



National Environmental Science Program

# Data management strategy

National Environmental Science Program

Climate Systems Hub



© National Environmental Science Program Climate Systems Hub

Unless otherwise noted, Copyright (and any other intellectual property rights) in this publication is owned by the National Environmental Science Program Climate Systems Hub.

### Creative Commons licence

All material in this publication is licensed under a [Creative Commons Attribution 4.0 International Licence](https://creativecommons.org/licenses/by/4.0/) except content supplied by third parties, logos and the National Environmental Science Program Climate Systems Hub.

Inquiries about the licence and any use of this document should be emailed to [info@nesp2climate.com.au](mailto:info@nesp2climate.com.au)



### Cataloguing data

This publication (and any material sourced from it) should be attributed as: National Environmental Science Program Climate Systems Hub Data management strategy.

This publication is available at [nesp2climate.com.au](https://nesp2climate.com.au).

The National Environmental Science Program Climate Systems Hub has exercised due care and skill in preparing and compiling the information and data in this publication. Notwithstanding, the hub, its employees and advisers disclaim all liability, including liability for negligence and for any loss, damage, injury, expense or cost incurred by any person as a result of accessing, using or relying on any of the information or data in this publication to the maximum extent permitted by law.

### Acknowledgements

The National Environmental Science Program Climate Systems Hub are grateful to staff of the Australian Government Department of Agriculture, Water and the Environment, and members of the Hub Steering Committee for valuable input in preparing this document.

### Document control

Version	Date of issue	Author	Reason for change
0.1	28/04/2021	Brad Murphy	Initial draft
0.2	14/05/2021	Brad Murphy	Revised
0.3	14/05/2021	Simon Marsland	Revised and file name changed for consistency across Hub strategies. Persistent identifier section revised. Research output submission process revised. Open access outputs revised. Accessibility revised.
0.4	14/05/2021	Amelia Tandy	Draft indicative budget
1.0	14/05/2021	Simon Marsland	Submitted draft
1.1	27/08/2021	Brad Murphy	Applied DAWE feedback
1.2	08/10/2021	Simon Marsland	Final draft resubmission
1.3	9/11/2021	Brad Murphy	Applied DAWE feedback
2.0	15/11/2022	Brad Murphy	Minor updates for 2022
3.0	13/10/2023	Brad Murphy	Minor updates for 2023
3.1	01/03/2024	Brad Murphy	Updated with status update
4.0	14/02/2025	Brad Murphy	Updates for 2024/5, including status update

# Contents

<b>Acknowledgement of Country</b> .....	<b>5</b>
<b>Background</b> .....	<b>6</b>
<b>Strategy aims</b> .....	<b>7</b>
<b>Data and information management approach</b> .....	<b>8</b>
<b>Hub roles and responsibilities</b> .....	<b>10</b>
<b>Types of research products and data</b> .....	<b>12</b>
<b>Ownership and intellectual property rights</b> .....	<b>14</b>
<b>Data licencing</b> .....	<b>15</b>
<b>Data management and storage</b> .....	<b>16</b>
<b>Exceptions to the open data policy</b> .....	<b>21</b>
<b>Accessibility</b> .....	<b>22</b>
<b>Acknowledgment of NESP</b> .....	<b>23</b>
<b>Project-level data management</b> .....	<b>24</b>
<b>Status of strategy implementation – tools, artefacts, and collaborative projects</b> .....	<b>26</b>
<b>Risks</b> .....	<b>28</b>
<b>Related materials</b> .....	<b>29</b>



# Acknowledgement of Country

Climate Systems Hub recognises and acknowledges First Nations people across the length and breadth of Australia land and sea Country and their ongoing connection to Country.

The Climate Systems Hub pay respect and acknowledgement to Elders past and present, as knowledge holders and keepers.

The Climate Systems Hub pay respect and acknowledgement to senior community leaders of their Country and communities.

# Background

The National Environmental Science Program (NESP) is a long-term commitment by the Australian Government to environment and climate research. The program is committed to promoting open access to public sector and publicly funded information, including optimising the use and reuse of data. Hence, open access to government-funded information is the default position with exception only for sensitivity reasons or underlying licensing constraints. The program works collaboratively with the private and research sectors to extend the value of public data for the benefit of the Australian public.

Open data is, at its best, freely available based on the *FAIR data principles* of being findable, accessible, interoperable and reusable under licences that allow reuse. When working with Indigenous data these guidelines require the complementary use of the *Global Indigenous Data Alliance's CARE principles* for Indigenous data governance of collective benefit, authority to control, responsibility and ethics. This will increase the capacity of the program to support a more collaborative, informed approach to managing Australia's environment.

In accordance with the funding agreement, the Climate Systems Hub is required to create a data management strategy that is compliant with the [NESP Data and information guidelines](#). The strategy also needs to be supported by resources and infrastructure, such as:

- the hub data wrangler, who has a role to work with the hub researchers, the department and other stakeholders to translate data and information into relevant databases and outputs that align with these guidelines
- allocation of resources for data management, from the initial data capture through to ongoing delivery and curation
- information technology infrastructure: hardware, software and other facilities that underpin data-related activities
- support services: resources allocated to support implementation of metadata management so that data records can be used for both internal and external purposes.
- The Climate Systems Hub will approach resourcing to distribute data in accordance with the FAIR data principles through:
  - using existing infrastructure wherever possible, including well recognised standards, metadata catalogues and existing repositories known to enable findable, accessible, interoperable and reusable data
  - each research project will devote a proportion of its resources to data management, including developing a project-specific data management plan, appointing data custodians for each dataset, and engagement with the Data Wrangler to facilitate appropriate management of all data and research outputs
  - providing researchers with clear guidance, templates and ongoing support to ensure consistency and efficient use of resources towards data management.

The NESP Climate Systems Hub will require a variety of data management approaches to cover the wide range of data volumes (from kilobytes to petabytes) produced in the course of climate research. While striving to achieve the same aims as other hubs, some of these approaches may be unique to the Climate Systems Hub.

# Strategy aims

The overarching objective of the Climate Systems Hub data management strategy (DMS) is to facilitate the hub's research having the greatest possible impact on environmental decision-making in Australia. The strategy will outline the key principles and provide guidance to researchers on how to best deliver on that main objective.

A key to achieving this goal is through the *FAIR data principles* - findable, accessible, interoperable and reusable NESP research products. This document will provide guidance for how the hub and its researchers will achieve this. Having a well-designed DMS for the hub also assists researchers in meeting requirements of the hub's legal agreement with the department.

The strategy will outline a variety of data management approaches to applying the *FAIR data principles* to cover the variety of methods for generating data, and their widely varying magnitudes of volume, unique to climate systems research.

A key to delivering data that is accessible is making them available through open accessing licensing wherever possible, and this strategy will embed this principle. It will also ensure all *Indigenous Cultural and Intellectual Property (ICIP)* will be managed in consultation with Traditional Owners, guided by appropriate data governance principles such as those of the *Global Indigenous Data Alliance's CARE principles* of collective benefit, authority to control, responsibility, and ethics.

The strategy will also outline the responsibilities of the different roles involved in managing the data and research outputs of the Climate Systems Hub. The hub leader provides oversight and line of reporting to the department. Project leads are responsible for the management of data within their projects, and will appoint data custodians who will manage individual datasets in line with this strategy. The data wrangler will provide advice to project leads and researchers on how to meet the data management aims, as well as working with researchers and stakeholders to ensure data are integrated into relevant repositories, digital systems and decision-making tools.

This strategy will also provide the framework for the co-design process that will design how to best deliver data into those repositories, systems and tools to have the greatest possible impact. Specific templates will be created to help researchers manage their data, keep intellectual property (IP) registers up-to-date, and publish compliant metadata to achieve the *FAIR* goals for data.

# Data and information management approach

This document outlines the approach the hub and hub researchers will take to manage their data and other research outputs before, during, and after a research project. The data management strategy will be reviewed periodically throughout the life of the program, but at least annually in line with the annual reporting process, to capture emerging trends in data management, and changing or new stakeholder requirements, and as the specifics of data outputs become clearer through co-design and project-design processes.

This strategy will provide guidance on the objectives, key steps and resources to ensure that hub research outputs meet the requirements of the NESP program guidelines and the [NESP data and information guidelines](#).

The data management strategy is a living document that describes:

- who will be responsible for each data related activity
- data generation or collection, analysis and management practices used
- types of data and research products created
- who owns and accesses the hub's data and products
- metadata standards used
- products and data storage, security, privacy and unique identifiers
- product legacy planning and responsibilities for ensuring enduring data and products
- facilities and equipment required.

The fundamentals of the data management approach to be followed by the hub are as follows:

1. Research data and outputs will be well-documented according to accepted and trusted community standards. A key requirement for all research products generated by the hub research will be to follow metadata standards based on accepted best practice. Refer to [Metadata](#).
2. Data will be made publicly available whenever legally, ethically and contractually possible, under an open licence policy except in particular cases. A record of these exceptions will be kept, and exceptions reported periodically to the Department. Refer to [Data licencing](#).
3. Data and other research outputs will be stored in repositories that are accessible to end-users and other researchers, and stored in such a way so as to endure and remain FAIR beyond the life of the project. The choice of repository will be based on practicalities of the data, its likely use, and the expected longevity of the hosting infrastructure. Refer to [Data storage and sharing](#).
4. Each project will develop its own data management plan, with identified data custodians responsible for managing the data. A standard template will be developed for project use (drawing upon material from the [Australian Research Data Commons](#)). Refer to [Project-level data management](#) and existing data management policies and tools from Australian universities.
5. The nature of research outputs, their formats and repositories, will be developed through a co-design process involving stakeholders and end-users to meet their requirements and ensure data are fit-for-purpose, and will evolve through ongoing consultation. Refer to [Data storage and sharing](#) section in this document.



6. Research outputs and data will be accompanied by documentation providing guidance to users on how best to make use of the data. Refer to [Data storage and sharing](#) section in this document.

A key principle to data management in the Climate Systems Hub is ensuring data and other research outputs meet stakeholders' requirements for their decision-making, including the type of information, is available at the required (possibly varying) technical level, and it can be accessed by stakeholders in the formats and through the tools that they can use effectively. To meet this goal, the hub is carrying out extensive stakeholder engagement through an on-going co-design process, to determine requirements and to fine-tune these through the life of the program. Project leads and team members will work with the Knowledge Brokering team, along with the data wrangler, through this process, to gain a thorough understanding of stakeholder needs, and to communicate with them the potential of the hub's science outputs. Project Leads also need to determine the level of Indigenous partnership in the project by following the 3 Category approach – refer to the Indigenous Partnership Strategy.

# Hub roles and responsibilities

## Hub data wrangler role

The data wrangler's activities include working with the hub, researchers, the department, and other stakeholders to translate data and information into relevant databases and tools and to help integrate research outputs into national information repositories, digital systems and decision-support tools. This includes ensuring data is discoverable, accessible and useable, to optimise the use and reuse of public data. The data wrangler is responsible for developing and updating this data management strategy and ensuring its implementation by all researchers in the hub. They will provide advice and guidance on the practical implementation of the strategy to project leads and researchers. The Data Wrangler will also maintain the IP and ICIP registers, a metadata catalogue for all hub datasets. They will also keep a register of exceptions to the open data policy and communicate these exceptions to the hub leader, who will report them to the Steering Committee and the department.

In order to ensure research outputs are integrated into national information repositories, digital systems and decision-support tools, and to meet stakeholders' requirements, the data wrangler will be closely involved in the co-design process for the hub's projects. Through this co-design process, stakeholders will be extensively consulted to determine the outputs they require, the appropriate formats, delivery methods, documentation, repositories, and long-term storage of the data.

## Project lead role

Accountability for project data management lies with the Project Lead(s). The project lead may choose to delegate the project data management activities to a project Data Lead. The activities that must be carried out in conjunction with the hub data wrangler and in accordance with the hub-level data management strategy are:

- develop and maintain a project level data management plan to ensure stakeholder requirements are met for the research outputs (type of information, formats, delivery methods, documentation, repositories and long-term storage, etc.), and those requirements are documented, along with details on how they will be delivered
- complete entries in the IP (and ICIP if relevant) registers and keep these up-to-date
- discuss any potential need for exceptions to the open data policy for any project output with the data wrangler
- assign a data custodian for each dataset the project will develop. A project may have more than one data custodian, with whom the Project Lead or Data Lead will work to ensure project data are managed appropriately.

## Hub leader role

It is the hub leader's responsibility to provide oversight of data management activities for the hub, and report instances of exceptions to these guidelines to the Steering Committee and the department. Collated lists are generally at project level and include basic information on the data generated or collected and the justification for its non-release. The publication of research outputs must comply with the [NESP data and information guidelines](#).

## Data Lead role

If a Data Lead is assigned by the Project Lead, they will be given responsibility for delivering the data management activities of the project. Accountability lies with the Project Lead,

## Data custodian role

It is the data custodian's responsibility to appropriately manage one or more dataset(s) of a project. Each project may have several data custodians. This person is responsible for ensuring the safety of data during the project, creating metadata records and keeping them up-to-date, and ensuring data are appropriately stored. A data custodian retains responsibility for a dataset until that responsibility is

passed to someone else. Before the end of the project, the custodian, in consultation with the data wrangler, should choose a suitable long-term repository for the dataset, deposit it and make sure it is published (i.e. made publicly available) following FAIR principles wherever possible. The data custodian will work with the data wrangler to ensure stakeholder requirements are met for the research outputs (formats, delivery methods, documentation, repositories, and long-term storage, etc.).

# Types of research products and data

To a large extent, Climate Systems Hub research tends to generate new and derived datasets rather than collecting raw data through field observations, and a broad range of research products are expected to be generated throughout the life of the hub. A characteristic of climate research data is the very large volumes of data generated, particularly output from climate models.

While derived datasets, application ready data and end-user products will be specifically targeted at meeting stakeholder requirements, not all data generated will be of direct value to different stakeholders. Generation of raw data will be essential to the research process, and some will be required by other researchers, so means of sharing such data between researchers within the hub and with other hubs is essential. However, it may be neither practical nor beneficial to make all such raw data accessible via public repositories.

This strategy will therefore differentiate between different technical levels of data and information, with each level requiring unique strategies for managing and sharing that data and varying levels of accessibility. The levels of data and information and their characteristics are listed below:

1. End-user products such as research papers, communication products (brochures, reports), schematics, presentations; aimed at broad range of end users. These present information at a high level for wide dissemination and are small in size.
2. Application-ready datasets, ready for ingestion into next- or end-user models, such as derived indices, targeted to specific next- and end-users, are small to medium data volumes, and delivered via data repositories with data services.
3. Raw research data such as climate model code and output, used by climate researchers but of little utility to others; often base data post-processed to level 2 data, have very large data volumes (petabytes) that are not easily shareable.
4. Interim and externally sourced data including model test output, third-party reanalysis data for model evaluation; superseded by later output or IP owned by third party; medium to large data volumes.

The hub's data management strategy treats the management, storage and sharing of each level of data and information appropriately in order to achieve the driving objective of the research outputs having the greatest impact on environmental decision making in Australia. Details of the outputs to be produced by each project, and how they will be delivered in a way that meets stakeholder requirements, will be recorded in the project's data management plan. These details will be developed as part of the ongoing co-design process between the hub and its stakeholders (refer to [Data storage and sharing](#)).

The research products produced by the Climate Systems Hub may include the following:

1. New raw datasets including spatial data and model outputs (type 3)
2. Data analysis and derived data products such as geographic information system-derived maps (type 2 or 3)
3. Models and other tools such as software created by the research process, including value-added digital products derived from off-the-shelf/open-source software (such as Decision Support Tools) as well as new apps and APIs (type 2)
4. Analysis code developed to produce data products and journal publications where appropriate (type 3)
5. Publications including scientific papers, reviews, books, book chapters (type 1)
6. Grey literature including fact sheets, project profiles, manuals, and technical reports (type 1)
7. Images, maps, photos, videos, animations (type 1)
8. Models and other tools, such as decision support tools or software (type 1)

9. Websites, mobile or tablet apps (type 1)

10. Indigenous Cultural and Intellectual Property. This may fall into one of several categories. See Indigenous Cultural and Intellectual Property (ICIP)

11. Unspecified emerging technology.

# Ownership and intellectual property rights

Registers will be kept for recording all inputs to projects (a Background IP Register), all outputs (a Project (or Foreground) IP Register), and, where relevant, an Indigenous Cultural and Intellectual Property Register. Project leads will be responsible for keeping IP registers for their projects up-to-date. The management of the registers is the responsibility of the Data Wrangler, with oversight from the hub leader.

At the commencement of each project, its project lead will prepare a Background IP (BIP) Register to record the agreed inputs into the project. The IP register will clearly indicate the type of IP and/or data, its owner, its purpose in carrying out the project and any contractual constraints around its use. Where a project lead identifies an input as relating to ICIP, a separate ICIP register must be populated.

During the term of the project, each project lead must regularly update the BIP Register and add items that are provided to the project or remove items that are no longer considered relevant.

Similarly, a Project IP (Foreground IP) register will be kept by the project lead and agreed upon by each IP owner to record agreed outputs of each project. The register should note a common understanding between the parties involved as to whether that IP constitutes project IP or an Improvement to an owner's Background IP. Where a project lead identifies any project IP as embodying ICIP, a separate ICIP register must be populated.

## Indigenous Cultural and Intellectual Property (ICIP)

ICIP will be managed by the hub including using the *Global Indigenous Data Alliance's CARE* principles for Indigenous data governance: collective benefit, authority to control, responsibility and ethics. As indicated above, any ICIP identified as either input or output of a research project must be captured in the ICIP register. Ownership of ICIP will remain with Traditional Owners, and will only be used in accordance with permissions granted by Traditional Owners. Such permission may be rescinded at any time.

The hub's *Indigenous Partnership Strategy* should be referred to when ICIP is identified, and any engagement with Traditional Owners should include the hub Indigenous Partnership Team. The *Indigenous Partnerships Strategy* has as one of its pillars for successful Indigenous partnerships the principle of upholding Indigenous Cultural and Intellectual Property of Traditional Owner groups from co-design to co-authorship. This principle will be applied through activities that apply free, prior and informed consent processes from initial engagement with Traditional Owner group, and throughout the research to final product, develop proformas in the co-design process where Indigenous Cultural and Intellectual Property are highlighted, through appropriate cultural and community protocols, and ensure the hubs' research is informed by relevant traditional ecological knowledge of weather and climate, consistent with Indigenous Cultural and Intellectual Property.

In order to ensure the hubs' Indigenous engagement is conducted to the highest ethical standards across all activities, all hub Program Management staff responsible for managing ICIP, or engaging with Traditional Owners about ICIP, will undertake relevant cultural-capability training.

# Data licencing

The Climate Systems Hub data licencing procedure is guided by the spirit of ‘Open Access Data’, supporting national and international research efforts through the free exchange of data. This aims to increase the impact of NESP research, by supporting the free exchange of data collected or generated by the hub, while ensuring that the hub retains any intellectual property rights in the datasets and is otherwise able to adequately manage the data in accordance with the NESP guidelines.

The hub will make all research outputs publicly available under the latest Creative Commons framework, using the Creative Commons – Attribution – Non-Commercial 4.0 International Licence (CC-BY-NC 4.0), unless a more restrictive licence is required for a particular dataset due to privacy concerns or the like. Refer to Open-access outputs and Exceptions to the open data policy sections in this document. This should allow for research outputs to be deposited into an appropriate repository. For code and software, open-source licences such as General Public License (a relatively restrictive copyleft license), MIT (Massachusetts Institute of Technology, a highly permissive license) or the CSIRO licence may be more appropriate (see <https://opensource.org/licenses> for a full list of open-source initiative approved licenses).

Note that ICIP and information shared by Traditional Owners may include Indigenous knowledge which cannot be protected by copyright and hence a Creative Commons licence. It has been agreed that any publication including ICIP will be published under Creative Commons Non-Commercial - No Derivatives 4.0 International licence, meaning appropriate attribution must be made to such knowledge, it may not be used for commercial purposes, and if modified in any way may not be redistributed. Please contact the Jeff Jones or Amelia Tandy for additional information, including an appropriate ICIP attribution statement.

# Data management and storage

This section provides guidelines and requirements for specific aspects of data management for the Climate Systems Hub, particularly those that will aid projects in meeting the *FAIR data principles*. It provides information around metadata and persistent identifiers that should be created for all datasets to make data findable; how data should be stored and shared (to enable accessible data); data formats and standards to make datasets more interoperable; and information on preferred licensing and how to ensure data are available beyond the life of the hub (to ensure data are reusable).

## Metadata

Hub researchers will take all reasonable steps to add high-quality metadata to all outputs resulting from NESP funding. This provides the contextual information needed to enable the end-user to use the data appropriately and consistently. High-quality metadata can facilitate findability of data, allow for more successful data integration and increase data value.

Metadata is intended to represent a definitive holding of the hub's data assets. A metadata record must be created for every dataset that a project collects or generates. A metadata record should include details appropriate to the research discipline, incorporated into the standard metadata schema appropriate to the discipline.

For spatial metadata, hub researchers will conform to the Australian New Zealand Land Information Council (ANZLIC) standard metadata profile. [ISO19115-1](#) is the international standard that underpins the ANZLIC spatial metadata standard (see [Related materials page](#)).

The Climate Systems Hub encourages the contribution of metadata to the Australian Research Data Commons (ARDC) data discovery portal, [Research Data Australia](#). This portal harvests information from data repositories around Australia. The metadata for the associated data should contain a link to related publications and to NESP. Best-practice guidance for linking to publications and grants from data (or the data's metadata) will be sourced from ARDC.

The hub data wrangler will maintain a metadata catalogue for all hub datasets. Such metadata statements are to be lodged by data custodians with the data wrangler as soon as is reasonably possible once the data are created/utilised for project purposes, and no later than the completion of relevant project deliverables/milestones. A standard metadata template will be developed and provided to all data custodians for this purpose.

At a minimum, requirements for effective metadata are that it:

- describes the nature and purpose of the data along with contextual background information
- describes the methods used in the dataset creation and all data attributes, and highlights the quality or limitations of the dataset
- is available online, linked to the data, code or documents (also online), and accessible by web search engines or discovery facilities
- provides links to key reports and papers that provide additional context and data details
- identifies the attributes, methods and procedures used for determining the attributes of all values within the dataset/stream
- uses standard or community accepted taxonomies and/or defines or links to online definitions of all terms used in the data
- provides contacts and access locations for the data
- provides provenance for any data that is used in generating the research output.

## Persistent identifiers

Where possible a [persistent identifier](#) will be assigned to all hub data, and, where possible, research outputs.



The identifier for products may vary, for example it could be an International Standard Book Number (ISBN), International Standard Serial Number (ISSN), an internal identification number, a handle or a Digital Object Identifier (DOI).

Assigning a DOI to finalised data and other research outputs facilitates citation and will be considered best practice.

Online tools to create DOIs for products that are not associated with a repository such as the ARDC DOI Minting Service or Zenodo will be used.

## Data Formats and Standards

Projects should use non-proprietary data formats (see <http://www.ausgoal.gov.au/open-formats>). For example, comma separated value (CSV) format is preferred to an Excel spreadsheet for storage of table data. PDF/A is preferred to a Microsoft Word file for storing text.

Self-describing data formats such as netCDF and HDF are encouraged for spatio-temporal data as they encapsulate much of the metadata required to use a dataset. When data are stored in files, there should be a data registration system or documented naming convention (defined by the data custodian) so that it is easy to differentiate between raw data and the various versions of a dataset as processing evolves. These systems should permit the allocation of IDs to data to link metadata and data file.

Domain data standards or conventions should be used wherever they exist. Examples are the *netCDF Climate and Forecasting (CF)* convention and the *Gridded Binary or General Regularly-distributed Information in Binary form (GRIB)* format. Where data standards or conventions do not exist, projects should develop and document their own conventions for data storage, for example defining non-cryptic variable, field and file names and using consistent standard units throughout the project. These standards should meet international conventions. Where disciplinary metadata taxonomies exist, these should be used and links provided, and digital metadata records (such as in netCDF headers) should be both machine- and human-readable.

## Open-access outputs

The Climate Systems Hub will provide up-to-date, high-quality data and information to decision-makers, environmental managers, other scientists and the community.

Hub outputs should follow the *FAIR data principles* and be available at no charge, using best-practice approaches specific to the research output type, and according to information provided in the [NESP data and information guidelines](#) expectations for publications, such as datasets, technical reports, grey literature, images, photographs and videos, models and software, websites, and mobile and tablet applications.

In the case of scientific publications, the hub will make reasonable efforts to either publish via Creative Commons or make scientific publications publicly accessible within 12 months of publication.

In some cases, open access to information may not be suitable, such as where that information is culturally, environmentally, commercially or socially sensitive, or could contravene privacy laws. Decisions to restrict access to sensitive research outputs will be justified, in accordance with the [NESP data and information guidelines](#). Such exceptions to the open-access policy will be lodged with the hub data wrangler as soon as possible for inclusion in a register that will be reported periodically to the department (see [Exceptions to the open data policy](#)).

In all other cases, datasets will be published under CC-BY 4.0 (or the latest version), subject to exclusions where appropriate related to ICIP data.

## Data storage and sharing

The key principle for the storing and sharing of data and research outputs is appropriate application of the *FAIR data principles*, and co-design of research projects with stakeholders to ensure fit-for-purpose research outputs that meet their requirements. NESP datasets should be published before

the end of a project and in compliance with specified deliverables/milestones of Annual Research Plans unless there are legal, contractual, ethical or privacy requirements that prevent this. Wherever possible, data and research outputs will be published as open access with open licencing, in data repositories, digital systems and decision support tools that enable outputs to have the greatest impact.

The choice of platform for publishing data will be driven by requirements of stakeholders, end-users and other researchers, as will the data format and standards followed, and a range of technologies can be used. At the simplest level, a link in a metadata record can point to a downloadable file that contains a bundle of the dataset licence, metadata and the data files, as long as the location of the files and the data server itself is persistent and well managed.

Co-design with stakeholders is required for all research projects, and in the data and research output context this is with the aim of identifying and clarifying throughout the research project's life:

1. stakeholders' decision-making processes
2. what data and research outputs could enhance those decision-making processes
3. the most practical and efficient methods of delivering data to stakeholders
4. appropriate formats and standards
5. how should stakeholders use, interpret and apply the data.

The co-design process that should be followed is described in detail in the Climate Systems Hub's knowledge brokering strategy. To a large extent, the stakeholder engagement work should occur at the work-package or project level.

Each metadata statement needs to have a description of how data will be shared, including access procedures, embargo periods (if any), outlines of technical mechanisms for dissemination and necessary software and other tools for enabling re-use, and definition of whether access will be widely open or restricted to specific groups. The metadata statement should identify the repository where data will be stored, indicating in particular the type of repository (institutional, standard repository for the discipline, etc.).

Some persistent data repositories can assign a Digital Object Identifier (DOI) to data to provide unique identification of the data and facilitate data access and citation. Data citation ensures that the scientists who acquired or developed the data are given due recognition for their work.

The process for transferring knowledge between researchers and research end-users, and integrating research outputs into national information repositories, digital systems and decision-support tools will vary depending on the level of information (refer to [Types of research products and data](#)). While specific processes will be captured in project level data management plans, this transfer and integration should also meet the following:

- Be designed through the co-design process for the project, determining the research outputs, their characteristics and means of sharing.
- All data should have a metadata record that is registered on the hub's catalogue and on other public metadata catalogues.
- Where appropriate (level 1 information) be published in a medium such as print or a scientific journal, captured in metadata catalogues, and a copy stored on the hub website.
- In the case of code and software, where appropriate these may be published and DOIs minted through CSIRO's DAP or Zenodo as well as linking to bitbucket or github (or similar) code repositories.
- Application ready datasets and those developed for ingestion into end-user models and systems should be published in public repositories such as the National Computing Infrastructure (NCI), CSIRO's Data Access Portal (DAP), or the Australian Ocean Data Network (AODN). The NCI Data Collection is available for widely used data, with data services available to enable ingestion into decision-support tools.

- Smaller datasets may be stored in repositories such as Zenodo, particularly when data services have not been requested for the data (noting that storage with Zenodo facilitates the creation of new data services at a later point in time if desired).
- Very large raw datasets should be stored on the infrastructure where they are generated and/or archived, with a preference for locations that can be shared with other researchers. NCI data projects are ideal for this purpose.

## Approach to legacy systems

Ongoing access to NESP research outputs will provide an enduring legacy of quality-assured data and information that will assist both decision-makers and the wider research community. In order to ensure the ongoing availability of research outputs beyond the Agreement Period, the following processes will be followed:

- Before the end of a project, datasets must be appropriately described and stored in an approved persistent repository so that they remain available for future re-use. It is the responsibility of the hub's partner agencies to ensure appropriate storage for the hub's data assets, including possible costs of maintaining data storage availability.
- Data should be stored in centralised repositories, but not in those which are funded by short-term funding. Additional means of data distribution outside of repositories are permitted, but not as the only form of data delivery – for example, data should not be delivered only via a ftp or Apache server which is not part of a formal repository.
- Repositories which provide DOIs to allow the citation of research products are recommended. AODN, DAP, Zenodo and NCI all offer DOI minting services. Code and software, where appropriate, may be published and DOIs minted through CSIRO's DAP or Zenodo as well as linking to bitbucket or github (or similar) repositories.
- Regardless of where data is stored, a hub project-hosted metadata entry must be created to indicate the location of the data. Whenever a dataset is moved, the metadata for that dataset must be updated to show the new location and any additional access information.
- A statement on how the products/tools developed will continue to be supported beyond the life of the project should be described in the project data management plan. If one or more projects identify that specific websites or tools need to be developed to store or serve their data outputs, these should be recorded in the project research plan.

Most research datasets do not have a defined lifetime. They may be of value for many years to come and should therefore not be disposed of until it is clear they have no further value (or can be readily recreated). Some exceptions will be datasets like model run outputs where the model or input data was found to be significantly flawed or otherwise of no lasting value.

Raw data should be retained for the life of the dataset. Storing a copy of the original data as well as any processed or quality-controlled version will enable researchers to go back and reprocess the data if the need arises. Any software or configuration information that is relevant to the dataset should be stored with the dataset wherever possible, and where possible processing provenance (e.g., versions of software/tools used) should be recorded alongside the data. In some instances, the cost of storage may exceed the computational cost of reproducing data, in this case it may be appropriate instead to retain the version of the code used to produce the data.

Paper records such as documents and photographs, including those that have been scanned, should be stored in a way consistent with the requirements of standard Records Management Service.

## Research output submission process

The Climate Systems Hub will adhere to the [NESP knowledge brokering strategy](#) and the [NESP communications strategy](#), which outline that the department and the hub will adopt a collaborative approach to communicating about NESP and the hub, in accordance with the 'no-surprises' approach that is central to this collaboration.

Appropriate members of the department and its portfolio agencies will be engaged throughout research and be provided with a timely opportunity to consider and, where appropriate, provide comments on draft outputs prior to public release. It is expected that the outputs in scope are the end-user products and application-ready datasets, whereas it would be impractical for raw research data.

The hub will email the Science Partnerships team at least 10 working days prior to the output's release or final publication and include the following:

- an electronic copy of the output
- completed research output submission template.

All research outputs should be provided to the department prior to release, or when publicly available for peer review products.

All NESP or hub communication and media products will comply with the Australian Government style manual and the NESP brand standards (refer to [NESP brand standards](#)).

All publications and communication materials must be reviewed/approved by the hub leader, or delegate, prior to release. This, along with processes by which the hub will communicate outputs to the department, are outlined in the hub's communication strategy.

# Exceptions to the open data policy

There may be instances where open access to information may not be suitable when information is culturally, environmentally, commercially or socially sensitive, or could contravene privacy laws. Decisions to restrict access to sensitive research products should be discussed with the data wrangler and justification provided for inclusion in a register of exceptions to the open data policy. Exceptions must also be recorded in project data management plans. In cases where restricted access applies, the project lead must ensure there is an enduring copy of the unaltered data and make freely available metadata that describes the data and why it has not been released. Sensitive data may include, but is not limited to:

- culturally significant data
- social data restricted by privacy law or considerations (for example, from end-user surveys)
- data with underlying licensing restrictions that cannot be lifted (for example, no on-sharing of data)
- data with geographical-restrictions.

Exceptions to the open data policy will be managed by the hub data wrangler, who will maintain a record of all exceptions and report these to the hub leader. They will also be recorded in project reports. The hub leader will be responsible for reporting these to the Steering Committee, and the department and as part of Annual Progress Reports.

Where a dataset may be significantly more utility and relevance for decision making or in the department's business were its restrictions be eased, the project lead should be open to discussions to finding options for reducing those limitations.

# Accessibility

The impact of Climate Systems Hub data outputs will be maximised by ensuring all relevant research outputs such as websites follow the Web Content Accessibility Guidelines (WCAG), an internationally recognised standard created by the World Wide Web Consortium (W3C), and include an accessibility statement.

Please see the below resources for additional information:

- [Web Content Accessibility Guidelines \(WCAG\) 2.0](#)
- [Web Content Accessibility Guidelines \(WCAG\) 2.1](#)
- [WCAG 2.1 checklist](#).

# Acknowledgment of NESP

Support from the Australian Government must be acknowledged in all research outputs, including data, publications, presentations, promotional and advertising material etc. See the *NESP brand standards* and *NESP knowledge and communication strategy* for more details.

The below is an excerpt from the *NESP brand standards*.

To acknowledge Australian Government funding, hubs must use one of the following funding acknowledgement statements:

- The Climate Systems Hub is funded by the Australian Government under the National Environmental Science Program.
- This project is supported with funding from the Australian Government under the National Environmental Science Program.
- This project is jointly funded by [\[insert organisation/program name\]](#) and the Australian Government under the National Environmental Science Program.

# Project-level data management

Every NESP project must have a project-specific research plan that includes an outline of the approach to data management, including specific tasks/activities and associated deliverables/ milestones to ensure compliance with the hub's data management strategy. This is aimed at providing a structured method of documenting the research datasets that a project will use or generate and ensures their ease of use and longevity. The subsection in the *NESP research project plan template* on the project-level data management approach must be completed for each project. A project data management plan template is available to project leads.

Developing and maintaining project data in a manner consistent with the hub's Data Management Strategy ensures that all aspects of the project's data (including appropriate disposal) have been adequately addressed in order to minimise the risks of data problems during or after the project. The data management provisions for each project outline data that the project will acquire and generate so that the resources required for data management can be effectively planned. Well-designed data management ensures that important aspects of using third party data (such as licencing) and generating new data have been considered, ensure correct identification and attribution of data, foster adequately resourced data storage planning and facilitate data publishing. In practice, the plan should be a living document, revisited as project requirements change.

As part of the process of developing a project proposal, thought needs to be given to the datasets that the project is likely to use as well as the datasets that will result from the project. If the project will use existing data or software then the conditions under which the project will be able to use them need to be ascertained, as well as any constraints these conditions will put on the project outputs, and whether these conflict with the requirements of the hub's head legal agreement with the department. Any negotiations should aim for access to an existing dataset that covers the widest set of users possible. This information must be entered into the project's IP registers.

The project-specific data management plan will include the following items:

- Details of research outputs
- Key datasets that will be of interest to stakeholders, next-users and other researchers (including those from other NESP hubs)
- Process of co-design with stakeholders to determine their data and research output requirements, and continuing engagement to refine requirements (refer to [Data storage and sharing](#))
- Resources from the project that will be allocated for data management purposes.
- A statement on how the project will endeavour for all datasets to meet FAIR Data Principles
- Datasets to be collected, generated and published by the project, including estimates of data volumes
- Entries in a background IP (BIP) Register to record agreed inputs into the project, and into a Project IP (Foreground IP) Register to record agreed outputs of each project. Any ICIP must be identified, and separate entries added to the ICIP Register (refer to [Ownership and intellectual property rights](#))
- Where applicable, for example when ICIP is identified, how the project will endeavour to meet the *CARE principles* for managing traditional knowledge and ICIP (see [Indigenous Cultural and Intellectual Property \(ICIP\)](#))
- Designated Data Custodians for each dataset, responsible for establishing and maintaining the dataset and producing metadata (refer to [Hub roles and responsibilities](#))
- Potential repositories for each of the project datasets, or other resources required to store and share data
- A statement on how the products/tools developed will continue to be supported beyond the life of the project, including where and how they will be stored and accessed (refer to [Approach to legacy systems](#))



- The project data management plan is a living document that must be kept up-to-date by The Project Lead, who will work with the data wrangler to develop the plan. The data wrangler will provide any advice needed, and review the plan on a regular basis.

# Status of strategy implementation – tools, artefacts, and collaborative projects

As many projects closed out at the end of 2024, the Data Wrangler worked closely with project leads and project data leads on capturing details on each dataset that was created and / or published. Project data information is recorded in the hub Data Catalogue and is detailed in project Data Management Plans (see table below). The principal objective of this work is implementation of this data management strategy to maximise the impact of the data being produced by the projects with our stakeholders.

New projects in RP2025 are still designing the data products they will produce, The Data Wrangler is working with each project to document as much detail as possible for their datasets, which will be an ongoing process during the remaining life of the hub. The Data Catalogue will serve as the initial database for recording this information, with project data management plans developed in parallel.

These activities are being augmented by a project the Data Wrangler and program managers have commenced in collaboration with the Australian Data Research Commons (ARDC). The project will develop and implement better metadata for hub data outputs to improve hub research accessibility. It will deliver:

- a complete record of all data products published by the hub with details of where they are published (repositories), and the metadata schemas that were used
- an approved list of data repositories for future data publication
- an approved metadata schema that captures all information required for those repositories as well as that needed by ARDC to harvest the metadata
- a set of persistent identifiers that must be used for each dataset, including ORCiDs for researchers, DOIs for data sets and PIDs for institutions and the hub.

The project will work with researchers, project leads and data leads to implement the consistent use of the metadata schema, the repositories, and the persistent identifiers. This will ensure that all data is made accessible and findable, and that all details of datasets can be correctly identified and harvested. Ultimately, this will benefit research users who will be able to more easily discover and access data sets, thus increasing their impact for environmental decision making in Australia

Artefact/tool	Purpose	Status
Project data management plan	To outline project's approach to data management and to record information about data products, software code and input data.	Each project has a data management plan. Updated periodically by data custodians and project leads
Data product submission form	To capture high-level information when data products are planned, and more detailed information when they are approved or published	Project leads input data and update as data products mature. Information captured stored in the data product catalogue

Data product catalogue	To record high-level information about the data outputs of the project, which informs planning for approval and publication. It also helps track the uptake of data products and their impact	Contains the planned data outputs of each project and is updated periodically
IP catalogue	To capture intellectual property and copyright information about data inputs and outputs. Also includes licensing information of inputs and outputs for determination of whether open licensing is appropriate	Updated periodically when new data sets are added to the data product catalogue
Metadata template	Provides a consistent set of requirements for metadata records of each dataset	Is available for researchers to complete when data is ready to be published

## Risks

Number	Potential risk	Risk treatment strategies	Relevant sections of this plan
1	Access to hub output restricted due to constraints on input data or software	Negotiate for less restrictive licence for output	<u>Ownership and intellectual property rights</u> <u>Data licencing</u> <u>Exceptions to the open data policy</u>
2	Outputs of hub fail to meet stakeholder needs or are not fit for purpose	Co-design process and stakeholder roadshow Work package specific advisory panels	<u>Data storage and sharing</u> <u>Project-level data management</u>
3	No access to Indigenous data	Establish relationships within the Indigenous community(ies) and ensure content is obtained prior to project commencement.	<u>Ownership and intellectual property rights</u>
4	Inadequate data storage capacity, security and/or management arrangements.	Projects determine data storage requirements and secure them as early as possible; record requirements in project data management plan	<u>Data storage and sharing</u> <u>Project-level data management</u>
5	Inadequate metadata management to ensure legacy data of hub is captured sufficiently after the life of the hub.	Metadata records will be required of all projects following agreed templates and recorded in a hub-level metadata repository	<u>Metadata</u> <u>Project-level data management</u>
6	High dependency on NCI infrastructure, whose long-term future is uncertain and whose availability cannot be guaranteed at any time	Close engagement with NCI to stay abreast of any changes to availability. Communicate through department very high value of ongoing existence of NCI	<u>Approach to legacy systems</u>
7	Potential loss of access to data from NOAA, NASA, and other US institutions and agencies that may affect some hub data products	Changes in access to relevant data will be closely monitored and impacts on projects minimized wherever possible through sourcing of alternate data	<u>Ownership and intellectual property rights</u>

# Related materials

The following related material is available on the internet or has been provided to hubs:

- [Australian Government branding–Guidelines on the use of the Australian Government logo by Australian government departments and agencies](#)
- [Australian Government Open Data Toolkit](#)
- [Australian Government public data policy statement](#)
- [Australian Government style manual](#)
- [Australian Research Data Commons](#)
- ARDC [Data management plans](#)
- ARDC [Working with data](#)
- [ARDC working with data](#)
- [NCI Data Management](#)
- [ANZLIC metadata profile](#)
- [Commonwealth library deposit and free issue schemes](#)

Data repositories:

- [Atlas of Living Australia \(ALA\)](#)
- [Australian Ocean Data Network \(AODN\)](#)
- [Australian Urban Research Infrastructure Network \(AURIN\)](#)
- [data.gov.au](#)
- [Integrated Marine Observation System \(IMOS\)](#)
- [Terrestrial Ecosystem Research Network \(TERN\)](#)
- [National Computational Infrastructure \(NCI\)](#)
- [Australian CMIP repository management \(NCI\)](#)
- [Australian ERA5 repository management \(NCI\)](#)
- [Geographic information metadata: ISO 19115-1](#)
- [NESP brand standards](#)
- [NESP grant opportunity guidelines](#)
- [Open Geospatial Consortium](#)
- Summary of Australian Universities [data management policies and tools](#)
- [University of Edinburgh: Sources of dataset peer review](#)
- [Web Content Accessibility Guidelines \(WCAG\) 2.1.](#)

This strategy should be read in conjunction with the:

- [National Environmental Science Program knowledge brokering and communications strategy](#)
- Climate Systems Hub *communications strategy*
- Climate Systems Hub *Indigenous partnerships strategy*
- Climate Systems Hub *knowledge brokering strategy*