

# Metadata Matters

*This guide provides a basic overview of metadata and outlines how the Landscapes and Policy Hub will meet its contractual obligations for ensuring research outputs are discoverable.*

## Data Management Plans

As part of our contractual obligations, we must have a Data Management Plan that secures the research outputs in perpetuity and provides open access. The overarching aim is that our research and outputs are discoverable. Research outputs include journal papers, grey literature, datasets, models and tools.

The traditional research outputs (the journal papers) are reasonably discoverable and we are publishing grey literature on our website. But datasets, models and tools aren't necessarily easily 'discoverable'.

The best way to ensure our research outputs are discoverable, is to create **metadata statements for datasets, models and tools**. Preferably, we would also publish the datasets, in the same way that research is published in journal papers, with a DOI (digital object identifier). Publishing models and tools isn't as straight forward or clear – but a metadata statement is a great start to becoming discoverable (and hence more cited).

Regardless of whether an output is published, our contractual obligation is to publish **metadata records** for all our datasets, ensuring our research is **discoverable** and **accessible** at a minimum, and **usable** where appropriate.

## Metadata: what is it?

Metadata is structured information that describes items. It can be used to describe physical items (for example, photographs, scientific samples), as well as digital items (for example, files, documents, images, datasets, databases). Metadata can describe individual items (for example, an individual image), or aggregations of similar items (for example, a collection of images).

In essence, metadata answers the 'who, what, when, where, why and how' about every facet of the documented item.

## Metadata: why create it?

Metadata enables research outputs to be discovered, connected with their owners and creators, linked to other related data or publications, contextualised in time and space, and to have the quality of the item assessed and research results validated. Metadata should also allow a user to acquire enough information about an item to use it without contacting the owner.

We only need to create metadata records for things we create. If you have a dataset from another source and clipped it to the study area for analysis, then no metadata is written, but the original must be acknowledged in any publication. But, if you create a dataset by merging two or more existing sets (eg New Alps Veg layer), then you need to write metadata for the new dataset, with full acknowledgements.

## Metadata standards and profiles

In many disciplines there are existing standards specifically designed for describing and sharing information. These standards form the basis of metadata 'schemas' or profiles.

As the Landscapes and Policy Hub has a significant amount of spatial datasets (especially through MCAS-S); we are using the ANZLIC Metadata Profile (Australia New Zealand Land Information Council). The ANZLIC Metadata Profile adopts established standards for the description of spatial data.

## ANZLIC Metadata Profile

The ANZLIC Metadata Profile is commonly used when geospatial location is an important feature of the material being described. As well as being ISO Compliant, the ANZLIC Metadata Profile is harvestable by the [Australian National Data Service](#) (ANDS). Being harvestable by the ANDS is a contractual requirement. The ANDS publishes a summary of metadata records (see page 3).



# Metadata Matters

## Metadata: elements

Metadata profiles are made up of individual fields or 'elements'. Not every research output will have or need every type of metadata element. The goal should be to provide as much information as possible, to facilitate future discovery and use of the material.

If you deposit your outputs in an institutional or domain-specific repository, you would probably be asked to fill out an online form or series of metadata fields or elements that describe your item or collection.

## Hub Metadata Information

Because you are a researcher with the Landscapes and Policy Hub, we want to make the process of making metadata records as painless and easy as possible.

We are negotiating with repositories to arrange batch uploading of the hub's metadata records, rather than single on-line entries created by individual researchers, which will take a lot of your time. We also know that there are many fields that are the same, and fields that are common across all our hub's datasets.

Therefore, we've created a spreadsheet to collect the information that we will use to create the records for you. The spreadsheet contains all the ANZLIC compliant elements and we have pre-filled the cells we know are consistent across the hub (funder statements, licences, distributor).

We are investigating the most cost effective and efficient way to publish the records, and publishing may not happen until early 2015. Therefore, we are collecting all the important information from researchers NOW to create the records later.

**Lauren Carter**, from the Bioregional Futures Team and **Suzie Gaynor** (Communications Manager) are coordinating the collection of the information and will arrange the generating of the compliant metadata record and publishing for you.

## Hints for writing good metadata

To create quality metadata, you need to include all the descriptive information necessary to locate, understand and use a dataset.

**Create a meaningful title:** Your title should provide meaning and convey as much information as possible using the least number of words. It should summarise important features of your data — *who, what, where, when and scale*.

**Abstract:** In about 500 words, your abstract needs to contain the best summary of everything, including important links, URLs and your dataset DOI if you have one. Sometimes people don't read any more than this and machines harvesting your record won't harvest all the fields, but they usually harvest the abstract.

**Data quality and purpose:** Establish the relevance and potential uses of your dataset. State any usage limitations and define fit-for-purpose — for example, 'not for navigational purposes'.

**Select descriptive keywords:** Convey the subject matter of your data. Choose both 'topic' keywords and 'place' keywords. For the hub's metadata records, we follow the ANZLIC theme keywords and qualifiers. Dropdown menus are included in our excel spreadsheet to guide you.

## Main Categories of metadata

Metadata is commonly arranged into broad, purpose-based categories.

**Descriptive metadata:** basic information used to find, identify and understand a dataset — title, author, abstract, subject keywords, resource type.

**Provenance (Lineage) metadata:** information about the origin of the data — Where did the data come from? Why was it collected? Who collected it, when and where? What instruments or technologies were used to collect the data, and how were they set up? What has been done to the data since it was collected?

**Rights and access metadata:** information about access and usage rules — Who is allowed to view, edit, or modify the data or metadata, and under what conditions? Who has authority over the data? Under what licence is the data available?



# Metadata Matters

## Harvesting Metadata Records

The Australian National Data Service (ANDS) harvests metadata records from many data repositories throughout Australia. Harvesting is an automated, regular process of collecting metadata descriptions from different sources to create useful aggregations of metadata and related services. The ANDS displays a summary of your metadata record that contains important links to the original ‘point of truth’ record. It only displays selected fields. An ANDS summary is likely to look like this:

The screenshot shows the Research Data Australia (RDA) website interface. At the top, there is a navigation bar with 'Research Data Australia' and 'ands' logos, and a search bar. The main content area displays a metadata record for a collection titled 'Climate Futures for Tasmania: runoff output from hydroclimatological model simulations'. The record includes a detailed description of the data, which are daily hydroclimatological simulations for the period 1961-2100. It also lists 'How to Cite this Collection', 'Identifiers' (Local ID: 01.tpac.org.au/ds/247), 'Related Publications', 'Related Websites', 'Spatial Coverage', 'Temporal Coverage' (From 1963-12-31 23:00 to 2101), 'Subjects' (Surfacewater Hydrology, EARTH SCIENCES, PHYSICAL GEOGRAPHY AND ENVIRONMENTAL GEOSCIENCE), and 'User Contributed Tags' (Oceans). On the right side, there are sections for 'Access' (URL: http://dl.tpac.org.au/pacpor...), 'Access rights', 'Connections', 'People' (James Bennett), 'Organisations & Groups' (Antarctic Climate and Ecosystems Cooperative Research Centre), and 'Services' (Tasmanian Partnership for Advanced Computing Oceans and Climate Digital Library Portal). At the bottom of the page, there is a footer with contact information and a 'Privacy Policy' link.

<http://researchdata.ands.org.au/climate-futures-for-tasmania-runoff-output-from-hydroclimatological-model-simulations/15215>

# Metadata Matters

## What the researcher needs to do

Each hub researcher will be asked to fill in the spreadsheet, to cover all the datasets, models and tools that they have generated in their research: one column per dataset.

Using the spreadsheet approach means that when you are entering the information, if you have two or more datasets almost identical, copying and pasting the column to create the new record is quick and easy.

For example, the metadata records for a suite of species distribution modelling outputs might possibly be very similar, with only the species name likely to be different for each record.

Remember, you only need to create a metadata record for something you have created. This includes if you have combined other datasets to create a new one.

As you fill in the spreadsheet – think about what information you would want to see to enable you to use someone else’s data.

## Publishing the Metadata Records

Using the completed excel spreadsheets, we will create compliant metadata records for each dataset and arrange for publishing in a recognised repository that is harvested by ANDS.

We are likely to have several options for publishing the metadata record, including ANU Data Commons, the Terrestrial Ecosystem Research Network (TERN) and Vivo at UTAS — at this point in time, where we publish the metadata record isn’t as important as making sure that we do publish metadata records.

## Options for Publishing Datasets

Hub researchers are responsible for the correct storage of datasets they generate as part of their hub research. The Australian National Data Service (ANDS) policy and code for responsible conduct of research expects that **datasets are stored securely in a durable, indexed and retrievable form.**

Our contractual obligation is that datasets are deposited with *an appropriate subject and/or institutional repository*, unless there are ethical, confidential or commercial reasons not to make the dataset public.

One of the benefits of publishing your dataset is that you receive a unique digital object identifier (DOI), which can be cited in papers, thus increasing your citation rates. Also, publishing your dataset in a recognised repository also takes care of your responsibilities to secure your datasets.

Note, metadata records don’t need to be published in the same place as the published dataset, and not all datasets with a published metadata record need be published.

Options available to Landscapes and Policy Hub researchers to store and publish their datasets include (but not limited to):

[ANU Data Commons](#) — This is where our MCAS-S input layers will be stored.

[TERN Discovery Data Portal](#) — The Terrestrial Ecosystems Research Network

[TerraNova](#) — The Australian Climate Change Adaptation Information Hub

[TPAC Digital Library Portal](#) — The Tasmanian Partnership of Advanced Computing

## Acknowledgements

This metadata guide was created specifically for the metadata collection process of the NERP Landscapes and Policy Hub. The information included was developed from information provided under open licences by other organisations, including:

[Best practices for writing metadata](#), courtesy of the US Geological Survey.

Isenor A, Bermudez L & Watson S (2010) [Writing Good Metadata](#). In *The MMI Guides: Navigating the World of Marine Metadata*.

Australian National Data Service. [Metadata \(awareness level\)](#) and [Metadata \(working level\)](#).

Griffith University (2014) [Creating Metadata for the NRM Climate Change Adaptation Information Management Support Project](#).