Climate Systems

National Environmental Science Program

Data Publishing Guidelines

By Brad Murphy, Data Wrangler

June 2026

This project is supported with funding from the Australian Government under the National Environmental Science Program. The Climate Systems Hub is a partnership of CSIRO, Bureau of Meteorology, Australian National University, Monash University, University of Melbourne, University of New South Wales and University of Tasmania and the Cross-jurisdictional Community of Practice for Climate Science. For more information visit www.nesp2climate.com.au.

Better metadata and persistent identifiers to improve Climate Systems Hub research accessibility (DOI:10.3565/espd-q195), a co-investment partnership with the Australian Research Data Commons (ARDC) through the Domain Data Portals program within the ARDC Planet Research Data Commons.

The ARDC is enabled by the Australian Government's National Collaborative Research Infrastructure Strategy (NCRIS).

Please cite as: Murphy, BF, (2026) *NESP Climate Systems Hub Data Publishing Guidelines,* Climate Systems Hub. National Environment Science Program, Australia. For further information about this report contact Brad Murphy, Bureau of Meteorology, brad.murphy@bom.gov.au

Acknowledgements

The hub would like to acknowledge the contributions of Julia Martin (ARDC), Aline Andrade (NESP Resilient Landscapes Hub), Emma Flukes (Marine and Coastal Hub), Chloe Mackallah (CSIRO), and Claire Trenham (CSIRO).**The Climate Systems Hub acknowledges the Tradition Custodians of the land across Australia where this work occurred. We pay our respects to Elders past, presents and future and recognise the important role traditional knowledge plays in understanding Australia's climate.**

*All images are creative commons unless otherwise referenced

Contents

Executive Summary

These guidelines provide practical advice for researchers publishing data from their hub projects. It provides context on the importance of good practice in data management, with the ultimate aim of maximising the impact of project data. Making data FAIR – Findable, Accessible, Interoperable, Reusable – is the key. The guidance describes the process for researchers to follow and the repositories they should consider for publishing their data. The inclusion of Persistent Identifiers in metadata, particularly Research Activity Identifiers, embeds useful information on who produced the data, enabling better tracking of impact and ingest of metadata into more data catalogues for increased discoverability.

Foreword from the Data Wrangler

As the Climate Systems Hub Data Wrangler, my role is to help ensure that the data we produce doesn't just exist, it gets used. One of the most important things I can encourage researchers to do is engage early. The decisions you make at the start of your project, about where your data will live, how it will be structured, and how it will be described, have a direct impact on whether that data can be found, understood and trusted by others. A short conversation early in your project can save significant time later and ensure your data is set up for maximum impact from the outset.

Good data publishing is not just a compliance exercise, it is a core part of how we extend the reach and value of our research. When data are published well, using appropriate repositories, rich metadata and persistent identifiers, they become far more than a project output. They become a resource that others can discover, interpret confidently, reuse in new contexts, and cite in their own work. Well-published data are far more likely to be found, understood, trusted, reused and cited, directly amplifying the impact of Climate Systems Hub research.

The opportunity for us as a Hub is to treat data publishing as an integral part of our research lifecycle, not something that happens at the end. By making deliberate choices early, and by working together across projects, we can build a stronger, more connected body of work that supports decision-makers, informs future research, and ensures the long-term legacy of what we produce.

At various points in these Guidelines I suggest you contact me to discuss your data. You can contact me anytime at brad.murphy@bom.gov.au

Introduction

This guidance is for all researchers working within the NESP Climate Systems Hub who are involved in producing, managing, or publishing project data outputs.

The document is intended to support researchers in making practical, informed decisions about how and where to publish data so that datasets have maximum impact for decision-makers, stakeholders, and the wider research community.

This guidance complements the hub Data Management Strategy and project-level Data Management Plans. Refer to these for more details on NESP and hub policies around data management.

Key message: Well-published data are more likely to be found, understood, trusted, reused, and cited – directly increasing the impact of Climate Systems Hub research.

Well-published data are more likely to be found, understood, trusted, reused, and cited – directly increasing the impact of Climate Systems Hub research

The aim of good data management – impact!

Managing and publishing data well are not just about compliance with funder requirements. When done well, they:

- Increase **visibility and discoverability** of your work
- Enable **reuse and citation**, extending impact beyond the project
- Support **transparency and trust** in climate science
- Make data easier to integrate into **models, assessments, and decision-support tools**
- **Preserve datasets** beyond the life of the project and the hub

Poorly published data (for example, unclear descriptions, missing metadata, or unstable hosting) may technically be “open” but are often **unused**.

Ultimately, the aim for all data you generate and publish is for it to be used by research users, industry practitioners, and policy makers. To make this a reality, the data needs to be both useful (i.e. serve a purpose) and useable (accessible and provided in a way that users can apply in their processes).

Key concepts

Core principles: FAIR and CARE

FAIR data principles

NESP and the hub adopt the **FAIR data principles**, which state that data should be:

- **Findable** – users (and machines) can discover the data
- **Accessible** – data can be accessed via clear, persistent mechanisms
- **Interoperable** – data can be combined with other datasets and tools
- **Reusable** – data are well described, licensed, and usable in the future.

FAIR is not about a single technical standard; it is about making practical choices that help others **find, access, understand, and reuse** your data.

Key enablers of FAIR data include:

- Trusted repositories
- Persistent identifiers
- Rich, standardised metadata
- Open, non-proprietary formats where possible.

CARE principles for Indigenous data

Where data involve **Indigenous Cultural and Intellectual Property (ICIP)**, FAIR must be complemented by the **CARE principles** for Indigenous data governance:

- **Collective Benefit** – data use should support Indigenous communities
- **Authority to Control** – Traditional Owners retain authority over their data
- **Responsibility** – data must be used responsibly and respectfully
- **Ethics** – data use must align with cultural protocols and consent.

Projects involving ICIP must engage early with the hub's Indigenous Partnership processes and the Data Wrangler. Additional restrictions, licences, or access controls may apply.

Persistent identifiers (PIDs): connecting people, projects and data

What are persistent identifiers?

Persistent identifiers (PIDs) are **stable, machine-readable identifiers** that permanently identify entities such as datasets, people, organisations, and projects.

They allow systems to reliably link data outputs to:

- Projects
- Researchers
- Organisations
- Funders.

Key PIDs used by the Climate Systems Hub

Researchers are expected to use the following PIDs wherever possible:

- **DOI (Digital Object Identifier)** – for datasets and data collections
- **ORCID (Open Researcher and Contributor ID)** – for individual researchers
- **ROR (Research Organization Registry)** – for organizations (e.g. universities, CSIRO)
- **RAiD (Research Activity Identifier)** – for hub projects.

RAiDs and why they matter

RAiDs are designed to record and maintain consistent and up-to-date information about a project. They can be modified throughout a project's lifecycle if information such as deliverables or personnel change, and outputs from the project can be added (via DOIs for example).

A **RAiD** represents your project as a whole and links together:

- Project title and description
- People contributing to the project (ORCIDs)
- Organisations (RORs)
- Outputs (datasets, publications, software)
- Other RAiDs (cross-referencing).

The RAiD is a single location for storing all information about a project, and is designed to be incorporated into metadata so project information can be retrieved easily.

Including the **project RAiD in dataset metadata** allows data catalogues such as Research Data Australia to automatically harvest from repositories and to group and surface all outputs from a project, the hub, and NESP overall. Having metadata in multiple catalogues increases the chances of users finding it and then accessing and using your data.

RAiDs are being minted for all Climate Systems Hub projects and are listed in Appendix 2. Inclusion of the RAiD in metadata is mandatory for new data publications. The Data Wrangler is also working with researchers to add RAiDs to some existing metadata records for existing datasets. Details on how to include RAiDs is provided in section 5 How to publish data.

Metadata: making data understandable and reusable

What is metadata?

Metadata is **structured information that describes data**. It provides the context users need to understand what the data are, how they were created, and how they should (and should not) be used.

Good metadata are just as important as the data themselves.

Why metadata matters

High-quality metadata:

- Enables discovery through catalogues and search engines
- Reduces misinterpretation and misuse
- Supports interoperability across systems
- Allows metadata to be harvested into national catalogues (e.g. Research Data Australia)
- Increases the likelihood of citation and reuse.

Recommended metadata approach

- All datasets must have a metadata record
- Metadata should follow repository standards
- For spatial data, ISO 19115 / ANZLIC profiles may apply.

At a minimum, metadata should clearly describe:

- What the data are and why they were created
- Who created them (with ORCIDs)
- Which project they belong to (RAiD)
- When and where the data apply
- How the data were produced and processed
- Limitations, uncertainties, and appropriate use
- Licensing and access conditions.

The Data Wrangler can provide templates and guidance. A hub-standard data management plan template for datasets will gather all the metadata needed when publishing to a repository.

Data Catalogues

For your data to be discoverable, its metadata needs to be included in data catalogues, and the more catalogues it is found in the more likely it will be found. The best way to get your metadata into many data catalogues is to publish it in a repository from which it can easily harvested, preferably automatically.

Research Data Australia (RDA) is a national data discovery portal operated by Australian Research Data Commons (ARDC) that harvests metadata from supported repositories into its centralised platform. RDA collects metadata from many Australian research entities, and a key to it being able to harvest your metadata and link it to the hub is for you to include your project RAiD and other PIDs in your metadata.

The choice of recommended data repositories listed in Section 4 is strongly driven by those that enable inclusion of RAiDs, as this tags your metadata as belonging to the NESP Climate Systems Hub. RDA is developing dedicated pages for each NESP hub, which will showcase our research output. **So please use one of the recommended repositories** – it will ensure your data is more findable and increase its impact. In addition, those repositories have data managers who will make the publishing and maintenance of your data much easier for you.

Where to publish data

Choosing the right repository is one of the most important decisions for data impact.

You should consider:

- Who needs to use the data?
- How large are the datasets?
- Are data services (e.g. APIs, OPenDAP, THREDDS) required and available?
- Does the repository support good metadata and PIDs?
- Can metadata be harvested into Research Data Australia?
- Is the repository stable and trusted?

Recommended and approved repositories

The following repositories are **approved and recommended** for Climate Systems Hub data, depending on data type and use case:

- **NESP Climate Systems Hub NCI Data Collection** – for datasets requiring wide access and strong metadata. Preferred.
- **NCI project storage + catalogue records** – for large climate model outputs, with separate metadata records for discoverability
- **CSIRO Data Access Portal (DAP)** – for projects led by or involving CSIRO
- **IMAS Data Portal** – for ocean and marine datasets
- **Institutional repositories** – where appropriate and compliant with metadata and PID requirements
- **GitHub** – for code (not as the sole location for data)

Repositories to use with caution and after consultation

- **Zenodo and Figshare** are designed for smaller datasets, particularly those linked to publications, but the NCI Data Collection was specifically designed for this purpose, and:
  - These do not fully support the hub's PID objectives (especially RAiDs and harvesting into the Research Data Australia catalogue)
  - Metadata quality depends heavily on researcher input and needs to be maintained. **So more work for you!**
- **Use these only after discussion with the Data Wrangler.**

Please do not publish data without consulting the Data Wrangler, especially for first-time or complex publications.

How to publish data

Details on how to publish data in each repository are provided in Appendix 1, but a summary of the standard approach follows.

Step 1: Early discussion

Before publishing, discuss your dataset with the **Data Wrangler**. This should happen:

- During project design, and
- Well before project completion
- Add the dataset to your Data Management Plan in your project's Project Notes OneNote

Topics to cover include:

- Data type, size, and users
- Suitable repositories
- Metadata requirements
- Persistent identifiers
- Access constraints or sensitivities
- Update or versioning needs.

Step 2: Prepare the data

- Finalise datasets intended for publication
- Use open, non-proprietary formats where possible
- Ensure files are clearly named and versioned
- Prepare/include readme or user documentation if needed
- Complete a data management plan entry for the dataset. Copy the Template (Data or Code) to a new sheet and rename for your dataset, e.g. "CS5.13 Dataset 1"): Data catalog 2026.xlsx.

Step 3: Prepare metadata

- Complete the required metadata fields for the chosen repository. All information required should already be included in the data management plan entry
- Include:
- Project RAiD
- Researcher ORCIDs
- Organisation RORs
- DOI (if assigned by the repository)
- Clearly describe methods, limitations, and appropriate use.

Step 4: Deposit data in the repository

- Upload data and metadata
- Apply the appropriate licence (default is open, unless an exception applies)
- Request a DOI if not automatically assigned.

Step 5: Register and link

- Ensure a metadata record exists in the relevant catalogue
- Confirm metadata can be harvested into Research Data Australia
- Link datasets to related publications, code, and reports
- Add dataset DOI to project RAiD and dataset data management record.

Step 6: Maintain and update

- If data are updated, ensure versioning is clear
- Update metadata if data locations or access conditions change.

APPENDIX 1 – Instructions for individual repositories

For generic data publishing steps, see Section 5. Specific instructions for each recommended repository follow, focusing on how to include RAiDs.

NESP Climate Systems Hub Data Collection

The NCI data collection was created specifically for hub researchers to publish smaller datasets. The Data Wrangler can guide you through this process, and the NCI data team will also provide assistance. If you do not have a NCI account they can do much of it for you. Or you can create an account.

1. Register for NCI project nd20 and nd20_w
2. Email help@nci.org.au to tell them you'd like to add a new dataset to the Collection (cc'ing the Data Wrangler brad.murphy@bom.gov.au)
3. Copy your data to /g/data/nd20/admin once you have nd20_w membership
4. Complete the NCI metadata template that they will send you.
5. To include your project's RAiD, add it under the "Funder - Grant/funding identifier"
6. Work with NCI on data QC and working out appropriate data services, and final publication of the data
7. Work with NCI on a promoting your new dataset

NCI Project Storage

If you are using NCI for your scientific compute and/or development, you will be a member of one or more projects. For example, the hub NCI data collection is NCI project nd20. Data produced in a project can be shared with other members of the project, so if someone requests access they need to apply for membership of the project. A metadata record can be created and added to the hub data collection (see above) to increase discoverability of any datasets you wish to promote, noting that only NCI users can access data under projects.

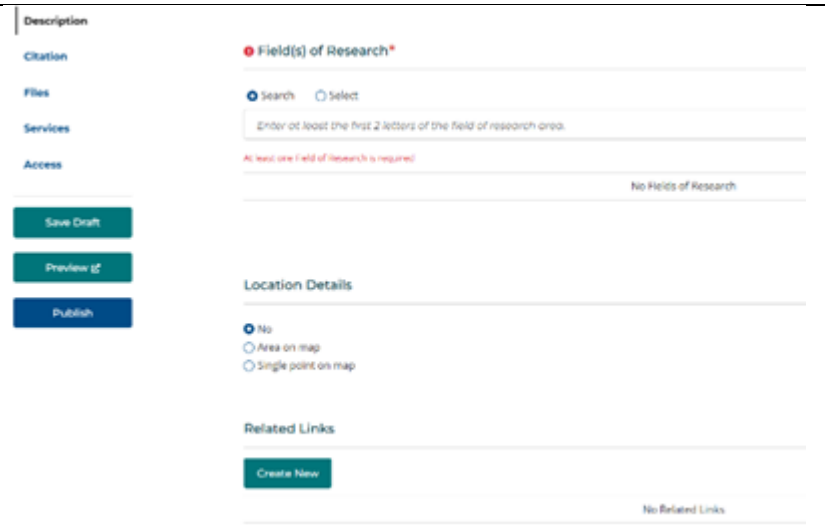
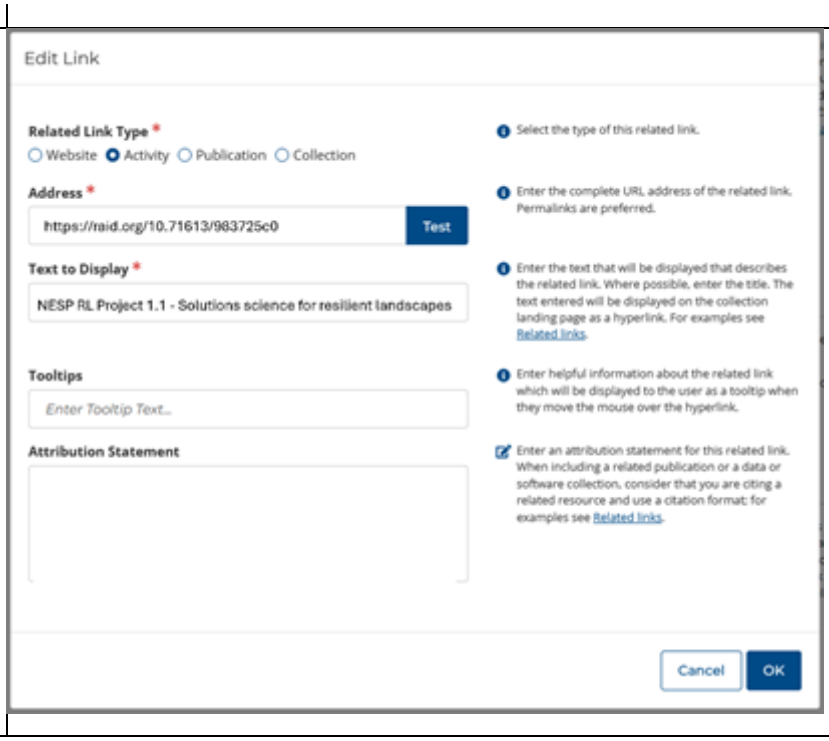
If data needs to be shared more broadly, especially to users who don't have a NCI account, the data can be added to the hub data collection (see above).

CSIRO Data Access Portal

Information about publishing in the portal is available in Collection Development Plan.

1. Follow instructions in the CSIRO Data Access Portal User Guide.
2. Your RAiD can be added as follows:

RAiD field	Section: Related links > Create New Values: - Address: [RAiD link] - Text to display: [RAiD Title]

RAiD instructions	
When uploading the dataset: - Select Description tab - Scroll down to Related Links section - Click the Create New button	
In the new box that appears: - In Related Link Type select Activity - In Address field insert the RAiD link for the project - In the Text to Display field, copy the title from the RAiD page or enter it using the following format: NESP RL Project <Project ID> – <Project Name>	

IMAS (Institute for Marine and Antarctic Studies) Data Portal

A comprehensive, easy to follow guide IMAS Data Submission Tool is available for publishing data on IMAS. Users need to register, but the tool is available for any researchers: https://data.imas.utas.edu.au/submit

RAiDs are added manually by the IMAS platform manager, so include your project RAiD when you submit your data via the tool.

University Repositories

Some universities will have their own data repositories for their staff to use. If a project is led from such a university then these repositories can be used:

University of Tasmania – has an instance of Figshare but is currently unavailable

University of Melbourne – has an instance of Figshare.

If you wish to use one of these repositories, the Data Wrangler can work with you to include the project RAiD in the metadata.

Github

If you are using github for your code development, it makes sense to share your code and data with others from github. This is a matter of making your code repository public, though you may wish to create a separate copy with only those files you want to make public.

Zenodo

This is an international data repository that is most often used for publishing relatively small datasets, particularly the open data linked to published research papers that needs to be made accessible. Metadata supports DOIs but not RAiDs. As it is hosted by CERN, it is expected Zenodo will exist well beyond the life of the hub and the useful life of most hub datasets. However, metadata from Zenodo cannot be easily harvested into other data catalogues like Research Data Australia, so discoverability will be limited. You also will not receive support from a data manager.

Unless there is a specific requirement to publish to Zenodo, the hub NCI data collection should be used instead. As always, discuss with your data wrangler.

Figshare

Like Zenodo, Figshare is designed as a data repository for small datasets and figures, particularly those related to research papers that need to be openly accessible. Metadata quality will directly depend on the level of detail provided by researchers when data are published. It supports DOIs, but not RAiDs. So metadata cannot be easily harvested into other data catalogues like Research Data Australia, and discoverability will be limited. Like Zenodo, Figshare is not recommended, and should only be used after discussion with your data wrangler.

APPENDIX 2 – Hub RAiDs

RAiDs have been created for all projects from Research Plan (RP) 2022. A current list can be found at https://static.prod.raid.org.au/raids/10.71790/

Other relevant RAiDs are:

- National Environmental Science Program Phase https://raid.org/raids/10.71821/23fcbc6f
- NESP Climate Systems Hub: https://raid.org/raids/10.71821/a945d761

Table 1: List of Climate Systems Hub project RAiDs

RAiD	Project name
https://raid.org/10.71790/883b2f55	CS2.1 - Enabling best practice adaptation
https://raid.org/10.71790/763d86ca	CS2.2 - Synthesis communication and data: Tailored information for stakeholders
https://raid.org/10.71790/af73740e	CS2.5 - Regional climate change guidance for local action
https://raid.org/10.71790/843e924e	CS2.6 - Extreme events explained
https://raid.org/10.71790/48b2d2bc	CS2.7 - Climate-effective management for threatened species and protected places
https://raid.org/10.71790/b058f9cc	CS2.8 - Extreme climate: dry, wet, hot-and-dry
https://raid.org/10.71790/11130ffd	CS2.9 - Transition to Net Zero
https://raid.org/10.71790/72de7bc0	CS2.10 - Oceans and coasts: connecting climate variability and extremes across scales
https://raid.org/10.71790/ef2fab6c	CS2.11 - Towards the next generation Earth System Model
https://raid.org/10.71790/05596784	CS4.1 - Adaptation planning approach for protected places
https://raid.org/10.71790/41dcc5e1	CS4.3 - Adapt Land & Sea
https://raid.org/10.71790/d20a98c5	CS5.1 - ACCESS development and delivery to CMIP7
https://raid.org/10.71790/fe86eab4	CS5.2 - Global carbon budgets and the role of terrestrial carbon sinks
https://raid.org/10.71790/2057ba78	CS5.3 - Uncharted climate futures
https://raid.org/10.71790/f0d5fb46	CS5.4 - Projection verification
https://raid.org/10.71790/534da9a2	CS5.5 - Marine heatwaves
https://raid.org/10.71790/27ff3254	CS5.6 - Extreme rainfall in compound events
https://raid.org/10.71790/7998ffee	CS5.7 - High-resolution rainfall extremes
https://raid.org/10.71790/741ed933	CS5.8 - Building for the future

https://raid.org/10.71790/655fe9ce	CS5.9 - Communicating coastal floods
https://raid.au/10.71790/93cbcc99	CS1.6 - The changing ocean's coastal and climate impacts
https://raid.au/10.71790/68b1cd17	CS1.8 - Australia's land and ocean carbon and acidification

APPENDIX 3 – Data Management Plan template

1	[Name of Data Product]
Description:	Describe the information that will be produced, including its type and characteristics, such as parameters, when available.
Project	Identifier and name of hub project, plus project RAiD (see Appendix 2)
Data Management Resources:	Describe the proposed approach to managing the data product
Data Custodian:	Who is responsible for managing this information, creating and maintaining metadata records, and ensuring data are appropriately stored?
Co-design process:	How was the utility of this information assessed with other users? How will stakeholders use it? What requirements do they have for accessing the data – format, repositories, services?
Licensing:	Provide the license under which the information will be made available?
Restrictions:	Identify any limitations on access or reuse (e.g., sensitive data, restricted data, software with license restrictions, etc.) and provide justification for restriction, including any period of exclusive use. Provide citation or documentation describing limitations if due to policies or legal reasons.
Persistent Identifier:	Provide a DOI or other identifier for the data
Metadata:	Identify the metadata record used to describe the data (URL where possible)
Quality Checks:	Identify the procedural steps for ensuring data quality during the project.
Data Processing & Scientific Workflows:	How was/will the data be generated (what model, observations systems, etc.) and how was/will it be processed. List other documentation such as papers, reports, etc. that describe how the data was generated
Format:	Identify the formats in which the data will be generated, maintained, and made available, and any associated tools/dependencies
Temporal and Geospatial information:	What period does the data cover? What is the time frequency? What spatial area, at what resolution, and on which projection?
Volume Estimate:	Estimate the volume of information generated: megabyte (MB), GB, TB, or PB.
Backup & Storage:	Describe the approach and locations used for backup and storage of the information during the project.
Repository for Data:	In addition to the NESP Climate Systems Hub register, identify any other repositories where you plan to share your data.
Data Services and Tools:	What services will be available for other users to access the data, e.g. NCI Data Services, OPeNDAP?
Legacy plan:	How will this information be stored and made accessible beyond the life of the NESP Climate Systems Hub?
Interpretation, applicability and limitations:	What information is available to guide users on how to make use of this data? For what types of applications should, and shouldn't, the data be used? What limitations and caveats about the data should users be aware of?
Citation:	How should users acknowledge and reference the dataset?

Climate Systems

National Environmental Science Program

For more information

www.nesp2climate.com.au

info@nesp2climate.com.au

Australian Government Bureau of Meteorology

CSIRO

Australian National University

MONASH University

THE UNIVERSITY OF MELBOURNE

UNSW SYDNEY

VICTORIA State Government

Government of South Australia

NSW GOVERNMENT

Queensland Government

GOVERNMENT OF WESTERN AUSTRALIA

Department of **Water and Environmental Regulation**

UNIVERSITY of TASMANIA