being shared.

Activity 1: Data repositories

Repositories enable discovery of data by publishing data descriptions ("metadata") about the data they hold - like a library catalogue describes the materials held in a library. Most repositories also provide access to the data itself, but not always.

1. Many Australian Universities have a repository in which data can be deposited. Have a look to see if your University has one. Often the Library manages the repository, or has staff who can help you to deposit your data there.
2. Explore some of the repositories in which you can discover and deposit ecological data:
   a. Atlas of Living Australia (ALA)
   b. Terrestrial Ecosystem Research Network (TERN)
   c. Biodiversity and Climate Change Virtual Laboratory (BCCVL) Data Portal
   d. Global Biodiversity Information Facility (GBIF)
   e. NatureServe Explorer
   f. TERN - AEKOS (Australian ecological plot data)
   g. Catalogue of Life
   h. Encyclopedia of Life
   i. Global Plant Trait database
   j. Fishes of Australia
   k. FishBase
   l. Ocean Biogeographic Information System
   m. ReefBase
1. Or search the Registry of Data Repositories:
   a. Click on Browse > By Country > click on Australia in the map
   b. There are a surprising number of data repositories listed for Australia. Does this present all the research data repositories Australia has to offer: is anything missing?

Consider: Find a dataset relevant to your interests in one of the repositories listed.
Activity 2: Australia’s national research data catalogue

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, including University repositories and national repositories like TERN. Data is not deposited into RDA, it harvests data descriptions ("metadata") from the other repositories, thus providing a discovery point for a wide range of data.

2. Have a close look at the record. Click on ‘Go to Data Provider’ to see the record in the original repository.
3. Spend a few minutes exploring RDA:
   a. Try browsing or searching on a topic of interest.
   b. See which institutions contribute metadata records to RDA.
   c. Explore a record or two in depth.

Consider: What could be the future impact of having a national research data catalogue?

Thing 8: DOIs, data citation and metrics

DOIs are unique identifiers that support data citation, metrics for data and related research objects, and impact metrics. Citation analysis and citation metrics are important to the academic community. Find out where data fits in the citation picture.

Also note that unique identifiers are being developed for people too! Universities, funders and publishers worldwide are starting to use ORCID to differentiate between people with the same name by assigning a unique identifier. Interested? Start by checking out the ORCID website.

Activity 1: DOIs

Digital Object Identifiers (DOIs) are a type of ‘persistent identifier’. They 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 in 2016 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
1. Start by watching this short 4.5 minute video Persistent identifiers and data citation explained from the 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 and follow the arrows to see how DOIs are attached to data sets and are used in data citation.

3. Let’s go to a CSIRO data record which shows how DOIs are used. Click on this DOI to ‘resolve’ the DOI and take us to the record: http://doi.org/10.4225/08/50F62E0D359D5
   a. Click on both the Description and Data tabs and note how many times the DOI is used in this data record. It will give you some ideas why DOIs are given such prominence.
   b. This same record has been syndicated to Research Data Australia.
   c. Click on the Cite icon on the upper left of the record (under the green Go To Data Provider tab).
   d. Now click on the DOI in the data citation. No matter where the DOI appears it always resolves back to its original dataset record to avoid duplication. i.e. many records, one copy.

4. DOIs can also be applied to grey literature, a term that refers to research that is either unpublished or has been published in non-commercial form, such as government reports. For example, reports like this: https://doi.org/10.4225/08/5ac520bd3fbc7

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).

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

Activity 2: Altmetrics

Alternative metrics or ‘Altmetrics’ count the number of views, number of downloads, social media ‘likes’ and recommendations associated with a dataset. Because of their immediacy, altmetrics can be an early indicator of the impact or reach of a dataset; long before formal citation metrics can be assessed.

1. Start by exploring the altmetrics for this Oecologia paper.

2. Now look at the altmetrics for this Phylogenomics article published in Science. Note the number and pattern of downloads for this article since it was published in November 2014.
   a. Now click on the “donut” or the link to ‘More Details’ to see the wealth of information available.
   b. Look also at the associated data in Dryad noting that the data has been assigned a DOI. Can you see how many times the record has been viewed?

By way of comparison, as of early April 2016:
the same dataset had been cited once in Thomson Reuters Data Citation Index.
- the article had been cited 143 times in Web of Science.

**Consider:** Do you think altmetrics for data have value in academic settings? Why, or why not?

**Activity 3: Citing research data**

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

1. Start by looking at this [Weddell Seal dataset](#). Check out how many times it has been cited. This citation count has been measured by the Clarivate Analytics Data Citation Index.
2. Scan through the ANDS introduction to data citation and the Global Biodiversity Information Facility (GBIF) [citation guidelines](#).

**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?

**If you have time:** The Force11 Joint Declaration of Data Citation Principles are a set of principles for citing data. They 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, 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](#).

Image from: https://www.enago.com/academy/china-open-science-open-data-manadate-released/
Thing 9: 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: Why license research data?

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? Alternatively, consider that you have published a dataset. How do you communicate to the secondary user what they can and cannot do with it? 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. 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.
3. Check out the data licensing policy and FAQ from TERN.
4. See what ALA says about data licensing.
5. Check out the licence chooser from Creative Commons, which walks you through the decision of which licence is appropriate for your purpose.

Consider: If you were considering licensing a dataset on something which may have commercial value to others - what licence would you apply?

Activity 2: Data licences: unlock data for innovation

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

1. Start by watching this 4.30-minute 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 by Prime Minister Malcolm Turnbull in December 2015. Note in particular, the last dot point.

Consider: What are the possible implications for data managers, researchers, librarians and others as we move forward with the National Science and Innovation Agenda? Does your institution have a policy or guidelines around data licensing?
Activity 3: Data licensing in practice

Not all research data that is shared is licensed for reuse. It should be!

1. Start by going to Research Data Australia and search for research data using keywords e.g. ‘water quality’ or ‘cancer’. Note the number of records in your result set.
2. Now “Refine Results” to limit your search to only those datasets with an ‘open’ licence. Note the number of records in this result set.
3. Look closely at the specific Licensing information on a small sample of those records with ‘open’ licences. How easy or difficult it is to work out if the data can or can’t be reused e.g. for commercial purposes? with international collaborators?

Consider: Assigning creative commons licences is not routine. Suggest one tip for encouraging uptake of these licences.

[Images of Creative Commons licences]

Creative Commons Licences Explained by by MaconEastLibraryProject is licensed under CC BY-SA 3.0
Thing 10: Spatial data

The importance of spatial data is ever increasing. Many of the societal challenges we face today such as food scarcity and economic growth are inherently linked to big spatial data. In fact, it is often said that 80% of all research data has a geographic or spatial component. It is useful then, for all of us to have an understanding of spatial data.

Activity 1: Spatial data: maps and more

1. Start by watching this incredible, inspiring video (3.59 min) from the University of Wollongong’s PetaJakarta project. It shows innovative ways of combining social media and geospatial data to save lives.
2. Now read how GIS can be used in Conservation Biology.
3. This video combines a range of different data visualisations depicting the human impacts on our environment.
4. Geospatial data is fundamental to Australia’s economic future. Check out this very short article about how GeoScience Australia is mapping the mineral potential of our continent - a world first!

**Just for fun:** Enter your address in the Atlas of Living Australia and see what birds and plants have been reported in your street or suburb. You may be surprised at how ‘alive’ your street is!

**Consider:** Why do you think these geospatial visualisations are so powerful?

Activity 2: Spatial Data concepts

There are many types and sources of geospatial data. If you are new to the world of geospatial data, you will probably appreciate some ‘busting’ of the jargon of geospatial data.

1. Start by reading this Fundamentals Chapter to learn more about maps, projections, coordinate systems, datums and GIS.
2. Want more? Continue with this blog about Finding and Making Sense of Geospatial Data on the Internet which explains some basic geospatial data file formats and concepts.
3. Prefer watching? Most of these concepts are also explained in this video.
4. Read more about two important aspects of spatial data: scale and resolution.

**Consider:** How would you give an explanation of two new terms you have just learnt?
Activity 3: Using and Visualising Spatial Data

Spatial data can be used in many ways, and there are many tools that you can use to manipulate and display spatial data.

You can try one of the tools below. Do one, or do them all and compare the results.

1. **13 Free GIS Software Options**: Map the World in Open Source
   Browse through this site for ideas for free, open source geospatial software; the descriptions often include discipline specific advice. Download one and try your hand at mapping.

2. **Spatial data visualisation with R**: for those who have done the R modules in Software Carpentry - this might be a good activity to flex your R muscles! Want more? Here are some more [R tutorials](#).

3. Create a map using [Google Fusion Tables](#): this offers lots of features, but you need a Google account. The excellent Google Fusion tutorial uses butterfly data to show you how to import data, map the data and customise your map.
The **Open Geospatial Consortium (OGC)** is an international not-for-profit organization that develops open standards for the geospatial community. OGC through their dedicated global members have developed several standards to share geospatial data. Some of the most commonly use standards are:

a. **Web Map Service** (WMS): a standard web protocol to query and access geo-registered static map images as a web service. The outputs are images that can be displayed in a browser application.

b. **Web Feature Service** (WFS): a standard web protocol to query and extract geographic features of a map, these are typically attributes of a map. The latest version of WFS (3.0, Dec 2017) has created a lot of excitement in the community.

c. **Web Coverage Service** (WCS): provides access to geospatial information representing phenomena that are variable over space and time, such as satellite images or aerial photos. The service delivers a raster image that can be further interpreted and processed.

**Geoserver** is the most popular open source reference implementation of WMS, WFS and WCS standards.

**Consider:** The data world is hungry for Geospatial tools and metadata and there is growing demand for people with these skills. How can these skills be encouraged in your institution?
What’s next?!

Well done! You have come to the end of this module and can now celebrate your new-found skills and knowledge, and hopefully put them into practice. But your EcoScience education journey can go much further! There are so many ways to keep in touch and up to date with data as well as familiarize yourself with other resources, tools and platforms that provide data and analytical tools. The info below is just a handful of resources for you to continue your journey.

Further learning

There is still so much more to explore and learn! Check out one of these resources:

- Want to learn more about research data management? Browse the [23 (research data) Things program](#), which includes data management issues not covered in this module.
- Explore other [ecoEd modules](#) on the website.
- Make connections with other people who ‘know data’ in your institution e.g. librarians, repository managers, IT specialists, researchers.
- Join a [Research Bazaar](#) event in your area! These events are a great way to learn more about digital skills and tools required to do their research better, faster and smarter.
- Sign up for one of the Carpentry workshops to increase your software and/or data skills: [Data Carpentry](#), [Software Carpentry](#), or [Library Carpentry](#). Data Carpentry has a specific workshop for [Ecology](#)!
- Find an Online Open Course, such as:
  - [Data Science](#) on Coursera.
  - [Data Science Essentials](#) on edX.
  - [Data Fundamentals](#) on School of Data.

Keep in touch

There are several ways to stay informed of what is happening:

- Sign up to the [ANDS-Nectar-RDS newsletter](#).
- Get social on Twitter and connect with [ecoEd](#), [ANDS-Nectar-RDS](#), [Research Data Alliance](#).
- Many universities organize regular Hacky Hours: an informal catch up aimed at improving a researcher’s understanding and implementation of the technical aspects of research. Google ‘hacky hour’ and your university name to find out more.

If you have questions about this module, or want to discuss opportunities for workshops at your organization or collaborations, we would love to hear from you! Feel free to [contact us](#) through the ecoEd website!