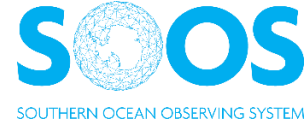


A BRIEF GUIDE TO PUBLISHING DATA FOR THE POLAR RESEARCH COMMUNITY



OBJECTIVE

This guide aims to provide you with tips to follow when publishing your data. Following these will help ensure your data is FAIR (findable, accessible, interoperable, and reusable). This means others can find, access, and reuse them and can combine them with other relevant data. It will also help ensure that you get appropriate citation/attribution when your published data are reused and increase the impact of your data.

KEY CONCEPTS

Metadata is the information used to describe your data. Many distinct types of metadata have been defined, but there are two important types. Firstly, metadata that helps others find your data. This might be called 'Discovery metadata' and includes things like the title, author, and keywords. Secondly, metadata that helps other use your data. This might be called 'Usage metadata' or 'Contextual metadata' and helps others access and understand your data. This metadata may be captured in a catalogue record, in a file associated with the data or within the data itself.

HOW TO MAKE YOUR DATA FINDABLE

1. PUBLISH YOUR DATA IN A RECOGNISED DATA REPOSITORY.

To make sure that your data will be findable, accessible, interoperable, and reusable (FAIR), we recommend using a recognised and trusted repository that focuses on your specific field of research. These specialised repositories are likely to offer better advice and support than general-purpose ones that rely on automatic submissions. For polar research, many nations have national polar/Arctic/Antarctic data centres you can approach.

Finding a repository:

- The Standing Committee for Antarctic Data Management provides a list of [National Antarctic Data Centres](#).
- You can search the [Registry for Research Data Repositories](#).
- The National Oceanic and Atmospheric Administration provide [a list of Data Centres and Institutions holding Arctic Datasets](#).
- You can also use the list of institutions in the [Mapping the Polar Data Ecosystem](#) tool.
- Publishers may also provide a list of preferred repositories such as the [list provided by Scientific Data](#).

2. DESCRIBE YOUR DATA TO MAKE IT EASY FOR SOMEONE TO FIND IT AND ASSESS ITS RELEVANCE

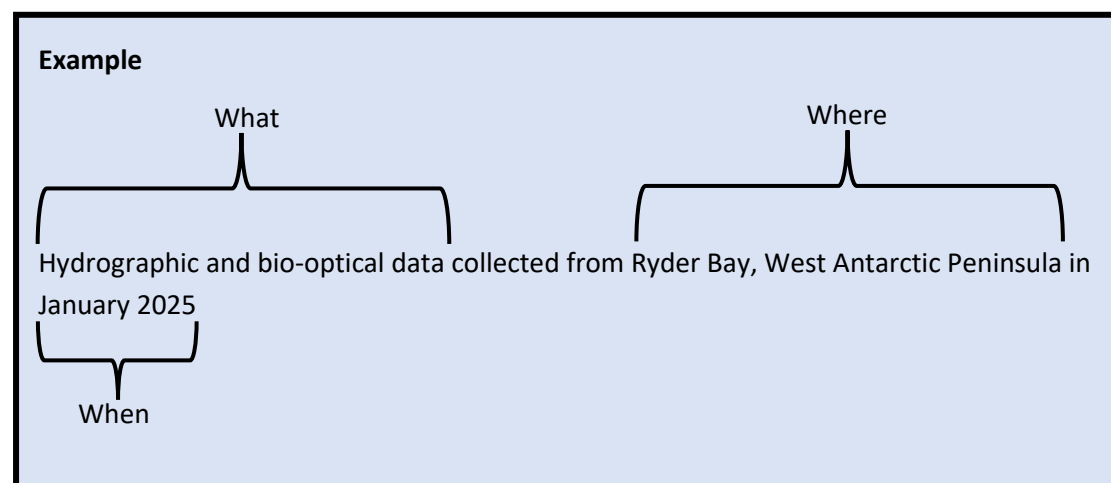
Discovery metadata refers to metadata elements that are generally indexed and searched in data catalogues. This typically includes fields such as: the title, abstract, authors (or creators), keywords, but may also include temporal and spatial information.

Tip 1: Ensure these elements include the terms that potential users might use when searching for data such as yours.

Tip 2: Structure the title and abstract information to make it easy for someone to assess whether the data it describes will be relevant to them.

Title

- Keep it brief and simple
- Follow the 'What, Where, When' structure
- Avoid use of acronyms unless they are also written in full
- Describe the data and not the project or activity that produced them



Abstract

- Provide a clear summary of the dataset in plain language for a non-expert user
- Use short concise sentences and consider using the 'what, where, when, how, why, who' framework to construct a comprehensive abstract:
 - ✓ **What** does the data contain
 - ✓ **Where** were the data collected
 - ✓ **When** were the data collected
 - ✓ **How** were the data collected
 - ✓ **Why** were the data collected
 - ✓ **Who** was involved (in the collection, management, publishing, and funding)
- Avoid use of acronyms unless they are also written in full
- Describe the data and not the project or activity that produced them

- Treat this as a stand-alone paragraph, where others should be able to determine exactly what was accomplished from this paragraph alone.

Example

This data release includes sensor data collected from Conductivity Temperature Depth (CTD) casts **[WHAT]** during a field campaign to Ryder Bay **[WHERE]** in January 2025 **[WHEN]** as part of the NERC funded SiCLING project (Silicon Cycling in Glaciated Environments), led by Dr K. Hendry from the British Antarctic Survey **[WHO]**. Four CTD casts **[HOW]** were carried out within the Bay. This research aims to investigate the release of biologically important nutrients and trace elements from glacial sediments in polar coastal environments, and the biogeochemical impact of ongoing glacier retreat **[WHY]**.

Funding:

NERC Pushing the Frontiers grant SiCLING (NE/X014819/1) **[WHO – funder]**.

Authors

- These may be called data creators, contributors, researchers, or Principal Investigators
- Provide a list of the people involved in the production of the data. This should include data collectors, processors, analysts, and anyone else that contributed to the production of these data.
- Think about who should be credited if the data are reused and cited
- Include permanent identifiers for people if you have them such as ORCIDs ([Open Researcher and Contributor ID](#))

Keywords

Keywords enable you to provide other terms that users might use if they were searching for your dataset that do not appear in titles or abstracts. They can be free text but often are terms taken from controlled vocabularies. Controlled vocabularies provide a consistent way of describing your data and help users find data that are all about the similar things.

Your chosen repository will be able to advise on keywords and use of controlled vocabularies.

Here are controlled vocabularies that you might come across:

- [Global Change Master Directory \(GCMD\) Earth Science and Earth Science Services Keywords](#)
- [ASFA Thesaurus](#)
- [INSPIRE Keywords](#) (these cover themes and topics)
- The [ICES Platform Codes](#)
- [SeaVox Platform Categories](#)
- Gazetteers for [marine regions](#) or place names such as the [Composite Gazetteer of Antarctica](#)

Time and Place

Including information in your metadata about where and when your data relate to is useful for catalogues that enable users to search using these parameters. This makes it easier for people to find data from a certain location or time. How you provide these may depend on the repository you are submitting to but standards exist for how to describe both time ([ISO 8601](#)) and geographic information and services (ISO 19115).

HOW TO MAKE YOUR DATA EASY FOR OTHERS TO USE

1. INCLUDE A LINK TO THE DATA FROM THE METADATA.

If a direct link to a resource is available, a web address (URL) that links to the resource should be provided. The URL should be accompanied by a title and function of the online resource, i.e. what the URL represents and what function it performs (e.g. download data, information about data, search for data). If a direct link cannot be provided, then a link to a contact point for requesting more information about the resource should be provided.

2. INCLUDE A LICENCE

A licence lets users understand the terms of use of the data. This will include what they can do with the data, whether it can be passed on to others and often stipulate that you have to acknowledge/credit the source. More information about licences and options can be found on the '[How to FAIR](#)' website. For open data, a Creative Commons licence such as CC-BY is often used.

3. INCLUDE A CITATION/ATTRIBUTION STATEMENT

This allows anyone who reuses your data to know how to reference use of your data. It also gives you a way of referencing the data in your publications. If your repository provides a DOI for your dataset then this should appear in the citation.

Your chosen repository is likely to provide you with a citation. If, however, you need to create your own then it should include creator(s) or contributor(s), date of publication, title of dataset, publisher, identifier (e.g. DOI) or URL of source, version (if appropriate), date accessed (if appropriate).

Example

Hendry, K.R. (2025). Hydrographic and bio-optical data collected from Ryder Bay, West Antarctic Peninsula, in January 2025 (Version 1.0) [Data set]. NERC EDS UK Polar Data Centre. <https://doi.org/10.5285/737cddb6-f3ee-4983-a117-b9d99d6e7f56>

4. PROVIDE THE DATA IN AN OPEN FORMAT

An open format is one that does not depend on a particular piece of software to open it. Ideally use a format which is machine-readable and self-describing such as netCDF ([Unidata](#) | [NetCDF](#)) for gridded data.

Formats likely to be accessible in the future are:

- Non-proprietary
- Open, with documented standards
- In common usage by the research community
- Using standard character encodings (i.e., ASCII, UTF-8)
- Uncompressed (space permitting)

Examples of suitable format choices:

- Image: JPEG, JPG-2000, PNG, TIFF, geoTIFF
- Tabular data: CSV, (TXT), XML
- Gridded data: NetCDF, HDF5, GRIB
- Audio: AIFF, WAV

5. ENSURE OTHERS WILL BE ABLE TO UNDERSTAND YOUR DATA AND ITS LIMITATIONS

It is important to provide rich usage metadata to ensure that potential users can understand the methodology used to collect your data, any processing or analysis undertaken, the instrumentation used and any quality control procedures. Where this information goes, will depend on the repository chosen and the metadata fields available. If you cannot include the information in the catalogue record, then add a README text file which contains the relevant information.

6. CHECK YOUR DATA HEADINGS

Make sure that the headers that you have used for tabular or gridded data clearly describe what the data are and include all relevant [units of measure](#). For spatial data, state which coordinate system was used. Where possible, use defined or standard conventions for the variables in your data (e.g. [CF Conventions](#) or [NERC Vocabulary Server parameter codes](#)). This enables your data to be combined with other data that measure the same variables.

FURTHER ADVICE AND TRAINING

Further advice and training are available from many sources. National polar data repositories (see choosing a repository section) can provide this. There are also a variety of online training course covering all aspects of research data management.

ABOUT THIS GUIDE

This guide was put together by the Polar Data Community including members of the Standing Committee for Antarctic Data Management (SCADM), the Arctic Data Committee (ADC) and the South Ocean Observing System (SOOS) Data Management Sub-Committee.

USER CHECKLIST

Use this before publishing your dataset to check it meets FAIR principles

✓ Repository Selection

- ☐ Deposit your data in a **recognised, trusted repository**.
- ☐ Verify that the repository supports polar research data and offers metadata guidance.

✓ Discovery Metadata

- ☐ Provide a **clear, concise title** following the "What, Where, When" structure.
- ☐ Write a **plain-language abstract** using the "What, Where, When, How, Why, Who" framework.
- ☐ List all relevant **authors/contributors** (with ORCIDs if available).
- ☐ Add **keywords** from a controlled vocabulary where possible (e.g., GCMD, NVS, INSPIRE).
- ☐ Include **temporal (ISO 8601)** and **spatial information (ISO 19115 or repository standard)**.

✓ Usage/Contextual Metadata

- ☐ Describe **data collection methods**, processing steps, instruments used, and quality control applied.
- ☐ Include this information either in the metadata record or in a **README file**.

✓ Licensing & Accessibility

- ☐ Apply an appropriate **open data licence** (e.g., Creative Commons CC-BY).
- ☐ Ensure a **direct URL or DOI** is provided for the dataset.

✓ Citation

- ☐ Provide a recommended **citation statement**, including:
 - Creator(s)
 - Year
 - Title of the dataset
 - Publisher or repository
 - DOI or URL
 - Version (if applicable)
 - Access date (if applicable)

✓ Data Formats & Standards

- ☐ Use **open, non-proprietary file formats** (e.g., CSV, netCDF, PNG, TIFF).
- ☐ Ensure **column headings** are clear, include units, and follow standard conventions (e.g., CF Conventions).
- ☐ Specify the **coordinate system** for spatial data (e.g., WGS84).
- ☐ Avoid proprietary or compressed file formats unless necessary.

✓ Final Check

- ☐ Review your submission for clarity, completeness, and adherence to your repository's standards.