CSIRO Oceans & Atmosphere Research Data Management Procedure

Discussion Version 1.0 – 9 August 2018

This O&A Data Management Procedure has been informed by the CSIRO Research Data Management Principles, which are currently in development. The Principles have been used to update and refine the CSIRO O&A Draft Data Management Policy 2016 to a Procedure.

Principles → Procedures → Implementation

A supplementary Implementation Guide will provide the information needed to apply Data Management processes at CSIRO O&A.

Contents

The purpose of the Oceans and Atmosphere Research Data Management Procedure is to:

- Outline the procedure for Research Data Management at O&A when collecting, processing and archiving data
- Support the CSIRO-wide Research Data Management Principles with guidelines

- Point to specific advice and guidance covered in the accompanying Implementation Guide

Research Data Management at O&A

Roles and Responsibilities

- Research Data need to be managed throughout its lifecycle, for example, from creation to archive or deletion. Roles for research data management should be allocated and defined before data creation.
  The following roles have responsibilities for data management:

Project Leader

Responsible for the day to day operation of the project and will:

- Implement the project data management plan and keep it updated as the project evolves and
- Assign data management roles for the project.

Project Proponent

Creates the initial project proposal that is used to seek approval to pursue the project. This person will:

- Include a data management plan and data related milestones as part of the project proposal, using a CSIRO template.

Data Originator

Creates, collects or collates the data, for example in the field, a lab, or at a computer.

Data Custodian

Is assigned formal administrative responsibility for the appropriate management of a dataset. This person:

- Is responsible for ensuring the safety of data during the project in accordance with guidance in the Implementation Guide.
- Must transfer custodianship to another person if the custodian leaves the organisation. Custodianship may extend beyond the end of the project.

Definitions

Research Data
Data obtained to inform research. Data may be in the form of facts, observations, images, computer program results, recordings, measurements or experience; they may be numerical, descriptive, visual or tactile. Data may be raw, cleaned or processed; generated, acquired, or collated for the work and may be held in any format or media.
Open Data Publishing
To provide open data access, normally via on-line self-service or connection to appropriate data portals.
Dataset
A set of related data values that forms a coherent unit, for example as collected by a common method for a particular study.
Data Products
The output from the application of Intellectual Property to data. See the Implementation Guide for examples.

Data are a critical asset for CSIRO

Ownership and Intellectual Property Rights

- Open access to data is the default position of CSIRO O&A.

Have you considered?

Project Leader: *Have data management roles been assigned before data collection starts?*

Link to O&A data management plan template

Data Custodian: *If this person leaves the project/organisation, have they transferred data custodianship?*

>> add a link to CSIRO Data Principles here

- Unless agreed upon with other parties (e.g. non-CSIRO research partners) and approved by CSIRO Business Development and Commercial (BD&C), datasets collected by O&A projects belong to CSIRO.
- Ownership and IP of all data and data products should be explicit in contracts.
- Projects which involve third party data must explicitly describe licensing and ownership of those data in contracts.

BD&C contact list

Project Proponent: *Are ownership and IP specified in research work contracts?*

Data need to be managed throughout their life

You will need to consider:

Data Formats

- When archiving data, projects should use non-proprietary and standardised data formats, ensuring long term accessibility of data and avoiding loss of support for licensed data analysis products.

Data Custodian: *Are archived data stored in non-proprietary and standardised formats?*

AUSGoal - Australian Governments Open Access and Licensing Framework

Standards

- Interoperable data relies on the use of accepted domain standards for data formatting, vocabularies, metadata creation and data services
- O&A aims to participate in the development and use of international and domain data standards and conventions.
- Where data standards or conventions do not exist, projects will clearly document, and where suitable, publish the procedures used for data storage.

Data Originator/Custodian: *Do data formats, metadata formats and data services meet domain standards and conventions?*

Data Originator/Custodian: *Do custom data/metadata/service conventions meet international and community conventions?*

Metadata

- Metadata represent a definitive description of O&A's data assets (and associated IP and licensing). Metadata should be appropriate to the research discipline and in a recognised, standard metadata schema.
- Metadata will be created before the end of a project and ideally as early in the data collection process as possible.
- A metadata record will exist for each of the research datasets a project uses, collects or generates and will be visible from within and outside of CSIRO.
- O&A datasets stored in external repositories will have linked metadata records in the O&A metadata catalogue pointing to the external location of the data and ensuring attribution.
- The O&A Information and Data Centre (IDC) also has services to deliver metadata records to other domain-relevant portals.
- Third party datasets without accessible metadata records should have metadata records created, ideally by the third party data originator.

Data Custodian: *Have metadata records been created before the end of the project?*

Data Custodian: *Are accurate metadata visible inside and outside CSIRO?*

Data Custodian: *Do metadata link to externally stored data?*

Data Custodian: *Do metadata need to be delivered to other catalogues?*

O&A IDC contact details

Data Custodian: *Do externally sourced datasets also have metadata?*

Data Object Identifiers (DOIs) and other linked identifiers

- DOIs are used to uniquely and persistently identify a dataset.
- Other unique identifiers exist for voyages, surveys, projects, authors, etc.

Data Management Plan

- All project proposals are required to include a Data Management Plan which will be followed though Project Management milestones. It clarifies the steps you take to manage your data from the start to the end of the project and outlines data the project will create, collect and collate. A Data Management Plan will consider third party datasets, licenses, IP, attribution, data storage and publishing.
- Data Management Plans will inform project reporting and should be living documents, revisited as projects progress.
- At the earliest opportunity during a project consider the following points:

Project Proponent: *Was a thorough Data Management Plan with milestones included in the proposal?*

Define O2D and emphasise good connections with DMPs

Add link here to O&A DMP Template

Project Leader: *Has the Data Management Plan been kept up to date as the project evolves?*

Data Storage and Management during a project

- Data storage must comply with CSIRO requirements, even during data collection, meaning that they are stored in a durable location with appropriate backup and ideally on a centralised server or a file share managed and backed up by CSIRO IM&T.
- Data files should follow a documented naming convention or use a data registration system so that it is easy to differentiate between raw data and later data versions. It must be possible to link metadata and data files.
- Read/write access permissions for datasets stored on shared infrastructure should be appropriate to protect against accidental or malicious corruption.

Data Custodian: *Does storage comply with* (>> add a link here to) CSIRO data storage policy and regulatory requirements

Data Originator: *Do file names follow a convention that allows metadata/data linkages?*

Data Originator: *During the project are dataset read/write permissions appropriate?*

Data Storage and Management at the end of a project

- Before the end of a project, datasets must be appropriately described (in a metadata catalogue) and stored in an approved persistent repository so that they remain available for future use. Suitable repositories for data exist within and outside of CSIRO.
- When data are sent to an external repository, an archival copy should be retained within CSIRO to ensure the continuity and persistence of CSIRO datasets.
- Repositories which provide DOIs to allow the citation of data are recommended.
- If a dataset is moved, the metadata record for that dataset must be updated to show the new location and any additional access information.

Data Custodian: *At end of project are data stored in an approved repository and do they have metadata?*

Include link to list of approved repositories

Data Custodian: *Is an archival copy of data retained at CSIRO?*

Data Custodian: *Does the repository provide DOIs?*

Data Custodian: *If the data are moved are metadata updated?*

Retention and disposal

- Most O&A datasets should be retained indefinitely. Some exceptions might be raw data in inaccessible proprietary formats or some deprecated model run outputs.
- Generally, raw data should be retained for the life of the dataset. Any software or configuration information that is relevant to the dataset should be stored with the dataset if possible.
- Paper records, including those that have been scanned, should be stored according to the requirements of the CSIRO Records Management Service.

Project Leader/Data Originator/Data Custodian: *Are datasets that won't be retained indefinitely identified and dealt with in the Data Management Plan?*

Data Custodian: *Is there facility and a plan to store software with data if required?*

Data Custodian: *Have paper records been stored according to Records Services requirements?*

CSIRO's research data are open by default

Data licencing – choosing a licence

- CSIRO data are open by default and CSIRO complies with the Australian Government Open Data Policy.
- Release of a research dataset and the choice of a licence are the responsibility of the nominated Data Custodian. The default licence for O&A datasets is the Creative Commons Attribution licence (known as CC-BY).
- The least restrictive licence appropriate to the material should be used.
- The licence should be clearly identified in any metadata records associated with the dataset.
- Licences should be chosen carefully as they will apply for the life of the dataset. They cannot easily be modified or retracted.

Governemnt Open Data Policy >> link to it

Data Custodian: *Have you chosen and applied the least restrictive data licence appropriate to the material?*

Creative Commons Licences

Data Custodian: *Is the licence identified clearly in metadata?*

Publishing and data citation

- O&A datasets should be published before the end of the associated project unless there are legal, contractual, ethical or privacy requirements that prevent this. Assigning DOIs is recommended where possible.
- Publications which reference the data should cite published datasets, ensuring appropriate attribution.
- Metadata must be lodged even for confidential or restricted data.

Project Leader: *If there are restrictions on data publication are these addressed in the Data Management Plan?*

Data Custodian: *Are data published electronically with web services, citation guidelines and DOIs?*

Be aware that there are obligations that come with data

- Sometimes there are legal, contractual, ethical or privacy reasons for not publishing data openly with a CC-BY licence.
- While planning a project, O&A Project Proponents should be familiar with CSIRO's position on the ethical requirements of using data, particularly privacy considerations where the data relate to individuals.
- Any sensitive data must be noted in the Data Management Plan.
- Contracts may include data access restrictions that are based on protecting privacy. Project Leaders and Data Custodians will ensure that all contractual and legal requirements pertaining to datasets are met.
- If in doubt about the data or dataset's actual or potential commercial value to CSIRO, the Data Custodian can consult BD&C for advice

Data Custodian: *Is CSIRO free to publish data without restriction from third parties or data co-owners? Has CSIRO considered possible risks (commercial, legal etc.) accompanying data release? Have all contractual/legal requirements been met?*

Project Proponent: *Have you considered CSIRO's ethical requirements and noted sensitive data in the Data Management Plan?*

The *CSIRO Ethical Conduct in Human Research Procedure* tells us how to comply with national legislation.

The *CSIRO Animal Welfare Policy* covers care and use of animals for scientific purposes.

The *CSIRO Privacy policy* covers management and protection of personal information at CSIRO.

Data Management Procedure Implementation Guide

The O&A Data Management Procedure Implementation Guide covers information about:

- *Creating and using IP and data ownership statements*
- *How to create and maintain an evolving Data Management Plan*
- *Data file naming and versioning*
- *Approved persistent data repositories*
- *Data formats, self-describing data files and centralised infrastructure for databases*
- *Creating and linking metadata records at CSIRO O&A*
- *Dataset retention and disposal*
- *Creative Commons licences and alternative licences available where required*
- *Publishing data including confidential data*

If you have any questions, you can talk to:

- *The Information and Data Centre*
- *The O&A Data Champion, Tara Martin*

Revision Schedule

Version	Date released	Revision Due
Discussion version 1.0	10th August 2018	September 2018