Data Management Plan

VERSION 6

July 2020

Version control revision history Data Management Plan			
Version	**Date revised**	**Reviewed by (Name, Position)**	**Comment (review/amendment type)**
V4.0		Aurel Moise	• 1st draft
V4.1	3/4/2018	Geoff Gooley	• Minor edits clarifying concept of the Data Management Plan
V4.2	10/4/2018	Marian Sheppard	• Reformatting and addition of version control table
V4.3	13/4/2018	Aurel Moise	• Incorporating edits and comment into final version
V5.1	13/7/2019	Aurel Moise	• Added new Project data managers as required
V6.0	18/5/2020	Claire Trenham	• Update approaching NESP1 close out
			•
			•
			•
			•
			•

About this document

Actual delivery for this deliverable:

Lead author: Claire Trenham (Chair, DMWG) and Geoff Gooley (ESCC HPMT)

Contacts: hubcentral@nespclimate.com.au
Visit us on: http://www.nespclimate.com.au/

Contents

Executive summary

The Data Management Plan (DMP) for the NESP Earth System and Climate Change (ESCC) Hub provides an overview of the arrangements for managing and accessing the Hub's research data, information and associated terms of use. The DMP reflects the current state of the discussions, plans and intentions of the NESP ESCC Hub partners, and will be updated on an annual basis (as part of the annual Research Plan) as work progresses. Responsibility for the DMP development and implementation is with the Hub's Data Management Working Group (DMWG), a representative forum across all of the Hub's partners and research projects.

The DMP sets the framework for the handling of data produced in the NESP ESCC Hub, from curation to dissemination, and assures management of ESCC Hub data beyond the lifetime of the Hub. It therefore describes the life cycle of modelling, reanalysis and observation data collected and processed as part of the NESP ESCC Hub research project delivery.

The focus of the DMWG is on encouraging good data management as an essential element of research best practice, noting that researchers are required to make all NESP research outputs publicly available on websites with a persistent and enduring link in compliance with specifications of the Hub's Research Plan and the associated NESP Data Management and Accessibility Guidelines.

1　Introduction

Why a Data Management Plan (DMP)?

The prerequisite for meaningful use, re-use or recombination of research data is that they are well documented according to accepted and trusted standards. Those standards form a key pillar of science because they enable the recognition of suitable data. To ensure this, agreements on standards, quality level and sharing practices have to be negotiated.

In compliance with the *NESP Data Management and Accessibility Guidelines V3.0 (2017)* (see Appendix A) the NESP ESCC Hub has a requirement to develop a 'data and information' management plan for all projects approved under the annual Research Plan. The purpose of the plan is to facilitate 'successful delivery of open access research', with emphasis on ensuring discoverability, accessibility and utility/re-use of all Hub 'research products' (that is, the Hub aims to follow FAIR data principles).

What kind of data are considered in the DMP?

For the purposes of this DMP, the definition of 'research products' includes:

- New raw data sets including model outputs and spatial data produced by ESCC-related work
- Data analysis and derived data products such as GIS maps
- 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 (e.g. Decision Support Tools) as well as new apps and APIs
- Analysis code developed to produce data products and journal publications where appropriate
- Websites and apps for mobile phones, tablets etc
- All publications including journal papers, books/book chapters, reviews
- All 'grey' literature (e.g. Fact Sheets, posters, technical reports etc)
- Visualised data, including images, maps, videos etc
- Unspecified emerging technology

A key requirement of this plan is for projects to routinely record, collate and communicate metadata statements based on a standard template for all 'research products' generated from Hub-funded research and outreach activities. The metadata statements are based on accepted best practice (including being consistent with the National Environmental Information Infrastructure and mandatory requirements of the Australian Standard ANZLIC Metadata Profile).

For the ESCC Hub, the metadata template (see Appendix B) was adopted from the Australian Climate Change Science Program (ACCSP). The ACCSP metadata template was constructed with support from CSIRO IM&T as part of the development of the CSIRO Data Access Portal

linked to relevant national repositories including the Australian National Data Service (ANDS) data discovery portal. Minimum requirements for metadata statements for NESP include:

- To be available online, with links to the relevant products (data/code/documents etc) and accessible via metadata discovery facilities
- Identifies attributes, methods and procedures used for determining all values within relevant datasets
- Uses standard taxonomies, and/or defines or links to online definitions of all terms used in description of the datasets
- Provides contacts and access locations for the data, and
- Provides provenance for any data that are used in generating research products.

Scope

The target audience of the DMP is all NESP ESCC Hub project team members, partners and other research organisations using the data and information produced by the Hub.

2 Principles

The following guiding principles apply for all NESP ESCC Hub data management activities:

2.1. All research data generated by the Hub is deemed 'open-access', and ownership of such data and associated IP created by researchers within the NESP ESCC Hub is determined according to and consistent with the relevant provisions of the NESP head legal agreement between DAWE and CSIRO as the host agency for the Hub. Hub data and IP are not the property of individual Hub researchers.

2.2. NESP ESCC research data will be made publicly available whenever legally, ethically and contractually possible, where necessary/appropriate subject to relevant licence conditions to manage NESP ESCC's intellectual property rights and/or other technical requirements (for example to ensure data are used appropriately within specified limitations).

2.3. A whole-of-life-cycle project approach should be taken to managing NESP ESCC datasets commencing with the project planning stage and ending with the data in an appropriate persistent repository for long term management and curation to ensure future identification and availability.

2.4. Each research dataset developed by a project must have an identified custodian, and responsibility for the dataset must be transferred to an ongoing custodian or generic email contact in the relevant Hub partner agency prior to the end of the project.

2.5 The privacy of human subjects will be maintained by managing all data involving human subjects in accordance with standard Privacy Policy and the Ethical Conduct of Human

Research Procedure, including full compliance with CSIRO ethics approval processes (unless otherwise agreed with the HPMT).

2.6.   All project datasets must be registered and adequately described in the NESP ESCC metadata catalogue using the specified standard Hub metadata template, whether the datasets are hosted by NESP ESCC or elsewhere. Such metadata statements are to be lodged with HPMT 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.

2.7.   Research data must be retained in a durable and retrievable form indefinitely, unless a case is made that the data are of zero potential value for future reuse (for example containing erroneous values, test data, model outputs of no long term significance, or data that is so easily re-created that long term storage is not justified). The responsibility for storage and curation of the Hub's data is with the relevant Hub partner agencies as determined by the relevant project LCI in consultation with the HPMT.

2.8.   Third party datasets acquired for project internal use should be acquired under the widest possible licence provisions (for example permitting NESP ESCC-wide reuse wherever possible), with appropriate acknowledgment to the source described in the relevant metadata statement. Datasets obtained from third parties for use in NESP ESCC research must be appropriately licenced for open publication of derived products.

2.9.   NESP ESCC will comply with all other (i.e. non NESP/Hub specific) relevant contractual and legal obligations and existing data licence conditions (e.g., Privacy Act) when deciding how to manage data (for example where, when and how to publish).

2.10.   Project LCIs and data managers are required to be aware of any underlying agreements and associated/other conditions relevant to data management/access whether or not they are explicitly stated in NESP Hub contracts.

2.11.   NESP ESCC DMWG will oversee data management policy for the Hub, including to facilitate where possible convergence of data management activities and infrastructure needs.

3   ESCC Hub Roles and Responsibilities

Lead Chief Investigator (LCI) is responsible for the day to day operation of the Hub's projects. This person will oversee implementation of the DMP at the project level and keep it updated as the project evolves. The LCI will appoint a data custodian to the datasets that the project will use, collect or generate. There may be more than one data custodian in a project. In most cases the LCI is the person who creates the initial project proposal that is used to seek approval to pursue the project and is therefore also the person responsible for including data management activities at the project level. The LCI is also the person to represent the project on the Hub DMWG, unless otherwise delegated (to one or more nominated data custodians).

Data Custodian is the person who is assigned formal administrative responsibility for the appropriate management of the project datasets by the LCI. This person is responsible for

ensuring the safety of data during the project, creating metadata records and ensuring it is appropriately stored. A data custodian retains responsibility for a dataset until that responsibility is passed to someone else, ensuring that information and metadata about the dataset is accurate and up to date. Before the end of the project, the Custodian, in consultation with the HPMT, should choose a suitable long-term repository for the data, deposit it and make sure that the data is published (i.e. made publicly available) wherever possible.

Suggested data custodians from current research projects:

ESCC Hub Project	Data Contact/DMWG member	Email
Project 5.1 (Rashid)	H. Rashid, P. Dobrohotoff, C. Mackallah	harun.rashid@csiro.au peter.dobrohotoff@csiro.au chloe.mackallah@csiro.au
Project 5.2 (Chung)	C. Chung, F. Delage	christine.chung@bom.gov.au francois.delage@bom.gov.au
Project 5.3 (Grose)	C. Trenham, T. Erwin, S. Narsey	claire.trenham@csiro.au tim.erwin@csiro.au sugata.narsey@bom.gov.au
Project 5.4 (Kirono)	C. Heady, H. Zheng	craig.heady@csiro.au hongxing.zheng@csiro.au
Project 5.5 (Dowdy)	A. Dowdy, M. Thatcher, H. Ye, K. Walsh, J. Evans	andrew.dowdy@bom.gov.au harvey.Ye@bom.gov.au marcus.thatcher@csiro.au kevin.walsh@unimelb.edu.au jason.evans@unsw.edu.au
Project 5.6 (Canadell)	P. Briggs	peter.briggs@csiro.au
Project 5.7 (Sloyan)	K. Stewart B.Sloyan	bernadette.sloyan@csiro.au, kial.stewart@anu.edu.au
Project 5.8 (McInnes)	C. Trenham	claire.trenham@csiro.au
Project 5.9 (Swearer)	S. Swearer, T. Konlechner	s.swearer@unimelb.edu.au t.konlechner@unimelb.edu.au

4   Data Management Plan

All project proposals must include specific tasks/activities and associated deliverables/milestones to ensure compliance with the Hub's DMP. The DMP is a structured method of documenting the research datasets that a project will use or generate and ensures their ease of use and longevity.

Developing and maintaining project data in a manner consistent with the Hub's DMP ensures that all aspects of the project's data (including appropriate disposal) have been adequately addressed so that the chances of data problems during or after the project are minimised. 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. Having a well-designed DMP for the Hub also assists researchers

ESCC Data Management Plan July 2020

in meeting requirements of the NESP Hub legal agreement with DAWE. 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 then the Project LCI/Proponent needs to ascertain whether the datasets exist and, if so, under what conditions the project will be able to use them, what constraints, if any, these conditions will put on the project outputs and whether these conflict with the requirements of the Hub's head legal agreement with DAWE If it is necessary to negotiate for access to an existing dataset, the LCI/project proponent should negotiate a licence that covers the widest set of users possible.

The Hub's head legal agreement with DAWE is explicit about the ownership of the intellectual property rights over any project outputs, including datasets and standard IP conditions therefore apply unless different arrangements are otherwise agreed.

A template for a project data management plan (Appendix B) is downloadable at the ESCC Hub SharePoint site (https://teams.csiro.au/units/nesp/_layouts/15/start.aspx#/). Some of the major considerations in the Hub's DMP are outlined in subsequent sections of this document.

A recent initiative from the European Community to unlock the potential of data resources is the implementation of the FAIR Data Principles (see Appendix C), abbreviating that data should be Findable, Accessible, Interoperable and Reusable (FAIR). The FAIR principles were generated to improve the practices for data management and data curation, and FAIR intends to describe the principles in a way that is domain-independent and hence can be applied to a wide range of data management purposes, whether it is data collection of individual researchers or data management of larger research projects regardless of scientific disciplines.

5   Ownership and Intellectual Property Rights

At the start of each project, a clear understanding of the ownership of the rights, including any IP rights associated with each acquired or generated research dataset and any contractual constraints must be documented in accordance with the *NESP Data Management and Accessibility Guidelines V3.0 (2017)*. All contracts for research work will include specific statements about the ownership of the intellectual property in data and data products from the project. Similarly, ownership of data should be incorporated into all contract and IP

negotiations. The IP/Rights statement should be included in the metadata record for the dataset so that it is visible to all.

If data belongs to a project which involves other IP such as patents and confidential information and early publication may risk IP rights, the data may not be made available unless it has been cleared for publication by the relevant IP Manager.

6 Data Licencing

The NESP ESCC Hub data licencing procedure complies with the spirit of government 'Open-Source Data'. It aims to increase the impact of NESP ESCC research by supporting national and international research efforts through the free exchange of data collected or generated by the NESP ESCC Hub, while ensuring that NESP ESCC Hub retains any intellectual property rights in the datasets or is otherwise able to adequately manage the data in accordance with the NESP guidelines.

NESP Hub funding agreements require all research outputs to be made publicly available under the latest Creative Commons framework using a Creative Commons Attribution licence. This should allow for research outputs to be deposited into an appropriate repository. For code and software, open licences such as GPL, MIT or the CSIRO licence may be appropriate.

7 Metadata

Metadata is intended to represent a definitive holding of NESP ESCC's data assets. A metadata record should be made for every dataset that a project collecting or generating – using the NESP-ESCC Metadata Template. Metadata captures information essential to understanding a dataset. It also enables discovery of a dataset, maximising re-use. Information that should be included in a metadata record includes details appropriate to the research discipline, which will be incorporated into the standard metadata schema appropriate to the discipline.

8 Data Storage and Management during a Project

The DMP addresses the issue of storage for data arising during the planning and operation of a project. All project data should be stored in a durable location with appropriate backup, ideally a centralised server or a file share managed and backed up by a central facility (e.g. NCI). If a portable device is used for data collection in the field, then it should be transferred to more robust infrastructure when back at the workplace (robust meaning that the data is not at risk of loss from a single hardware loss or failure). It is the responsibility of the Hub's partner agencies to provide appropriate storage for the Hub's data assets.

When data are stored in files, there should be a data registration system or documented naming convention 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.

When datasets are stored on shared infrastructure, the custodian should ensure that the access permissions for those files are appropriate to protect against accidental or malicious corruption. Write permission to these files should be restricted to the people who need to make changes and locked to future changes once data is published. Read access to these files can and should be open unless they contain personal information or there are contractual constraints requiring restrictions.

9   Data Formats

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 (and specifically 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 as they encapsulate much of the metadata required to use a dataset.

Any data in a non-digital form, such as field sheets or traces from instruments, should be turned into a digital form for ongoing management. Paper records should be scanned into PDF files, with sufficient resolution to ensure that the information is readable. These PDF files should be archived in the corporate record keeping system, TRIM, and the TRIM reference included in the metadata record. They can also be published in the same way as any data files. The data content of the records should be keyed into digital files of an appropriate format for analysis.

10   Standards

The NESP ESCC Hub will work to accommodate domain data standards or conventions, and these 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.

11   Data Storage and Management at the End of a Project

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. Repositories managed by local data management groups are recommended as the default for long-term storage of datasets. It is

the responsibility of the Hub's partner agencies to ensure appropriate storage for the Hub's data assets. Examples of suitable repositories include:

- repositories for international programs such as Argo and the World Ocean Database
- Accredited repositories such as the AODN (ocean data) or CSIRO's Data Access Portal (smaller data sets and other research products)
- The Oceans and Atmosphere Data Warehouse
- NCI's repository (large scale data)

Data stored in centralised repositories should also be published via appropriate webservices including for direct download (FTP or HTTP) and OPeNDAP for netCDF data. WMS and other OGC web services should also be provided where possible.

Repositories which are funded by short-term funding are not considered to be suitable repositories. Data distribution outside of repositories is not recommended as the only form of data delivery – e.g. data should not be delivered only via an 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 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 repositories.

Regardless of where data is stored, a NESP-ESCC 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.

12   Retention and Disposal

Most datasets in domains of research do not have a defined lifetime. They may be of value for many years to come and should therefore be retained indefinitely. 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

ESCC Data Management Plan July 2020

processing provenance (e.g. versions of software/tools used) should be recorded alongside 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.

13   Publishing and Data Citation

NESP ESCC datasets should be published before the end of the project and in compliance with specified deliverables/milestones of Annual Research Plans unless there are legal, contractual, ethical or privacy requirements that prevent this. The default licence for data is the *Creative Commons Attribution CC-BY 4.0*.

Datasets can be published using a range of technologies. 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.
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 the data are given due recognition for their work.

14   Data Sharing

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.). In case the dataset cannot be shared, the reasons for this should be mentioned (e.g. ethical, rules of personal data, intellectual property, commercial, privacy-related, security-related).

15   Definitions

Research data   qualitative or quantitative attributes of a variable or set of variables. In the context of this document, data includes measurements, observations, values derived from measurements or observations and the output of models, as well as economic, social and other data types acquired for research purposes. Data may be digital (for example numeric values, text, digital imagery and video) or non-digital (for example field or laboratory data sheets and printouts, maps, analogue imagery).

Dataset   a set of related data values that forms a coherent unit, for example as collected by a common method for a particular study, such as all CTD data recorded on a particular voyage or the output from a particular run of a model.

Data management   the execution of practices and procedures to ensure that datasets are suitably designed, stored, protected, described and available.

Metadata   information that properly and adequately describes a dataset.

ESCC researcher   all references to staff in this document mean people within NESP ESCC Hub projects as well as affiliates.

Affiliate   a person associated with a NESP ESCC project; including contractors, consultants, seconders, visiting scientists, fellows, students and trainees.

Project Proponent   the person who creates the initial project proposal that is used to seek approval to pursue the project (typically the LCI).

Lead Chief Investigator (LCI) the person responsible for the day to day management of the project (often also referred as the 'project leader').

Data Custodian   the person with administrative responsibility for the appropriate management of a defined dataset.

Persistent data repository   a persistent repository is a facility where a collection of datasets is stored and maintained in a rigorous way so that they remain discoverable and accessible over the long term using the same address. The repository technology is managed so that it remains current and data integrity is regularly checked.

Data publishing   to make data freely available and accessible to others, normally by on-line self-service or connection to appropriate data portals. Also referred to as data sharing.

Appendix A: Excerpt from NESP Data Accessibility Guidelines regarding peer-reviewed publications

Peer reviewed publications

Peer-reviewed publications include: peer reviewed scientific papers; books; and other peer-reviewed published material. An electronic copy of all peer-reviewed articles (as accepted for publication after peer-review) must be made openly and freely available on the internet, if not immediately, then within/upon 12 months of publication. Researchers are not permitted to grant a licence for NESP funded articles to publishers if that licence does not allow the hub to place a complete copy of the article (not just an abstract or citation) on its website within twelve months of publication.

Understanding pre-print, post-print and re-prints: Which should be made publicly available?

Journal copyright arrangements often distinguish between three versions of a published peer-reviewed article - pre-print, post-print and re-print. A pre-print refers to a paper that has been submitted for publication but which has not yet undergone peer-review. The manuscript is termed a post-print after the author has re-written the manuscript based on the comments of reviewers or editors. The re-print refers to the publisher's final PDF version of the paper encompassing the journal-specific formatting and which contains the same figures and text exactly as the post-print.

Tools for understanding journal copyright arrangements

It is the responsibility of the research hubs to understand the copyright and associated licensing arrangements being entered into when signing agreements with publication houses. Internet tools such as SHERPA/RoMEO can be used to determine the default position of the publication house.

Where the default position of a publication house does not conform to the NESP requirements for open access, hubs may be in a position to negotiate licensing arrangements by stipulating that their funding body has a requirement for all research products to be open access. Many journals will negotiate the licence conditions applying to individual journal articles in cases where research funding organisation has an open access policy. Some publications allow researchers to pay a fee to make the article open access. Payment of open access fees is the responsibility of the research hub. The Department's preference is for research funding to be used for research, in preference to publication fees.

Alternatively, it is often possible to negotiate with the publisher to publish under an alternative licence scheme that meets the Australian Government's open access policy. AusGOAL provides licences for a range of information access situations that are designed to manage legal risks associated with making information available for reuse and are endorsed by the Office of the Australian Information Commissioner. These include a non-commercial licence which may be

acceptable to the publisher. Figure 1 is a flow diagram, which may aid in determining whether publishing in a specific journal conforms to the NESP funding agreement specifications.

Figure 1. Choosing a journal that conforms to the NESP funding agreement publication specifications.
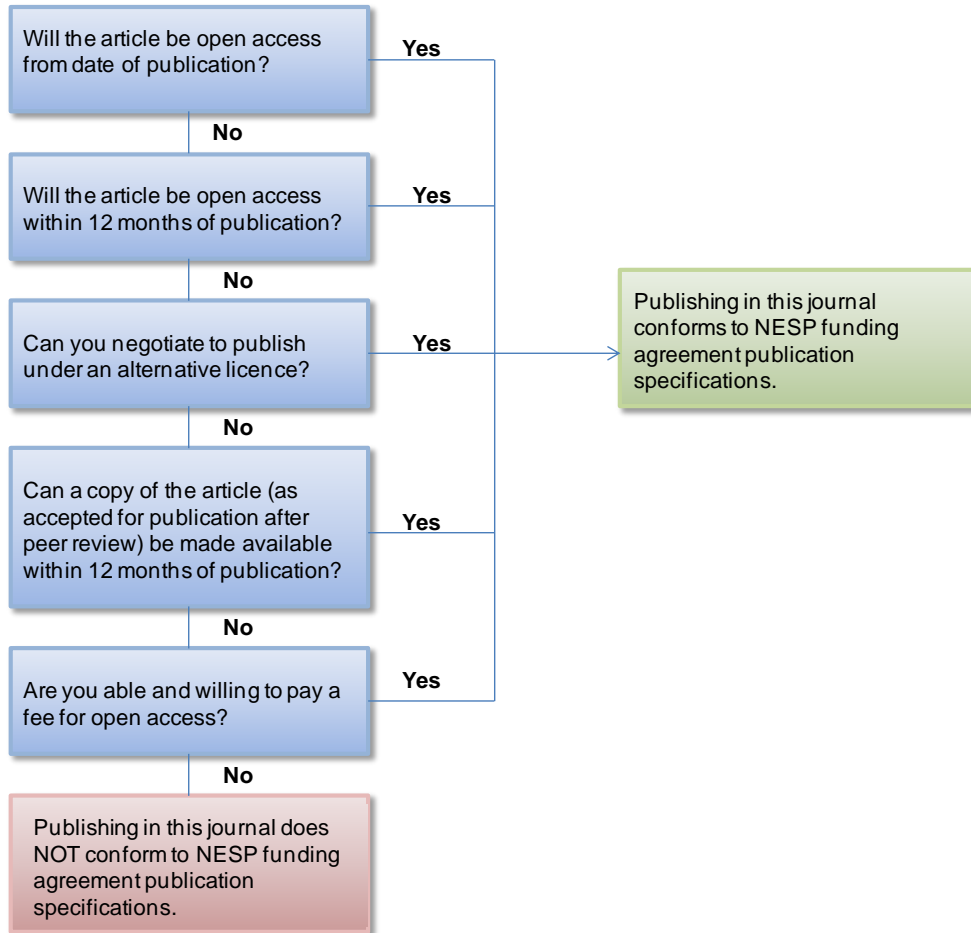
Appendix B: NESP ESCC Hub Metadata Template

The link for ESCC Hub researchers to the current meta-data template is here:
http://teams.csiro.au/units/nesp/_layouts/15/start.aspx#/SitePages/ProgrammeManagement.aspx

Example of a completed metadata form for a dataset

Project details	
ESCC Project No. & Title(s)	2.1 Preparing ACCESS for CMIP6 5.1 ACCESS evaluation and application
Project Leader(s)	Simon Marsland/ Harun Rashid
Deliverable(s)	ACCESS-CM2 model output data for CMIP6 publication
Data/software Manager	Chloe Mackallah

Research output data collection, tool or software URLs if applicable
Data located at NCI's ESGF node (https://esgf.nci.org.au)

Description	
Title	ACCESS-CM2 model output prepared for CMIP6 under 'CMIP'
Description	These data includes all datasets published for 'CMIP6.CMIP.CSIRO-ARCCSS.ACCESS-CM2' according to the Data Reference Syntax defined as 'mip_era.activity_id.institution_id.source_id.experiment_id.member_id.table_id.variable_id.grid_label.version'. Experiments include the DECK (*piControl, abrupt-4xCO2, 1pctCO2, amip*) and *historical* simulations.
Temporal and Spatial Extent	Global, various year ranges.
Lineage	Met Office HadGEM3-GA7.1, ACCESS1-OM2, CABLE2.5, CICE5.1.2
Credit	CSIRO Coupled modelling team & ARCCSS
Keywords	Coupled modelling, CMIP6, ACCESS, climate, climate change
ABS Fields of Research Category / Subcategory*	37 Earth Sciences 3701 Atmospheric sciences 3702 Climate change science 3708 Oceanography

*These are listed in
https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1297.02020?OpenDocument

Attribution/IP (complete for data, tools and software)	
Owning Organisation	CSIRO
Collaborating Organisations	CSIRO & ARCCSS
Primary contact for this data	Chloe Mackallah access_csiro@csiro.au, chloe.mackallah@csiro.au
Lead Researcher	Dave Bi dave.bi@csiro.au

Contributors	CSIRO coupled modelling team & ARCCSS
Access	Publicly available at NCI's ESGF Node (https://esgf.nci.org.au)
Licencing	CC BY-SA 4.0
	Consult https://pcmdi.llnl.gov/CMIP6/TermsOfUse for terms of use governing CMIP6 output, including citation requirements and proper acknowledgment.

Related materials: publications, tools, websites, related input data. Please provide full citations for publications, data and software.	
Details	**URL**
ACCESS-CM2 model description paper	TBA (Bi et al. 2020, JSHESS, under review)
Local data copy at NCI on Gadi in project fs38	gadi.nci.org.au:/g/data/fs38/publications/CMIP6/CMIP/CSIRO-ARCCSS/ACCESS-CM2
Forcing data based on 'input4mips'	https://doi.org/10.22033/ESGF/input4MIPs.1118 https://esgf-node.llnl.gov/projects/input4mips
Model configuration at Met Office Science Repository Service (MOSRS)	https://code.metoffice.gov.uk/trac/um
CSIRO Research website for ACCESS	https://research.csiro.au/access/cmip6-submission/
DKRZ citation & DOI	https://cera-www.dkrz.de/WDCC/ui/cerasearch/cmip6?input=CMIP6.CMIP.CSIRO-ARCCSS.ACCESS-CM2 https://doi.org/10.22033/ESGF/CMIP6.2281
ES-DOC model documentation	https://github.com/ES-DOC-INSTITUTIONAL/csiro-arccss-bom/
NCI Data Catalogue record	https://geonetwork.nci.org.au/geonetwork/srv/eng/catalog.search#/metadata/f1481_6421_2540_3569

Technical details	
For data: Total size of this data collection	18 TB

For data: **Total number of** **Files**	9,038
Current location of **files (data or** **software)**	gadi.nci.org.au:/g/data/fs38/publications/CMIP6/CMIP/CSIRO-ARCCSS/ACCESS-CM2
Format(s)	netCDF4
Associated tool(s)/ **dependencies**	Rose and cylc, netCDF
Proposed **publication host**	NCI; Earth System Grid Federation (ESGF)

Appendix C: FAIR principles adopted from NCI

FAIR is integrated into NCI's Data Quality Strategy as it provides a reference framework for measuring the potential contribution of data collections and data services to enhance research outcomes. The following describes how the FAIR principles have been applied at NCI:

Findable: The datasets on the NCI NERDIP have catalogue entries that are accessible via human and machine harvestable interfaces. The metadata standard used is conformant with the ISO 19115 standard for discovery of geospatial information and can be cross-walked with the RIF-CS profile of ISO 2146 used by ANDS Research Data Australia, as well as the Dublin Core and Data Catalog (DCAT) metadata standards used by data.gov.au. Conforming with multiple metadata standards and profiles significantly increases the discoverability of NCI datasets, both nationally and internationally. Both metadata and data are assigned a globally unique and persistent identifier. Data records are in a searchable resource and harvestable for external cross-walk.

Accessible: Datasets on the NCI platform are made accessible for general research access (e.g., data download for small file sizes), as well as being suitable for advanced techniques and multiple applications, including virtual laboratories, portals, common desktop tools, and programmatic access via well-known network protocols. Metadata are retrievable via standardised international CS/W repository protocols.

Interoperable: Wherever possible international data standards for interoperability are applied including metadata standards at both data services and at the data level; controlled vocabularies and interchangeable self-describing data formats (e.g., NetCDF4/HDF5); and accessible via network protocols and community standard APIs.

Reusable: Rigorous QA/QC procedures are used to validate the data against domain relevant community standards so that users are assured that the data can be accessed in consistent ways. The QA/QC validation also demonstrates that the data works across different (non-domain specific) packages, tools and programming languages deployed by the various user communities thus extending the use of the data across domain silos. All datasets have assigned license / terms or use, including access controls to allow access to data (currently only internally within NCI).

The QC and QA reports for data published at NCI record compliance with the NCI Quality and FAIR data criteria.

Appendix D: DMWG Terms of Reference (TOR)

Mission:

The Data Management Working Group (DMWG) is established by and reports to the NESP ESCC HLT. The primary objective of the DMWG is to assist the ESCC Hub in the coordination and facilitation of data management activities/issues within each of the projects. Membership consists of a data management expert from each project (plus a deputy where appropriate).

Membership:

Each ESCC project will be invited to nominate a data management expert and a deputy (where appropriate) to serve on the DMWG. The DMWG Chair will be appointed by the ESCC HLT (i.e. the relevant HLT functional coordinator) and will report to the HLT and the DMWG will be supported by the HPMT as appropriate. In addition, the Chair will be responsible to report on a monthly basis as required, to the HLT on DMWG recommendations and progress on action items approved by the HLT. The normal term for each DMWG member will be twelve months and can be renewed. The current membership is listed in a Table on page 9.

Mode of Operation:

The DMWG will normally conduct its activities by correspondence (electronic or regular mail); however, it will meet in person when and where appropriate.

The major responsibilities of the DMWG are:

1. to serve as the principal advisory group in all matters pertaining to data management activities/issues and the coordination and exchange of data/metadata across the ESCC Hub projects; with specific emphasis on facilitating Hub compliance with requisite provisions of the NESP *Data and Information Management Guidelines*.
2. to act as point of contact for data management questions on behalf of the ESCC Hub or assist (whenever possible) in data requests. The DMWG will also coordinate and/or support data management activities/issues with other organizations where appropriate (e.g. NCI, IMOS, etc.).
3. to develop a draft Data and Information Management Plan template for implementation at project level by Lead CIs in compliance with the *NESP Data and Information Management Guidelines* This Data Management Plan will address specific issues regarding identification of available data sets at the existing holdings at NCI, etc. (i.e. formats, resolution, frequency, parameters, etc.) and establish specific working protocol procedures among the ESCC Hub participating members. The final data management plan will be approved and distributed by the ESCC HLT.
4. To design, draft, and periodically review a Hub Data Management "Home Page" sitting on the NESP ESCC web page to provide current information on data policies, availability, access/links, and data set status of the various projects. This page will be approved by the HLT and linked to the ESCC Hub's "Home Page".