Sustainable Communities and Waste

National Environmental Science Program

Data management strategy

National Environmental Science Program Sustainable Communities and Waste Hub

Strategy aims

The Sustainable Communities and Waste (SCaW) Hub data management strategy will provide a framework for how the Hub and its researchers will achieve findable, accessible, interoperable and reusable (FAIR) research products. It is recognised that discipline-specific standards of data management will also apply, and researchers will be encouraged to apply these standards wherever possible.

This strategy aims to:

- guide data wrangling activities in the Hub, in particular outlining how the Hub will manage data at all stages of research
- ensure that FAIR principles are embedded in all Hub activities, and that Hub activities are consistent with the NESP data and information guidelines
- provide clarity on the activities that data wrangling actor(s) in the Hub will undertake.

The aims of the Data Wrangling function in the Hub are to:

- manage all data collected through Hub activities in accordance with the FAIR data principles in a sustainable manner for the long term
- make all Hub data and data products freely and openly available unless specific restrictions apply (indefinitely or for a limited time) for reasons of privacy, ethics, sensitivity, or commercial confidence
- ensure research outputs are presented in an accessible form that encourages reuse and maximises impact on management and policy decision making
- ensure published data acknowledges the Hub and associated researchers and requires subsequent citation and acknowledgement when reused by implementing appropriate licensing, persistent identifiers, and acknowledgment guidelines
- apply existing established research data management policies, standards, and guidelines
- promote collaboration and support the work of management agencies, researchers, Indigenous peoples, organisations both for profit and not for profit operating within the sustainability and waste sectors, and community groups, through its data management approach.

Data and information management approach

In accordance with the funding agreement, the Sustainable Communities and Waste (SCaW) 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, DAWE, 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.

This 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
- 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
- facilities and equipment required.

For the SCaW Hub, data management is guided by several existing programs and platforms, including the Australian National Data Common (ARDC), Australian Urban Research Infrastructure Network (AURIN), and Research Data Alliance groups on Data Management, Physical Samples, and Research Data Collections. In addition, the projects in this Hub would benefit from the Research Data Alliance recommendations on Data Management Plans (DMP).

The SCaW Hub brings through its founding members an experienced and knowledgeable data management team already embedded in the SCaW research community. Additional resources from the ARDC brings standards, infrastructure knowledge and economies of scale.

This strategy will be implemented through a combination of action plans for communication and knowledge brokering, as well as plans for Indigenous participation, data management, and individual projects. These are in addition to and will sit alongside annual research plans and the Hub's guiding 'strategy, governance and operating model' document.

Data and information management planning is essential to achieve the successful delivery of open access research. A Data Management Plan (DMP) will be developed for the Hub and for research and project plans to bring into effect this strategy.

Fundamental to all data management will be adherence to Indigenous cultural and intellectual property (ICIP) guidelines and principles. Overseeing this requirement will be a collaborative effort between the Hub's Data Wrangler and the Hub's Indigenous Facilitator.

The Australian Research Data Commons (ARDC) provide a helpful overview of data management plans and the FAIR data principles. Many Australian universities have data management policies and tools available for use by researchers needing to create a data management plan at the start of a research project.

As outlined in the Australian Code for the Responsible Conduct of Research, Hub researchers will have primary responsibility for managing their research data. The Hub will provide a Data Management Framework as a practical guide for Hub researchers outlining requirements and processes required to effectively manage and publish their data and provide resources to assist them in doing so.

Metadata and data collected or collated by the Hub where applicable will be published in publicly accessible data repositories complying with FAIR data principles. Some data may be subject to restricted access based on its classification under privacy legislation or for ethical, commercial or sensitivity reasons (see exceptions examples to the open data policy are found here)).

In this early stage of the Hub's establishment phase, the Hub Data Wrangler will be working closely with the Knowledge Broker and others to establish a process to enable potential use of institutional, national, and discipline-specific repositories for data storage. Furthermore, this strategy aims to enable and encourage dissemination of research outputs metadata through Research Data Australia services provided by Australian Research Data Commons (ARDC).

The strategy will be revised annually following a process of review and consultation with DAWE and other key stakeholders.

Given research planning for the Hub is still at the early establishment phase, more information will be available in the future revisions of this document.

Hub roles and responsibilities

Hub Data Wrangler role

The Data Wrangler's activities will include working with the Hub, researchers, DAWE, 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 the data is produced and used in accordance with the FAIR principles of being discoverable, accessible, and useable, to optimise the use and reuse of public data.

The Data Wrangler is responsible for coordinating and conducting data discussions with research projects, providing guidance to projects on best practice data management, reviewing project data management plans, and tracking data management milestones.

The Data Wrangler will also work with other Hub staff; including the Knowledge Broker and Communications and Media Officer, researchers, and other stakeholders to maximise the usability of the Hub research and increase the accessibility of data to policymakers.

The Data Wrangler role is a part-time position.

Researcher role

The Sustainable Communities and Waste (SCaW) Hub researchers will be responsible for developing data management plans for each project, in conjunction with all project team members and the Hub Data Wrangler, and in accordance with this strategy.

Researchers will have primary responsibility for managing their research data and must comply with agreed data management and publication milestones as outlined in their project plan.

Researchers are responsible for ensuring intellectual property requirements in the use of third-party data are appropriately managed to not restrict access to NESP funded research outputs.

Hub Leader role

The SCaW Hub Leader will be responsible for providing oversight of data management activities and access to related infrastructure for the Hub. The Data Wrangler will be responsible for recording any exceptions. Records will generally be at project level and include information on the data generated or collected and the justification for its non-release. The publication of research outputs must comply the NESP data and information guidelines.

Partner/ external organisation data specialists

Where partners or external organisations provide access to data management infrastructure or other specialised data management roles, they must comply with the relevant sections in this document and the SCaW Hub Data Management Framework (document in progress).

Given that research planning, including the finalisation of key roles and responsibilities for the Hub is still at the early establishment phase, more information will be available in the future revisions of this document.

Types of research products and data

A broad range of research products are expected to be generated throughout the life of the program by the Sustainable Communities and Waste (SCaW) Hub. These products can be broadly categorised as written outputs or data outputs and may include the following:

Written outputs:

- publications including scientific papers, reviews, books, book chapters
- grey literature including fact sheets, project profiles, manuals, and technical report
- communication, knowledge brokering and Indigenous partnership strategies, procedures, and frameworks.

Data Outputs

- raw datasets including spatial data
- analysis and data products such as geographic information system-derived maps
- images, maps, photos, videos, animations
- models and other tools, such as decision support tools or software.

The SCaW Hub is still at the research planning phase, more information about research outputs will be available following further specification of project details.

In future revisions, new information will be available about types of raw datasets and other new data formats.

Ownership and intellectual property rights

At the commencement of each Sustainable Communities and Waste (SCaW) Hub project, a clear understanding of the ownership of the rights, including any intellectual property rights associated with each acquired or generated research dataset and any contractual constraints, must be documented in accordance with NESP data and information guidelines.

To achieve this overarching aim the Hub Leader and Impact Priority Leaders will ensure that the ownership of rights invested in a project, including any intellectual property rights associated with each acquired or generated research dataset and any contractual constraints, will be in accordance with NESP data and information guidelines.

This strategy does not seek to alter existing intellectual property ownership rights defined or assigned under agreement, including the NESP SCaW Hub Agreement and agreements between staff or students and their research organisations or educational institutions.

In general:

- Intellectual property ownership of data collected in the execution of Hub-funded projects will remain with the party or partner carrying out the project, and may be governed by employment conditions or other agreements binding individuals, or additionally guided by principles such as the Global Indigenous Data Alliance (GIDA) CARE principles (see below).
- Owners of intellectual property vested in data generated in the execution of Hub funded projects will be required to make their data freely and openly available in such a way that complies with the FAIR data principles and NESP data and information guidelines (unless it meets the exceptions to the open data policy).
- Researchers will be required to prepare a data management plan for all Hub funded projects which will include details of intellectual property ownership and licensing for data generated as part of a project or introduced as background intellectual property.

Indigenous cultural and intellectual property (ICIP)

NESP adheres to the objectives of the Global Indigenous Data Alliance (GIDA) with respect to Indigenous data, especially in relation to access of data by non-Indigenous users. Although the NESP data and information guidelines follow the FAIR data principles, when working with Indigenous data these guidelines require the complementary use of GIDA's CARE principles for Indigenous data governance, which consider both people and purpose as part of open data and information.

Knowledge held by Indigenous peoples will be recognised, valued, and protected during the operation of the Hub. This will be guided by the CARE principles for Indigenous data governance: Collective benefit, Authority to control, Responsibility, and Ethics.

The Hub and its researchers will ensure legal obligations are understood before collecting information (including free and prior informed consent) and be guided by the overarching GIDA objectives.

Just as NESP acknowledges the rights of Indigenous peoples to their cultural and intellectual property (ICIP) the obligation to recognise, value and protect any ICIP rights will be incorporated into all the SCaW Hub project agreements in accordance with the funding agreement with the Hub.

The Hub acknowledges that all research that is undertaken, irrespective of its nature, will affect Indigenous people. As outlined in the Indigenous Partnership Strategy and supported by this Data Management Strategy, projects will apply the NESP 3-Category Approach to Indigenous Engagement and be guided by these recommendations for planning, designing, collaborating, sharing, and

communicating with Indigenous partners. CARE principles will be applied at all stages from initial project planning through to final data curation. For Categories 2 and 3 projects, all decisions regarding the curation and publication of data will be made in consultation with Indigenous partners. While the Hub's Data Management Strategy aims to apply FAIR data principles alongside CARE principles, if conflict occurs, CARE principles will take precedence.

Given that research planning, including ownership and intellectual property rights is still at the early establishment phase, more information will be available in the future revisions of this document.

Data licencing

All Hub research data are required to be made available through an appropriate data repository under an appropriate Creative Commons or Open Source Initiative approved (software) license to enable flexible public reuse, unless specifically exempt under the exceptions to the open data policy or the funding agreement. Third party material collated by, or supplied for use in, Hub research activities, is also subject to these guidelines, unless data use agreements between third party data providers and Hub researchers (or organisations) explicitly prohibit this.

At this early research planning phase, the Sustainable Communities and Waste Hub does not have a central data repository; however, the Data Wrangler will work with the Steering Committee and the Hub Knowledge Broker to establish the requirements for open access infrastructures and address the need of making research outputs publicly accessible.

Data management and storage

Metadata

The Sustainable Communities and Waste (SCaW) Hub expects individual Hub researchers to create high-quality metadata records for all data resulting from NESP funding. Appropriate training will be provided to researchers to achieve this expectation. This approach will be incorporated into our research planning and project specifications. This provides the contextual information needed to enable the research 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.

At this early stage of the Hub, the Data Wrangler is planning to work closely with the Hub Knowledge Broker and researchers to establish a pipeline for publishing the Hub's research outputs metadata through the Australian Research Data Commons (ARDC) data discovery portal, Research Data Australia. This pipeline would enable unified publishing and discovery of all data collection using the Registry Interchange Format - Collections and Services (RIF-CS) based on ISO 2146.

Persistent identifiers

Persistent identifiers are globally unique numeric and/or character strings that reference a digital object and are guaranteed to be managed and kept up to date over a defined time (intended for the long term).

Aligned with the FAIR data principals, the Hub Data Wrangler is working with the Hub Knowledge Broker to create a process for assigning a persistent identifier to all NESP data and, where possible, research outputs, to allow trusted, uninterrupted access.

Persistent identifiers will vary depending on the type of research output and may include an International Standard Book Number (ISBN), International Standard Serial Number (ISSN), a Digital Object Identifier (DOI) or a web URL. Assigning a Digital Object Identifier (DOI) to data and other research products facilitates data citation and is considered best practice. A DOI is a type of persistent identifier that indicates a dataset will be well managed and accessible for long term use. It is now routine practice for publishers to assign DOIs to journal articles and for authors to include them in article citations.

Research output submission process

The NESP knowledge brokering and communications strategies outlines that DAWE and the SCaW Hub will adopt a collaborative approach to communicating about NESP and the Hub. A 'no-surprises' approach is central to this collaboration.

This allows DAWE and other Australian Government agencies the opportunity to prepare, where necessary and appropriate, a response to the research prior to its release. It also a allows the Science Partnerships team to manage all incoming outputs and have a timely copy on hand.

Appropriate members of DAWE and its portfolio agencies will be engaged throughout Hub research and be provided with a timely opportunity to consider and, where appropriate, provide comments on draft outputs prior to public release.

Written outputs are to be submitted to the Hub website as drafts for approval by the Hub leadership. Submission of a draft output instigates the Hub's approval and publishing processes. This will include assignment of an ISBN and (once approved) submission to national digital libraries for public access. Additional guidance will be provided in the Hub's Data Management Framework (document in progress).

Consistent with this approach the Hub Data Wrangler, Knowledge Broker and / or Communication and Media Officer will, where possible, email the Science Partnerships team at least 10 working days prior to the output's release or final publication and include the following information:

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

A summary of research outputs from the preceding month will be provided by the Hub to the Science Partnerships team. The summary of research outputs will be consistent with NESP guidance on research output submissions.

As referenced in the Hub's communication and knowledge broking strategies, all Hub communication and media products will comply with the relevant departmental guidelines and the NESP brand standards including NESP brand standards for correct acknowledgement of funding sources.

Open-access outputs

The Hub will, wherever possible, follow Open Access principles and make its research outputs available at no charge, using best practice approaches specific to the research output type.

In the case of scientific publications, researchers should make reasonably practicable efforts to either publish via Creative Commons license or make scientific publications otherwise publicly accessible within 12 months of publication.

The Hub will apply FAIR principles to all data outputs. Hub data will be accompanied by descriptive metadata records published using a standardised, interoperable schema allowing outputs to be discovered and harvested by national data aggregation services (e.g. Research Data Australia).

Open access to information may not be suitable in cases where that information is culturally, environmentally, commercially or socially sensitive, or could contravene privacy laws (see sections *Indigenous cultural and intellectual property* and *exceptions guide to the open data policy*), or the Hub funding agreement. Decisions to restrict access to sensitive research outputs must be justified.

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.

The SCaW Hub will leverage existing individual university partner and research capabilities. However, all research outputs will be published via the most relevant institutional, Government, publication or domain specific services that demonstrate sustainable support for the long term. Where necessary, contribution to the development, configuration, or sustainability planning to meet Hub requirements may be made, but the Hub will not directly operate, or own infrastructure used to publish and/or archive research outputs other than its own website platform.

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 justified and made by those closest to the source. In cases where restricted access applies, an enduring copy of the unaltered data and metadata, describing the data and why it has not been released should be made available.

Sensitive data may include, but is not limited to:

- location information for highly desirable or collectable species
- location information for rare species
- culturally significant site data
- social data restricted by privacy law or considerations
- other heritage or sensitive Indigenous matters.

Hub Project Plans and Data Management Plans will be required to identify potentially sensitive data and information that will be generated during a project, and should trigger discussions in that regard (e.g. PhD theses and publications). These plans must also contain strategies for how exceptions to the open data policy will be managed to ensure adequate security of the information and meet any legislative or other requirements (e.g., GIDA's CARE principals).

It is the researcher's responsibility to communicate and justify requests for exceptions to the open data policy to the Data Wrangler. The Hub's Data Wrangler will collate instances of exceptions to these guidelines. Proposed exceptions will be reported to the Steering Committee and DAWE, with information about the project and justification for the exception. It will then be captured in the Annual Progress Report by the Hub Leader. Ideally, owners of the data that constitute the exception will be open to discussions with DAWE regarding that data, the potential utility and relevance of the data to the Department's business. If relevant, portions of that data could be provided so as to assist decision-making.

The NESP data and information guidelines will be provided to all project teams to enable determination of both open access requirements, and exceptions to, and conditions around this policy. A further action will be outlining how exceptions to the open data policy will be managed out-of-session as part of an internal consultation process on data yet to be developed. This will be undertaken as and when projects start to formulate over the next year. For example, this is likely to involve seeking written endorsement from the Steering Committee and DAWE where exceptions may be raised.

Accessibility

The Sustainable Communities and Waste (SCaW) Hub Data Wrangler works with the Hub Communication and Media Officer to ensure all web content complies with the Web Content Accessibility Guidelines (WCAG) 2.0.

All online systems used by the Hub will comply with the Web Content Accessibility Guidelines in accordance with Australian Government requirements under the Disability Discrimination Act 1992 to ensure that online information and services are accessible by people with disabilities.

Furthermore, for those to be involved in publishing information to the Hub's website, we will ensure relevant Hub staff undergo web content accessibility training, including the process for products to be developed, reviewed and amended for correct/consistent accessibility. The Hub will consider user design and testing by people with disabilities.

The Hub will also consider how the Hub will ensure its products are accessible for varied disabilities/products: viz videos and animation (e.g. using closed captions), images (e.g. using alt text), webinars (e.g. using an interpreter), and types of documents used (e.g. PDFs are not compatible with JAWS so ensuring each document has a Word accessible version).

Acknowledgment of NESP

Support from the Australian Government will be acknowledged in all research outputs, including data, publications, presentations, promotional and other public material in accordance with the latest available NESP funding acknowledgement documentation.

As referenced in the Hub communication strategy, copies and summaries of these requirements will be provided to all to all Project teams to enable determination and application of the correct brand standards.

Project-level data management

The Sustainable Communities and Waste Hub will have project-specific research plans that include the outline of the project-level approach to data management. As the Hub is in its establishment phase, the Hub Data Wrangler is working with the Hub Data Knowledge Broker and the project leaders to determine a process to review and approve the project-level approach to data management. More information will be available in the future revisions of this document.

Related materials

Note: Further information will be available following the confirmation of the hub's research plan and the research projects.

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
- ANZLIC metadata profile
- Commonwealth library deposit and free issue schemes
- Data repositories:
  - Atlas of Living Australia (ALA)
  - Australian Urban Research Infrastructure Network (AURIN)
  - data.gov.au
  - Terrestrial Ecosystem Research Network (TERN)
- NESP brand standards
- NESP grant opportunity guidelines
- 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*
- Sustainable Communities and Waste Hub *communications strategy*
- Sustainable Communities and Waste Hub *Indigenous partnerships strategy*
- Sustainable Communities and Waste Hub *knowledge brokering strategy*