Australian Research Data Commons

Planet Research Data Commons

FEEDBACK FROM CONSULTATIONS

By Kerry Levett

The Australian Research Data Commons

December 2022

NCRIS National Research Infrastructure for Australia
An Australian Government Initiative

The ARDC is enabled by NCRIS

DOI: 10.5281/zenodo.7455648

Suggested citation:

ARDC Ltd. (2022). Planet Research Data Commons feedback from consultations. Viewed online at: *http://doi.org/10.5281/zenodo.7455648*

CONTENTS

Cover Image — themorningglory - 391783296 / stock.adobe.com

The ARDC is enabled by NCRIS

BACKGROUND

ABOUT THE PLANET RESEARCH DATA COMMONS

A Thematic Research Data Commons is a vehicle for the Australian Research Data Commons (ARDC) and our national partners to collaboratively develop and deliver sustainable digital research infrastructure on a national scale. It will enable us to best meet the needs of our diverse national research communities in a strategic and comprehensive way.

The Planet RDC will support earth and environmental science researchers to develop cross-sector and multi-disciplinary data collaborations on a national scale. It will integrate underpinning compute, storage infrastructure and services with analysis platforms and tools that are supported by expertise, standards and best practices. And it will bring together data from a range of sources to tackle the big questions.

We're taking a co-design approach to developing the Planet RDC through collaborative partnerships between research institutions, industry, government agencies and other relevant national stakeholders.

The Australian Research Data Commons (ARDC) has undertaken consultations with national stakeholders for the Planet Research Data Commons (or 'Planet RDC'). The purpose of these consultations was to:

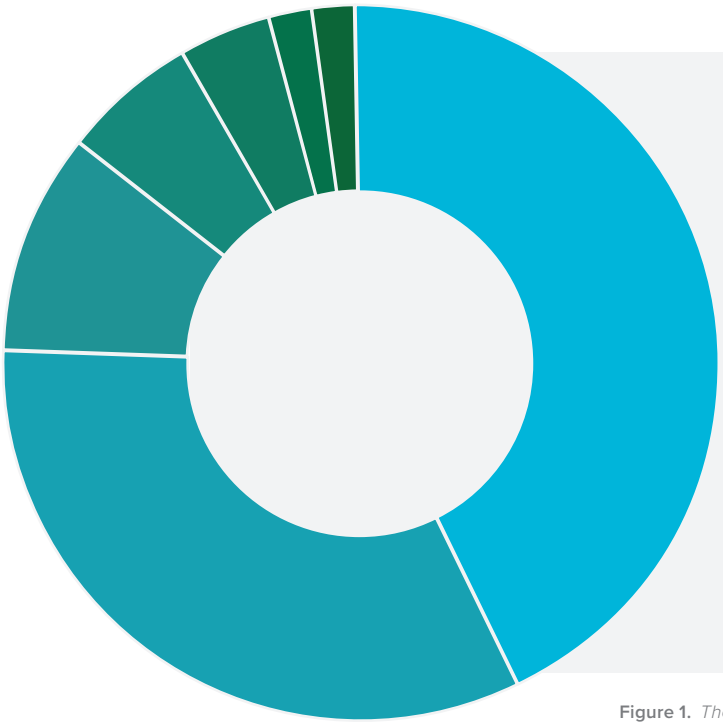
- identify the key data challenges in earth and environmental research, as well as the associated priorities
- help define the value proposition of the Planet RDC
- learn where initial efforts need to be focused in establishing the Planet RDC.

The ARDC identified relevant data challenges to be addressed in the Planet RDC through an analysis of the national research and research infrastructure priorities related to environmental research and management.

We conducted the consultations in phases, with initial feedback sought from the ARDC Research and Technical Advisory Committee, and the National Earth and Environmental Sciences Facilities Forum (NEESFF).

We had individual consultations with 37 researchers and research infrastructure providers from the earth and environmental sciences.

We hosted 3 public roundtables on 7, 10 and 12 October 2022, which were open to all interested stakeholders. There were 49 attendees across the 3 sessions. The attendees were predominantly researchers and research infrastructure providers from NCRIS facilities and universities. There were also representatives from eResearch providers, industry, government, and non-government organisations (NGOs).



ALL ATTENDEES: SECTOR

- University - 42.9%
- NCRIS - 32.7%
- Government - 10.2%
- PRFA - 6.1%
- eResearch Provider - 4.1%
- Industry - 2.0%
- NGO - 2.0%

Figure 1. *The distribution of the 49 attendees of the three Roundtable sessions by sector*

ALL ATTENDEES: ROLE

- Researcher- 36.7%
- Infrastructure Provider - 36.7%
- Research Support - 16.3%
- Decision Maker - 8.2%
- Other - 2.0%

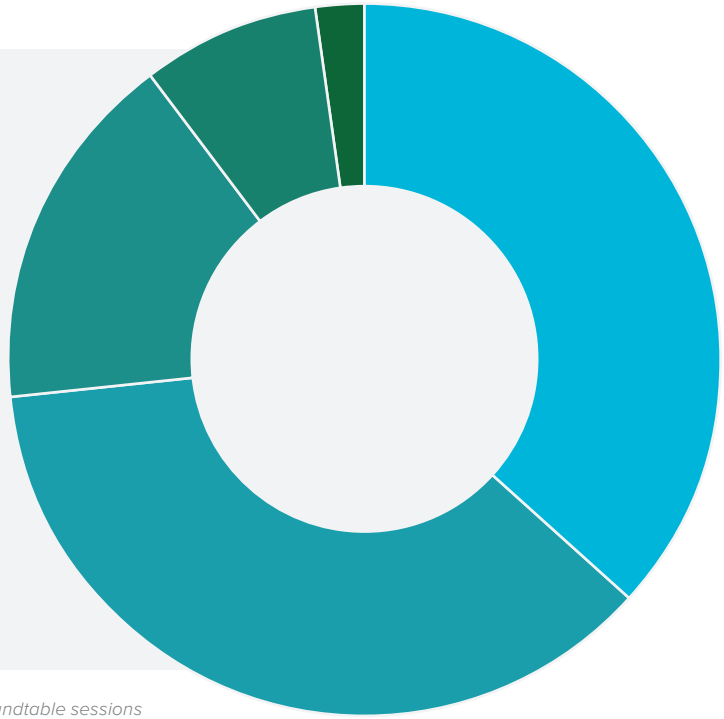


Figure 2. *The main roles of the attendees of the three Roundtable sessions*

BACKGROUND

Image — René Riegal - 0VFr6ub7HIA / unsplash.com

The focus of the initial consultations was on "***What*** *the Planet RDC needs to deliver*" and "***Why*** *the challenges addressed by the Planet RDC would be transformative for research".* This report summarises the insights gained from the initial discussions with stakeholders and the roundtables.

The consultations confirmed that providing nationally connected, shared infrastructure that supports data access, analytics and community collaboration across disciplines is needed to address the complex challenges facing our planet. It was noted that the ARDC's vision complements and builds upon existing world-class infrastructure provided through NCRIS, governments and universities.

The next stage of Planet RDC implementation process will be focused consultations with NCRIS facilities, research infrastructure providers, government, industry and research institutions. The consultations will validate identified challenges and ask "***How*** *can the Planet RDC deliver infrastructure, services and capability to address identified challenges by extending existing capabilities, both within the ARDC and across the sector?*"

WHAT – CHALLENGES AND PRIORITIES

The Planet RDC roundtables tested the relevance and priorities of the following data challenges in earth and environmental research, which are detailed in the Appendix:

1. Enabling federated discovery and access
2. Making it easier for researchers to collect FAIR-ready data
3. Enabling processing of large or streamed datasets
4. Facilitating curated and integrated datasets
5. Developing networked modelling infrastructure
6. Providing guidance on appropriate repositories
7. Supporting the use of decision support tools
8. Consistent data governance and data sharing arrangements

At each roundtable, the attendees were polled to determine the relevance of the 8 proposed data challenges and their priority. The results from these polls showed that all 8 challenges were relevant (Figure 3). The challenges in order of priority (Figure 4) were:

1. Curated and integrated datasets
2. Federated discovery across
3. FAIR-ready data collection
4. Consistent data governance and sharing arrangements
5. Processing of large or streamed datasets
6. Networked modelling infrastructure
7. Decision support tools
8. Guidance on appropriate repositories

Participants were asked to suggest additional challenges. The majority of the suggestions could be characterised as more specific examples of one of the 8 proposed broad challenges. For example, specific datasets that could be integrated, or what is needed in modelling infrastructure. The other suggestions were:

- Foundational infrastructure–the need for short and long-term storage–for data, models, workflows, model outputs, and orphaned data
- Data preservation–actively managing and curating data to ensure long-term access
- National approaches to data governance and data supply chains–the ability to link and integrate data from multiple sources including research, government and industry, and update data assets
- Data provenance–the ability to track provenance information is critical for data (re)use in terms of quality, versioning and correct attribution.

WHAT – CHALLENGES AND PRIORITIES

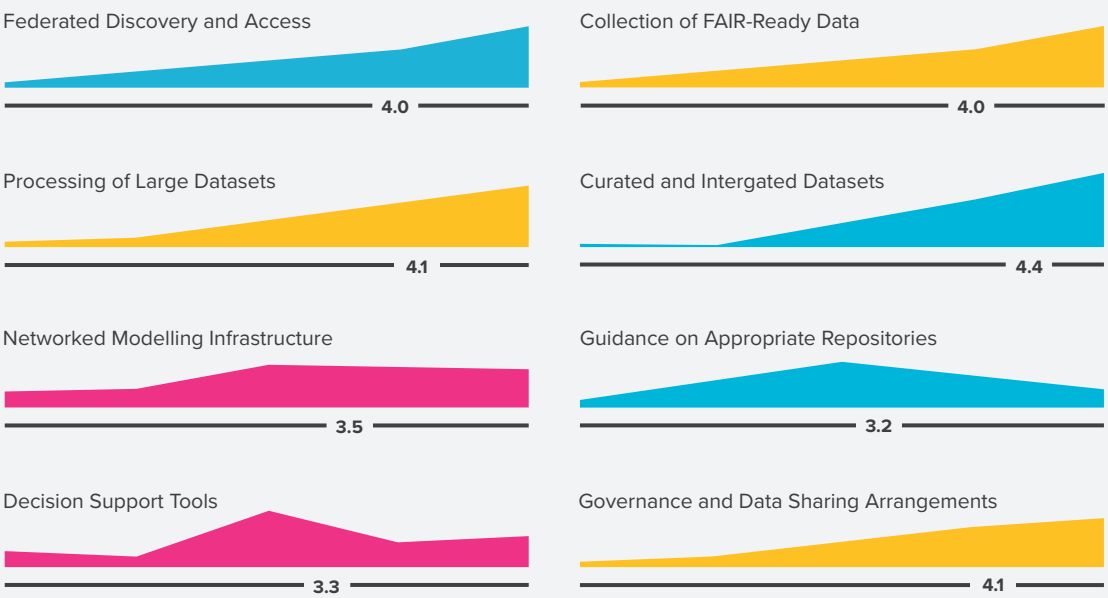


Figure 3. *Poll results for the question "How relevant are these data challenges" from the 3 public roundtables (49 respondents). The circles show the average "relevance" score from 0 (not relevant) to 5 (extremely relevant).*

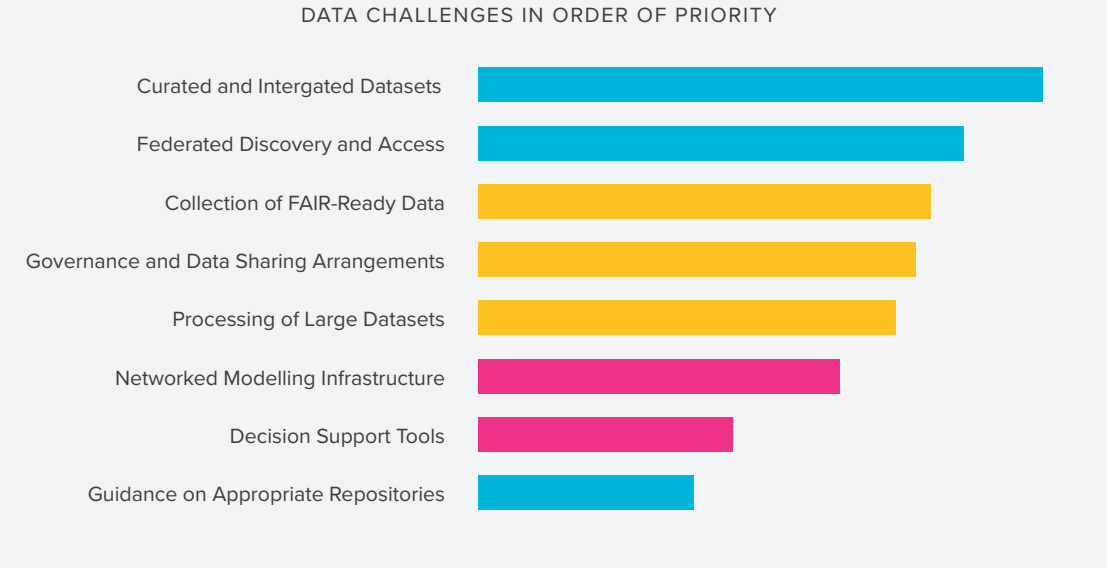


Figure 4. *Poll results for the question "Rank the data challenges in order of priority" from the 3 public roundtables (49 respondents).*

CASE STUDY – PROCESSING CAPABILITY FOR AUTOMATED MONITORING DATA

Image — Ioan Panaite - 91242387 / stock.adobe.com

The Planet RDC could work with the national research infrastructure facilities that are collecting diverse climate and environmental data and predicting future changes.

Automated acoustic, imaging and video monitoring of the environment is set to revolutionise the broad-scale monitoring and management of threatened species and ecosystems by capturing a permanent, direct, scalable and objective record of the environment. Extracting meaningful insights from diverse automated monitoring and remote sensing data has the power to enable continental-scale monitoring and management, which was identified as a gap in the 2021 National Research Infrastructure Roadmap (2021 Roadmap), and will be critical in informing Australia's adaptation strategy to environmental and climate threats. The 2021 State of the Environment report also highlights the need for new ways to manage and process the massive volumes of remote sensing and other data to "observe multiple facts of biodiversity and ecosystems" (SOE 2021: Overview. p26).

To address one of the step-changes outlined in the 2021 Roadmap, the Planet RDC could work with the national research infrastructure facilities that are collecting diverse climate and environmental data and predicting future changes. The Planet RDC and partners could operationalise and scale existing data processing platforms by providing underpinning compute and storage. This stability would allow connections to other infrastructures, such as modelling platforms and other types of data collections. This model could then be used to develop solutions for other types of dynamic data streams.

WHY - IMPACT AND VALUE PROPOSITION

The feedback from the consultations elicited some of the benefits and value of addressing the identified data challenges.

CHALLENGE - Collection of FAIR-Ready Data

IMPACT AND VALUE

- Multiple actors and institutions collect data on the environment. Enabling digital data to be born standardised and machine-readable allows it to easily flow through the data lifecycle.
- Standardised metadata models built into field data collection will greatly aid data integration, federation and interoperability.
- With more consistent data standards, land management groups that collect their own data could:
  - contribute their data to broader uses and;
  - more readily use other data as it would now match better with what they collect.

CHALLENGE - Processing of Large Datasets

IMPACT AND VALUE

- Automated monitoring has the potential to transform biodiversity and threatened species monitoring through continental-scale monitoring. This requires large-scale data analysis (close to the collections) using deep learning and other machine learning techniques.
- Will enhance data-driven approaches to reporting and monitoring programs such as natural capital accounting, State of the Environment and sustainability initiatives.

CHALLENGE - Curated and Intergated Datasets

IMPACT AND VALUE

- Large, national, comprehensive, geospatial and temporally integrated datasets are fundamental for research advances. They make it possible to address complex, multidisciplinary, cross-jurisdictional problems (e.g. fisheries, food security, biodiversity and biomass at national scale).
- Large scale novel data integrations, such as genome, species and location, are needed to meet emerging research areas in species population management, rehabilitation and biosecurity.
- Curating historical plot data would enable analyses of change through time and space, crucial to forecasting changes for ecosystem functions and services such as fire activity, stream flow, fauna habitat, and carbon sequestrations.

WHY - IMPACT AND VALUE PROPOSITION

CHALLENGE - Networked Modelling Infrastructure

IMPACT AND VALUE

- In the geosciences, networked modelling infrastructure would facilitate an integrated interdisciplinary approach. This would provide a compelling way of deriving knowledge about our planet's past using the power of distinct complementary datasets, which would not be possible by examining individual datasets in isolation.
- Minimise duplicated modelling efforts, which are very time-consuming and produce valuable IP, particularly if model outputs are treated as data assets and made FAIR.

CHALLENGE - Decision Support Tools

IMPACT AND VALUE

- Data could be utilised by decision makers to help them make more robust, data-driven decisions.
- Help deliver prioritised responses to emergencies for biodiversity and natural resource management such as bushfires and floods.
- Enhance the ability to translate research data, models and applications to decision making processes and analysis.

CHALLENGE - Governance and Data Sharing Arrangements

IMPACT AND VALUE

- Data governance is a fundamental concern and needs to be considered across research, government and industry data supply chains.
- Consistent data sharing arrangements will greatly reduce the barriers and increase the confidence of researchers actively contributing towards FAIR data.
- Proper data governance will allow Indigenous data custodians to appropriately share their data and knowledge, and for researchers and land managers to make appropriate use of Indigenous data and knowledge.
- Having access to restricted-access data (e.g. on threatened species) will enable researchers and governments to make much better conservation assessments, hence know where to invest into recovery.

CASE STUDY – ECOASSETS

Image — primipil - 462128013 / stock.adobe.com

Through the ARDC's EcoAssets project, the ALA, TERN, and IMOS are working together to deliver access to integrated environmental data in forms that can support national, state and territory environmental assessment and reporting activities

Australia's environmental research infrastructures and their partners collect large amounts of data each year on all aspects of the environment. In the past, this data has not been integrated across infrastructures or provided at a scale or format suitable for environmental reporting at national or state and territory levels. Through the ARDC's EcoAssets project, the ALA, TERN, and IMOS are working together to deliver access to integrated environmental data in forms that can support national, state and territory environmental assessment and reporting activities. Some facets of these data are of particular importance for large-scale environmental assessments such as State of Environment reporting. These assessments will draw on EcoAssets for information on trends for threatened species (those that are of highest conservation concern) and introduced species (one of the most significant drivers of biodiversity change), and how well Australia's protected area system is preserving the country's rich biodiversity.

NCRIS National Research Infrastructure for Australia An Australian Government Initiative
The ARDC is enabled by NCRIS

WHO - NATIONAL STAKEHOLDERS

The range of national stakeholders for earth and environmental research and management is diverse across sectors including research, research infrastructure, government, industry, and NGOs.

The table below summarises the feedback and comments from the 49 roundtable participants and the 37 researchers, research infrastructure providers, government and NGO staff consulted individually, mapped to the stakeholder categories. This is presented in a strengths, challenges, risks and opportunities framework regarding national digital research and decision support infrastructure aligned with the earth and environmental science research domains.

Researchers

STRENGTHS

- Australia is good at monitoring ecosystems across large areas
- Some researchers have existing relationships with state monitoring agencies
- Strong digital expertise and capability

RISKS

- 'Reinventing the wheel' or 'not adding value' due to the lack of alignment between the initiatives underway nationally and internationally

CHALLENGES

- Protocols and data standards are foundational for a commons and need to be (developed and) implemented
- Lack of time-series data collected and made available
- Navigating NCRIS is difficult for researchers
- Identifying appropriate storage facilities for storing large amounts of data collected (e.g. from automated monitoring)
- Lack of infrastructure and skills to build and publish data products and decision-tools
- Researcher-developed tools and models need to scale to become national-scale infrastructure

OPPORTUNITIES

- Delivering research data supply chains for environmental management, environmental accounting and policy making
- Linking with international initiatives
- Implementing FAIR and CARE principles for earth and environmental data
- Providing underlying infrastructure for researcher-developed tools to operationalise them
- Linking research and data specialists together to address particular challenges or data needs
- Providing analysis ready cloud data storage and tools

WHO - NATIONAL STAKEHOLDERS

Research Infrastructure Providers (NCRIS, University eResearch/IT, eResearch Organisations)

STRENGTHS

- Domain expertise
- Existing relationships with researchers and other users
- Mature sampling / data collection infrastructure
- Existing digital research infrastructure (repositories, platforms, tools, data transfer networks, etc)
- Strong connections with international domain communities
- Experienced in use of domain (meta)data standards and sampling methodologies

RISKS

- The fast-pace of technological advances could mean infrastructure is obsolete before it is complete

CHALLENGES

- Different levels of maturity in digital research infrastructure
- Need for infrastructure to capture and store different data types
- Need to pivot from only delivering data for researchers, to also delivering information for decision makers
- Current (seed) funding landscape is difficult for maturity and sustainability
- Cannot currently manage and make accessible (long-tail) research data

OPPORTUNITIES

- Increased alignment National Earth Environment Science Facilities (NEESF) strategies and programs would enable a step-change
- Research data integration programs that combine data from different facilities into new data products

WHO - NATIONAL STAKEHOLDERS

Government - Federal, States and Territories, Agencies

STRENGTHS

- Favourable policy environment- multiple government policy and strategy highlights need and willingness to make data available
- Significant initiatives underway at federal and state levels, ie: the Commonwealth Nature Positive Plan that will establish a Data Division that provides access to high quality authoritative environmental data sources.
- Strong existing relationships and translation programs between NRI and government

RISKS

- Governments may not have sufficient trust in research infrastructure to integrate it into their operational processes
- Lack of easy access to government-held data is a critical risk for research

CHALLENGES

- Governments need enduring infrastructure they can rely on
- Training needed for government users to use models/decision support tools
- Data sharing agreements are ad hoc, and licencing varies across agencies and jurisdictions
- Privacy and data sensitivity issues remains a challenge without standardised access and identity frameworks

OPPORTUNITIES

- Data supply chains for implementing national environmental economic accounting, including time series and fine-grained data
- Working with research sector to provide government data to researchers, and getting access to decision tools in return
- Assisting government to adopt persistent identifiers that are common in the research sector
- Government initiatives primarily focus on the data requirements for government. The Planet RDC can help bridge the gap for research engagement with government data initiatives
- Estabalishing translation pathways between research, government and industry

WHO - NATIONAL STAKEHOLDERS

NGOs

STRENGTHS

- Domain expertise
- Large data collection programs across different biota of Australia
- Strong public engagement and interest

RISKS

- Multiple NGOs makes it difficult to engage as a sector
- Access to technical and digital infrastructure

CHALLENGES

- Inadequate resources to implement standards or new processes
- Access to storage, platforms and connected compute infrastructure

OPPORTUNITIES

- Assisting NGOs to collect FAIR-ready data and become part of a data supply chain
- Partnering with NGOs who work with Indigenous Land and Sea Country managers

Industry

STRENGTHS

- Big data assets in environmental impact assessments
- Emerging programs that showcase the value of industry data to research and government such as Shared Environmental Analytics Facility, WA
- Significant investment from commercial cloud providers in sustainability and environmental modelling tools

RISKS

- Lack of a strategic long term approach to building industry engagement

CHALLENGES

- Need to define the value proposition of a commons for industry
- Commercial and privacy sensitivities on data collected
- Cumulative regional impact analysis requires trusted data supply chains and robust, secure analytics

OPPORTUNITIES

- Assisting consultants to collect FAIR-ready data and become part of a data supply chain
- Partnership with commercial cloud providers to meet data and analytics challenges

HOW - IMPLEMENTATION FACTORS

Australia has world-class data infrastructure from our NCRIS partners Atlas of Living Australia (ALA), Australian Plant Phenomics Facility (APPF), Australian Urban Research Infrastructure Network (AURIN), AuScope, Bioplatforms Australia, Integrated Marine Observing System (IMOS) and Terrestrial Ecosystem Research Network (TERN), and publicly funded research agencies and institutions.

The Planet RDC will complement and collaborate with existing infrastructure to connect and extend capabilities.

The feedback from the consultations related to the delivery of the Planet RDC can be summarised as follows:

- A co-design process that supports the community to collaborate, identify priority datasets, models and identify infrastructure gaps is important
- The Planet RDC should be interoperable with the ARDC's People and HASS RDCs, and other infrastructures by using data, metadata and technology standards and best practices
- There is a need for alignment between data infrastructure initiatives across NCRIS and government, to reduce overlap of investments in the same types of infrastructure independently by the NCRIS facilities in earth and environment
- Existing infrastructure should be leveraged
- The social infrastructure needed to enable long-term sustainable and interoperable data infrastructure must be considered
- The CARE principles must be implemented, and Indigenous Cultural IP must be incorporated and preserved
- The Planet RDC must be compatible with similar international groups: EarthScope, ENVRI, EPOS, GBIF
- There is a need for consistency in data governance and data sharing arrangements, particularly across different jurisdictions
- Skilled workforce is a key concern–how to attract and retain staff to support these challenges.

The roundtable attendees also suggested certain capabilities and technologies that could be utilised in the commons:

- A sector-wide, agreed reference architecture that supports data storage, retention, availability, modelling and analytics tools
- Supporting common technology solutions by running common/shared services, which are added to as new solutions are found by a partner in the commons
- GPUs, scalable computing, shared file system, large swift storage, ability to scale the compute organically
- Framework for operationalising tools into web services (e.g. an easy to use "Amazon Lambda" for research)
- Better tools for deriving information out of large, integrated datasets (presumably via artificial intelligence, machine learning, etc)
- Establishing innovation pathway to take things from innovation, to product, to commoditised utility
- Ways to identify the primary data source to prevent republication or duplication of datasets
- Integration of geospatial tools and data
- Provide multi-cloud access to data and modelling infrastructure
- New and improved tools for working with novel data types
- Creation of analysis-ready, cloud-optimised data (ARCO)
- Explore new semantic data linkage methods and infrastructure.

REVISED DATA CHALLENGES

Based on the findings from the consultations we have revised and grouped the initial data challenges into 3 streams:

1. Curated, integrated FAIR datasets and services
2. Reliable and trusted data supply chains
3. Modelling and decision support infrastructure.

Figure 5 shows how the initial proposed data challenges fit into these streams.

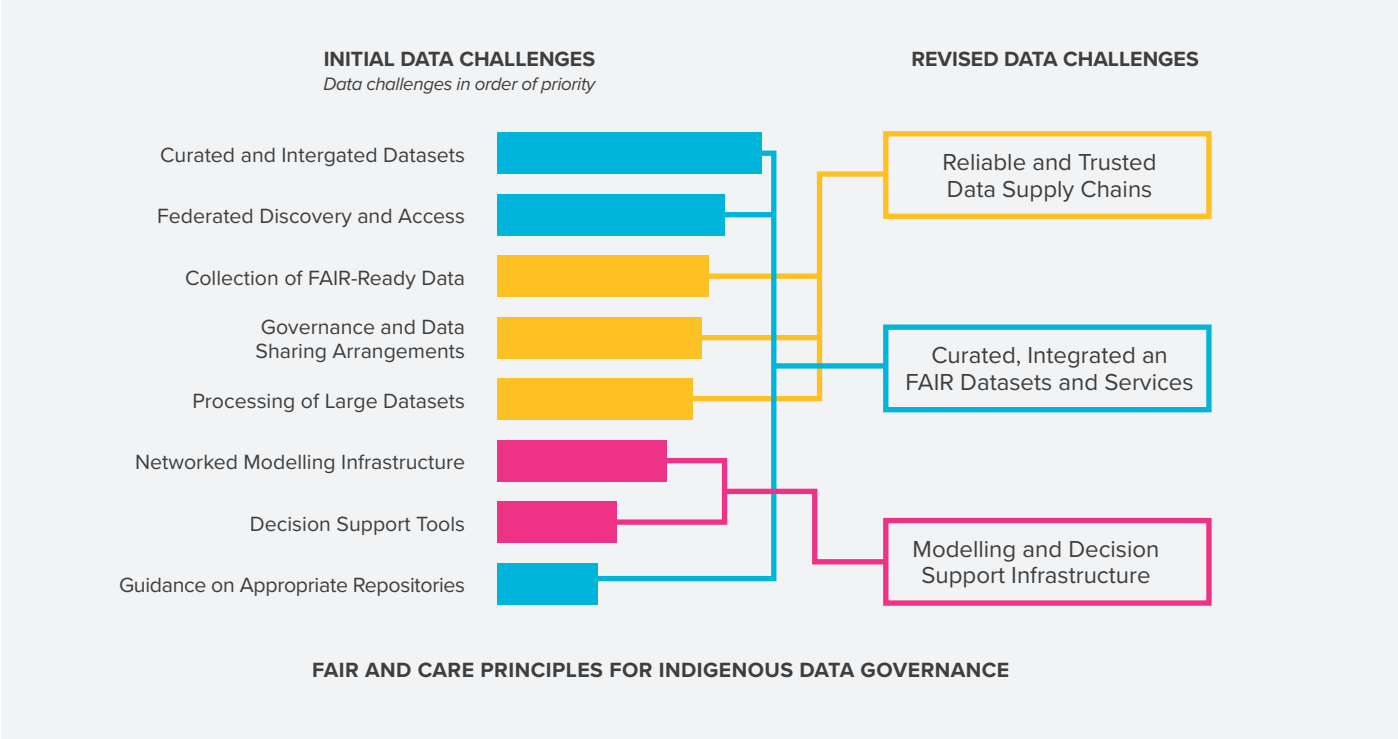


Figure 5: *Revised data challenges*

1. Curated, Integrated FAIR Datasets and Services

Integrated datasets were seen as the highest priority, with integration needed both within and between Planet research domains. Integrated datasets must bring together both geospatial and temporally enabled, and be ready for new analysis methods, including machine learning.

2. Reliable and Trusted Data Supply Chains

Data supply chains encompass FAIR-ready data collection, consistent data governance and sharing, and processing capability for large or streamed datasets. Our consultations showed that government and industry are significant generators of data, but it is difficult for researchers and others to access these data. Government users also have trouble accessing data they can trust to make decisions. There needs to be a two way value flow between research, industry and government. Data quality is a very important attribute for all data users.

There is an opportunity for continuous environmental monitoring from new automated monitoring equipment. However, enduring infrastructure is needed across the earth and environmental science disciplines to manage and process streamed data to turn it into information that researchers and decision makers can use.

3. Modelling and Decision Support Infrastructure

Modelling and decision support infrastructure that allows researchers to utilise and build upon previous models, and enable research to feed into decision making, was considered relevant. This infrastructure should include shared services and reusable platforms, and enable the cheap, rapid and repeatable deployment of tools across all cloud providers.

There was strong support for the implementation of the FAIR principles to enable data reuse. Likewise, the CARE principles must be implemented, and Indigenous knowledge must be incorporated and preserved.

MORE INFORMATION

For more information on the Planet RDC, please visit the ARDC website. For updates on the Planet RDC, please subscribe to our newsletter.

FEEDBACK

We welcome your feedback on this guide. Please email contact@ardc.edu.au with any comments or questions.

ABOUT THE AUSTRALIAN RESEARCH DATA COMMONS

The Australian Research Data Commons (ARDC) enables the Australian research community and industry access to nationally significant, data intensive digital research infrastructure, platforms, skills and collections of high quality data. The ARDC is supported by the Australian Government through the National Collaborative Research Infrastructure Strategy (NCRIS).

GLOSSARY

TERM	DESCRIPTION
2021 Roadmap	2021 National Research Infrastructure Roadmap, outlines national research infrastructure required over the coming decade
ABIS	Australian Biodiversity Information Standard
APPF	Australian Plant Phenomics Facility
ARCO	Analysis-ready, cloud-optimised data
ARDC	Australian Research Data Commons
AURIN	Australian Urban Research Infrastructure Network
CARE principles	Collective Benefit, Authority to Control, Responsibility, Ethics
FAIR principles	Findable, Accessible, Interoperable, Reusable
GPUs	Graphics processing units
IMOS	Integrated Marine Observing System
NCRIS	The Australian Government National Collaborative Research Infrastructure Strategy
NEESFF	National Earth and Environmental Sciences Facilities Forum
NESP	National Environmental Science Program
NGO	Non-government organisation
People RDC	People Research Data Commons for health and biomedical research
Planet RDC	Planet Research Data Commons for earth and environmental research
RDC	Research Data Commons
PFRA	Publicly Funded Research Agency
TERN	Terrestrial Ecosystem Research Network

APPENDIX - PLANET RDC PROPOSED DATA CHALLENGES

The Planet RDC should support both research and decision-making to protect and restore Australia's environment. Relevant data challenges to be addressed in the Planet RDC have been identified through an analysis of the national research and research infrastructure priorities related to environmental research and management. The process followed was analogous to that used for the People RDC. A list of the reports consulted during the analysis process is provided at the end of this document.

This analysis has identified a set of data challenges that are in scope for the Planet RDC to address and an additional overarching challenge. The data challenges have been mapped to the research lifecycle (Table A1).

Essential Attributes

These are a number of recurring statements in the documents reviewed that indicated 'essential attributes' of a Planet RDC:

- FAIR and CARE principles (for Indigenous Data Governance) must be implemented
- Continuity of data storage, access and management is essential
- Indigenous knowledge and data are essential for caring for Country.

These are not specific challenges, but will need to be embedded throughout the activities and outcomes of the Planet RDC.

Proposed Data Challenges for the Planet RDC

These are the identified data challenges that are in scope for the Planet RDC to address:

- Enabling **federated discovery and access**
- Making it easier for researchers to **collect FAIR-ready**[1] **data**
- Enabling **processing of large and/or streamed datasets**
- Facilitating **curated and integrated datasets**
- Developing **networked modelling infrastructure**
- Providing **guidance on appropriate repositories**
- Supporting the use of **decision support** tools
- Implementing **consistent data governance and data sharing arrangements**

1 *FAIR-ready = A digital object that has the attributes of FAIR that can be ascribed to the dataset itself (e.g. is in an open, documented, structured format, uses published ontologies/vocabularies to describe the data elements). Note: a dataset cannot be fully FAIR by itself–it requires delivery infrastructure to make it fully Findable, Accessible, etc.*

Data Challenges Mapped to the Research Lifecycle

To ground the data challenges in research practice, we have taken a generic research lifecycle and mapped each challenge to a particular phase (except for 'data governance and data sharing arrangements,' which is considered an overarching challenge). For each phase we have identified:

- The data challenge
- The elements that might be part of a solution (NOTE: These are examples of elements that could be integrated into a Planet RDC or extended as part of one)
- The impact for a researcher.

Figure A1. *Data in the research lifecycle*

APPENDIX - PLANET RDC PROPOSED DATA CHALLENGES

Table A1. *Proposed data challenges mapped to the research lifecycle*

STAGE - Discovery
DATA CHALLENGE - Federated Discovery and Access

FROM ISSUE TO IMPACT

Challenge: There are many repositories (such as institutional, discipline, government, national) where data can be stored and made accessible for re-use. Finding relevant data or even knowing where to look is a challenge for time-poor researchers. (See also Publishing ➜ Guidance below).

Solution Element(s): The Planet RDC could augment the existing Research Data Australia to federate a wider range of repositories, enabling multiple repositories to be searched as a single step.

Researcher Impact: The ability to search across repositories saves researchers time and increases findability, especially outside their home discipline.

STAGE - Collection
DATA CHALLENGE - Collection of FAIR-Ready Data

FROM ISSUE TO IMPACT

Challenge: Data that is not collected according to an agreed (meta)data standard is harder to make FAIR.

Solution Element(s): Using standard formats and terms, agreed at (ideally international) discipline level, aids interoperability. Digital tools (i.e. apps) for data collection can automate and enable this standardisation. For example, soils data collected by farmers using the Pilot Soils Monitoring Incentive Program (PSMIP) Data App (created by the FAIMS project) is standardised from the time of capture, enabling streamlined submission to the Australian National Soil Information System (ANSIS) database. This approach could be extended to a wider range of field-based domains.

Researcher Impact: Data that is collected in standard formats and described using standard terms avoids wasting researchers' time in having to apply standards to make their data FAIR at the end of the project. This greatly increases the likelihood that data will be published at all. Standard (meta)data schemata improve interoperability and reusability, increasing the likelihood that data will be used outside the collecting discipline.

STAGE - Processing
DATA CHALLENGE - Processing of Large Datasets

FROM ISSUE TO IMPACT

Challenge: Automated monitoring[1] and remotely-sensed[2] data are challenging to process and manage due to their volume, velocity and unstructured nature. Machine learning is often assumed to be the answer to annotating these data, but most researchers do not have the capability to perform this themselves, or the resources to access commercial options. Additionally, there are storage issues to address.

Solution Element(s): The Open Ecoacoustics platform is an example of how to provide data storage, data submission tools, a metadata standard, and algorithms ("Recognisers") to analyse ecoacoustic data, training on how to use the Recognisers, and a portal to share the data and results. This could be extended or used as a model to deal with a wider range of automated monitoring and remote sensing data.

Researcher Impact: Researchers do not need to worry about storage and are more easily able to analyse data without becoming experts in machine learning techniques, and thus focus on the research questions. Extracting meaningful insights from automated monitoring and remote sensing data has the power to enable continental-scale monitoring and management.

1 *Automated photo, video and audio capture of species and ecosystems, and automated sensors capturing environmental variables.*

2 *Remote sensing is the process of detecting and monitoring the physical characteristics of an area by measuring its reflected and emitted radiation at a distance (typically from satellite or aircraft), e.g. photography, infrared, RADAR and LiDAR.*

APPENDIX - PLANET RDC PROPOSED DATA CHALLENGES

STAGE - Integration
DATA CHALLENGE - Curated and Intergated Datasets

FROM ISSUE TO IMPACT

Challenge: There is a great deal of data collected on Australia's ecosystems and issues related to protecting and restoring the environment; however, datasets collected for a particular purpose or in a particular discipline are often hard to integrate together. This can be because of differences in sampling methods, variables studied, units used, in how the objects of study are identified, and in temporal and spatial scales.

Solution Element(s): EcoAssets is an existing example of how to provide open, integrated data delivered by the Atlas of Living Australia (ALA), Integrated Marine Observing System (IMOS) and Terrestrial Ecosystem Research Network (TERN). EcoAssets datasets are publicly available for biodiversity reporting and assessment, covering Australian terrestrial and marine systems. This approach to pre-integration could be extended to a wider range of data providers.

Researcher Impact: Integrated datasets facilitate and encourage interdisciplinary approaches to problems and allow linking of sectors. By making it simpler and faster for researchers and managers to obtain integrated data, EcoAssets will enhance reporting efforts, and support improved environmental decision-making.

STAGE - Analysis and Modelling
DATA CHALLENGE - Networked Modelling Infrastructure

FROM ISSUE TO IMPACT

Challenge: Environmental modelling allows researchers to understand how ecosystems respond to change and future risks and uncertainties, which is essential for environmental management and policy development. But it is often hard to identify the correct models, find compute resources to run them on, and data to input into them, and configure and install the models correctly. Modelling infrastructure that allows models to be shared and reused across domains, and also allows access to compute infrastructure, will create world-class infrastructure to support decision-making, adaptation planning and intervention strategies.

Solution Element(s): EcoCommons provides a user-friendly environment that aims to become the platform of choice for analysing and modelling ecological and environmental challenges. It brings together curated models, relevant data and compute. This is already being extended to Biosecurity (see Decision Support below) and provides a general foundation for application to a wide variety of problem domains.

Researcher Impact: EcoCommons will empower Australian practitioners and researchers to use trusted datasets, modelling tools, and training resources to produce high-quality research that will inform and accelerate evidence-based policy and decision-making for the environment.

APPENDIX - PLANET RDC PROPOSED DATA CHALLENGES

STAGE - Publishing
DATA CHALLENGE - Guidance on Appropriate Repositories

FROM ISSUE TO IMPACT

Challenge: It is often hard to decide on the right repository for a particular dataset, or even what repositories are available. Researchers need clear guidance about what to use, and the pros and cons of the available options.

Solution element(s): The ARDC could develop a national system that guides users to the right repository to deposit their results, based on the ANZSRC-FOR code for their research, the policies of the institution where they are based, and the requirements imposed by the source of their research funding.

Researcher Impact: Researchers have a single location to go to when depositing their data, so the right decision also becomes an easy decision.

STAGE - Reporting / Informing (Translation)
DATA CHALLENGE - Decision Support Tools

FROM ISSUE TO IMPACT

Challenge: Environmental managers and decision-makers increasingly need to rely on complex models and analytics to inform their decision making, but most do not have the specific skills required, nor access to the required data and compute resources. Additionally, for the government and the public to trust the model outputs, and thus in the decisions themselves, the workflows must be repeatable and transparent.

Solution element(s): Biosecurity Commons is developing a solution that will empower researchers and decision-makers to produce consistent and transparent models and analytics without coding experience or high-end IT equipment. The platform will offer users everything they need to collaboratively solve common biosecurity problems, like an intuitive point-and-click web interface, high-performance computing, trusted datasets, repeatable scientific workflows, a secure workspace and cloud storage.

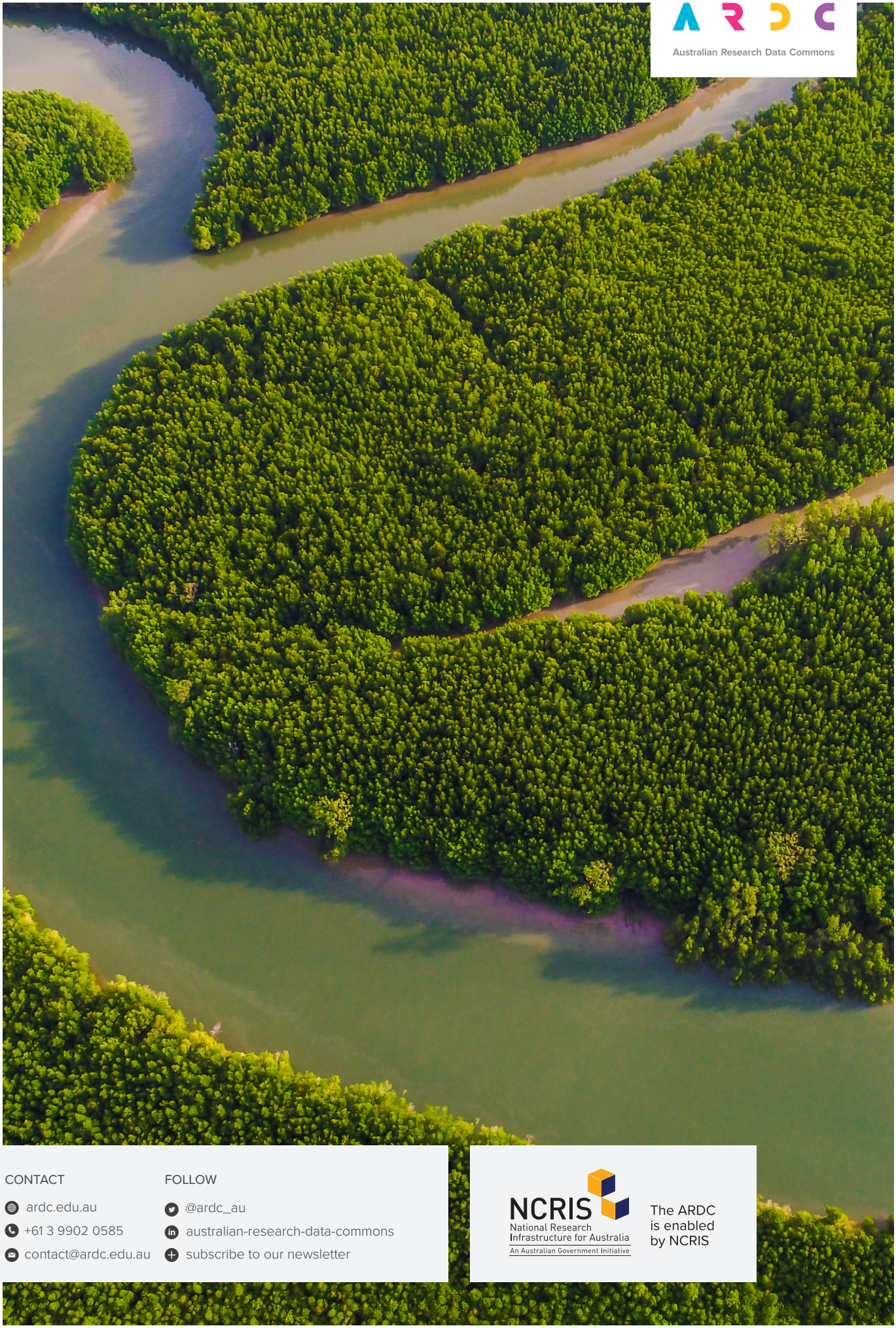
Researcher Impact: Researchers, managers and decision-makers are able to derive the information they need from data, without getting blocked by lack of technical expertise and access to infrastructure, and are able to support their findings with repeatable workflows and trusted datasets.

APPENDIX - PLANET RDC PROPOSED DATA CHALLENGES

The policy and strategy analysis to determine the proposed data challenges for the Planet RDC covered the following documents:

- 2021 National Research Infrastructure Roadmap
- Australia's Data-Enabled Research Future: synthesis report
- National Climate Resilience and Adaptation Strategy
- 2019 Australian National Outlook
- Australia's data-enabled research future: Science
- Australia's Data-Enabled Research Future: Technology and Engineering
- State of the Environment Report 2021
- Discovering Biodiversity: A decadal plan for taxonomy and biosystematics in Australia and New Zealand 2018–2027
- Foundations for the future: A long-term plan for Australian ecosystem science
- TERN Strategic Framework: 2016-2025
- ALA Strategy 2020-2025
- AuScope 10 year Strategy: 2020-2030
- AuScope 5 year Investment Plan
- Decadal Plan for Australian Geoscience
- National Marine Science Plan 2015-2025: Driving the development of Australia's blue economy
- Independent review of the EPBC Act, Chapter 10 - Data, information and systems

Information was also drawn from consultations with Hub Leaders of the National Environmental Science Program.



Australian Research Data Commons

CONTACT

ardc.edu.au
+61 3 9902 0585
contact@ardc.edu.au

FOLLOW

@ardc_au
australian-research-data-commons
subscribe to our newsletter

NCRIS
National Research Infrastructure for Australia
An Australian Government Initiative

The ARDC is enabled by NCRIS