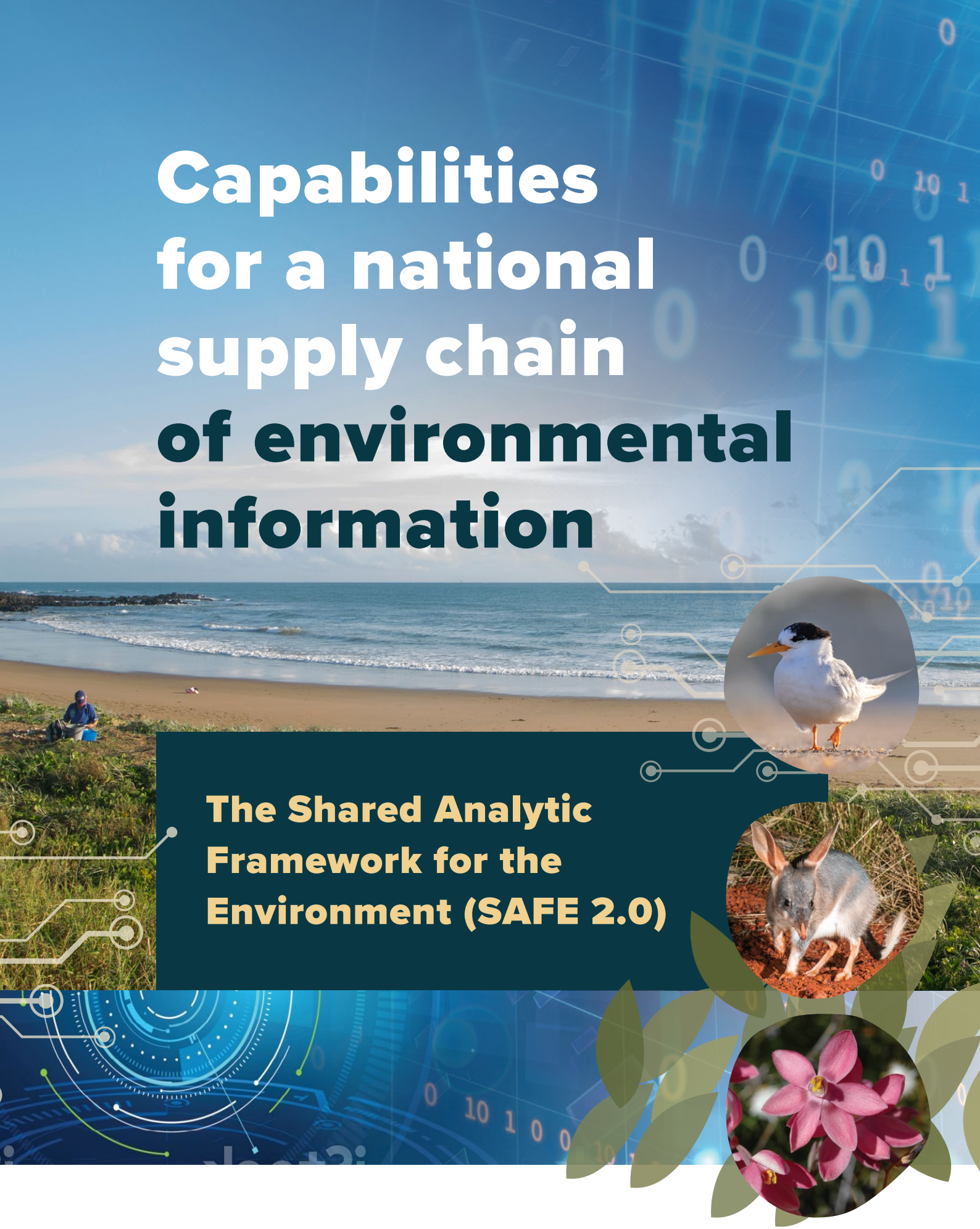
Capabilities for a national supply chain of environmental information



The Shared Analytic Framework for the Environment (SAFE 2.0)

The Western Australian Biodiversity Science Institute
Western Australian Marine Science Institution
Australian Research Data Commons

Acknowledgements

Thank you to our many partners across industry, government, research and community for their contributions through the process of updating SAFE. We specifically acknowledge the work of Chris Gentle (WABSI), Greg Terrill, Hamish Holewa, Kerry Levett and Robin Burgess (ARDC), Simon Ferrier, Lesley Wyborn, Rob Freeth; the support of Owen Nevin (WABSI) and Luke Twomey (WAMSI); and Preeti Castle (WABSI) for assistance with this document.

Photo acknowledgements:

- Lochman Transparencies
- Judy Dunlop
- WA Museum
- Megan Hele
- Preeti Castle
- Renee Young
- Robert McLean
- Claire Greenwell

COVER IMAGES: Claire Greenwell, Lochman Transparencies, Megan Hele

ISBN 978-0-646-87869-0

Citation:

The Western Australian Biodiversity Science Institute and Western Australian Marine Science Institution 2023, *Capabilities for a national supply chain of environmental information SAFE 2.0*, The Western Australian Biodiversity Science Institute, Perth, Australia.

Published May 2023

Acknowledgement of Country

We acknowledge the traditional custodians throughout Australia and their continuing connection to, and deep knowledge of, the land and waters. We pay our respects to Elders both past and present.

WABSI and WAMSI supported by:

GOVERNMENT OF WESTERN AUSTRALIA

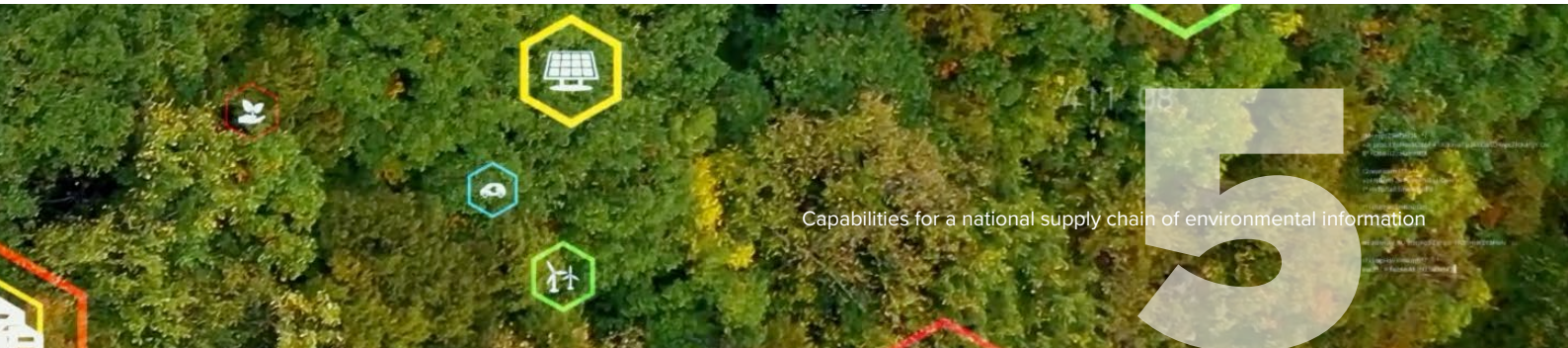
Department of **Jobs, Tourism, Science and Innovation**

ARDC supported by:

NCRIS National Research Infrastructure for Australia
An Australian Government Initiative

The Shared Analytic Framework for the Environment (SAFE) is a means to understanding the complexity of the environmental data and analytics landscape.

Contents



Introduction

The environmental data and analytics landscape is complex and fragmented. Many organisations are involved in the creation, curation, integration and analysis of environmental data and information.[1]

The Western Australian Biodiversity Science Institute (WABSI), the Western Australian Marine Science Institution (WAMSI) and the Australian Research Data Commons (ARDC), along with several partners including the Australian Government and state government, are working on initiatives which need to be integrated and interoperable across multiple domains, organisations and data holdings. To succeed, they require a common view of all elements of the environmental data and analytics landscape.

[1] The terms 'data' and 'information' are used sometimes interchangeably, sometimes with distinct meanings, in different Australian Government contexts. This document does not draw a strong distinction, though 'data' is generally used to numeric and 'information' to word formats.

The updated Shared Analytic Framework for the Environment (SAFE 2.0) presented in this document provides a common vocabulary to show how the many components of the environmental data landscape work together.

It will help understand how the components might best work together to deliver specific reporting and analysis needs; to understand areas where further development is needed; and help identify what is required to deliver a a national supply chain of environmental information.

The Framework is intended to:

- facilitate a consistent view of the required capabilities and their interdependencies across varied stakeholders;
- develop an organisational, maturity and investment view tailored to specific needs;
- individual projects determine the capabilities that they need; and
- align effort, reduce fragmentation, and prioritise investment across the capabilities that support information supply chains.

The Framework is intended to be used by:

- **individuals:** to understand the complexity of the environmental data and analytics landscape;
- **institutions:** to better understand institutional contributions to the national supply chain of environmental information, as well as dependencies between such institutions and their capabilities; and
- **funding bodies:** to promote a consistent, coherent view of a complex landscape, better enabling the mapping of maturity and the planning of investment.

The Framework is primarily focussed upon the environmental data and analytics landscape with its many domains and multiple institutions. It has not been developed for use at a detailed level within a single domain or institution, though it may be of use in some instances.[2] It is of most use when mapping a complex landscape across domains and institutions.

[2] SAFE v2.0 is based upon SAFE v1.0, developed by WABSI, WAMSI and others to accelerate the move to devolved robust, repeatable and transparent decision making by proponents, regulators, Indigenous groups and the community for environmental assessments — https://wabsi.org.au/wp-content/uploads/2021/07/SAFE-Guide-V1.1P.pdf. SAFE v2.0 revises and extends SAFE v1.0.

What is SAFE?

SAFE is a means to understanding the complexity of the environmental data and analytics landscape.

SAFE depicts the capabilities – the building blocks – which work together across the information supply chain to provide data and data-driven decision-support and reporting tools for environmental research, decision-making, management and policy. SAFE helps maintain an overall perspective across the many components of, and dependencies between, elements of the supply chain.

SAFE provides an overview which enables all elements of the environmental data and analytics landscape to be seen in relation to each other.

The matrix:

- depicts capabilities; institutions and individual organisations can be mapped to the layers and boxes, and in some cases will cover several;
- is agnostic as to data complexity, model type or infrastructure scale;
- helps put in context data providers, analytic and other frameworks; and
- is not country-specific. This document references Australian capabilities, though the matrix is generic and could be applied to other national or international contexts.

Capabilities for a national supply chain of environmental information

The need: A national supply chain of environmental information

The environmental data and analytics landscape is very complex, and there is a significant challenge in maintaining a shared overview of its elements.

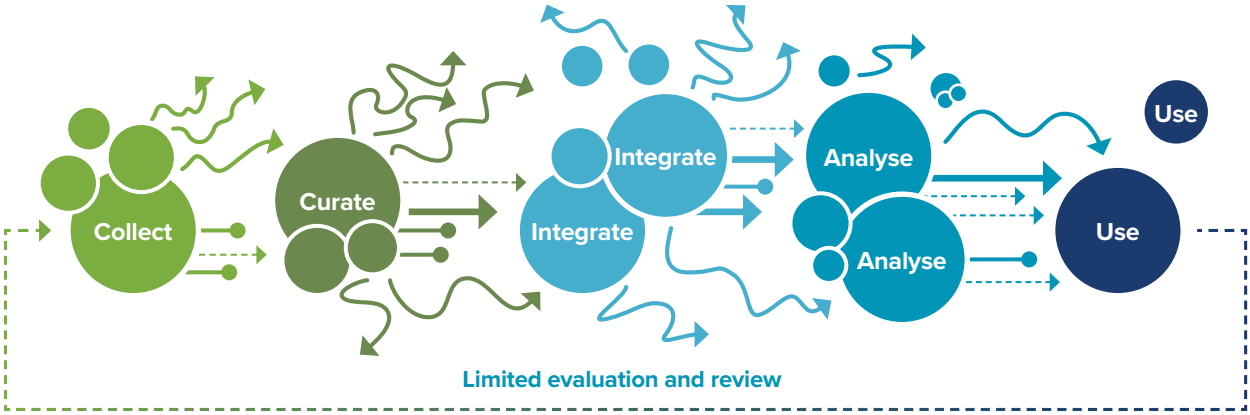
To address the complexity of the environmental data and analytics landscape, the major Independent Review of the Environment Protection and Biodiversity Conservation Act (the primary Australian Government environment legislation) called for the development of a 'national supply chain of information':

- 'A national supply chain of information will deliver the right information at the right time to those who need it. This supply chain should be an easily accessible, authoritative source that the public, proponents and governments can rely on';
- 'The opportunity to derive benefit from a national supply chain for environmental information is broader than just the EPBC Act. While the focus should be on delivering to the National Environmental Standards, incremental effort can provide a supply chain that delivers to the broader national system of environmental management'.[3]

[3] Samuel, G 2020, Independent Review of the EPBC Act—Final Report, Department of Agriculture, Water and the Environment, Canberra, October 2020 p22, p164, https://epbcactreview.environment.gov.au/resources/final-report

One of the challenges in reforming the national environmental information supply chain is maintaining an overview of all the elements which comprise it. SAFE provides a consistent view of the required capabilities and their interdependencies across the varied stakeholders.

SAFE provides a consistent view of capabilities that together constitute a national supply chain of information. Its five layers underpin the depiction in the report of the Independent Review of the EPBC Act of the supply chain[4], enabling a view of both the current state (Figure 1) and the future state (Figure 2):



- Duplicated effort
- Single use
- Survey gaps

- Limited capture and sharing
- Inconsistent quality assurance
- Unclear which data are best to use
- Manual data update

- Multiple bestoke methodologies
- Lack of transparency
- High cost

- Not searchable
- Evidence not used or reused
- Inconsistent, not trusted

FIGURE 1: Current state of the national environmental information supply chain

[4] SAFE v1.0 was used by the Independent review of the EPBC Act to underpin analysis of the national supply chain of information — Samuel, G 2020, Independent Review of the EPBC Act—Final Report, Department of Agriculture, Water and the Environment, Canberra, October 2020 pp22, 165.

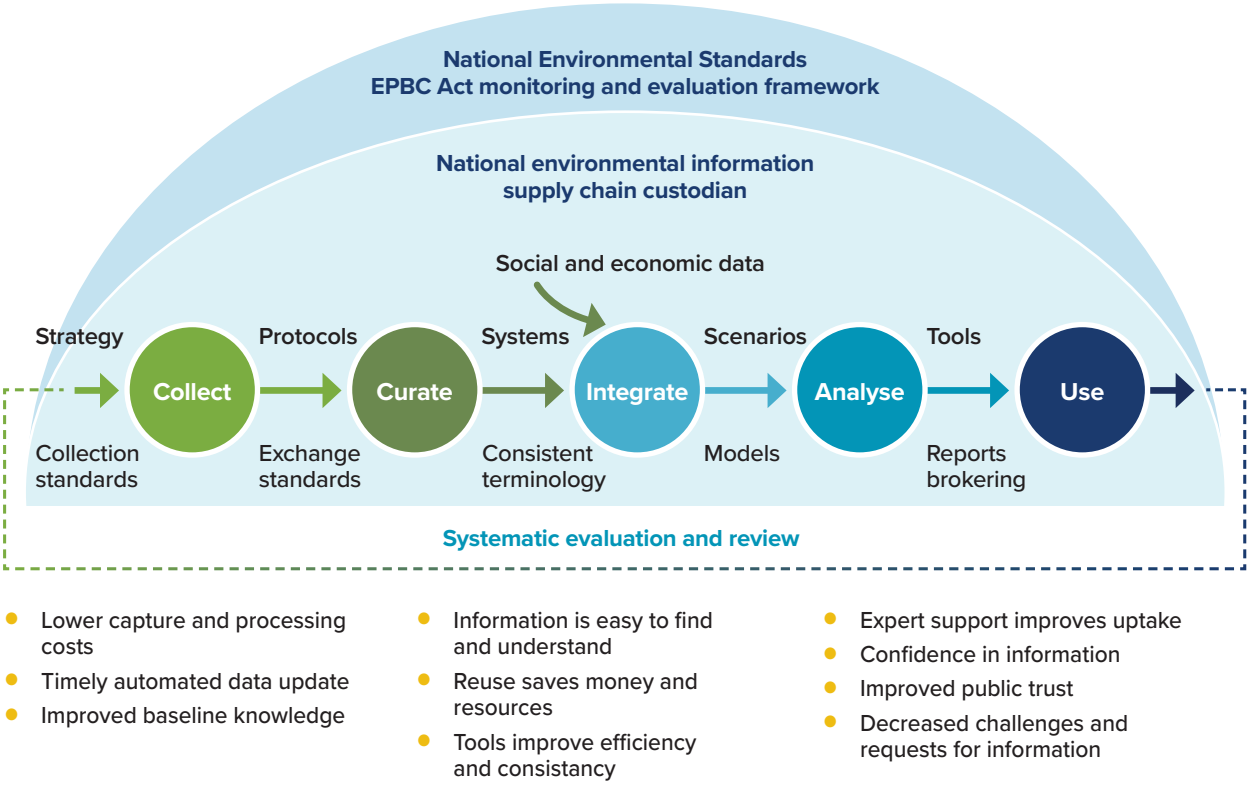


FIGURE 2: Future state of the national environmental information supply chain[5]

A national supply chain of environmental information can lower the cost of businesses, researchers, government entities and the community individually undertaking the intensive tasks of data discovery, curation, and integration for analysis. It is a public good that would assist a range of uses including:

- business and industry analysis and reporting on environmental impacts, eg at the international level under the Taskforce on Nature-related Financial Disclosures (TNFD)[5] and domestically through environmental approval processes;
- researcher access to diverse data sets to undertake, e.g. exploratory modelling of cumulative impacts of proposed developments upon a changing environment;
- reporting on the state of Country and the environment, development of environment-economic accounts; and
- regulation of development, natural resource management and more.

A successful national information supply chain depends upon inclusion of the research sector and research infrastructure — to produce and supply data and derived products, and to improve analytic approaches. The research sector can also benefit from and further develop the data and analytic products developed and used for operational purposes.

[5] https://tnfd.global/. TNFD is one of the major international risk management and disclosure frameworks for organisations to report and act on evolving nature-related risks. It is developing a market-led, science-based framework to enable companies and financial institutions to integrate nature into decision making. References to it are included throughout this document to provide an international perspective on the capabilities covered by the SAFE framework. TNFD represents institutions with over US$20.6 trillion in assets under management and a footprint in over 180 countries. It has been recognised by G7 Finance, Environment and Climate Ministers and through the G20 Sustainable Finance Roadmap.

Case study: TNFD

The Taskforce on Nature-related Financial Disclosures (TNFD) has developed an integrated assessment process for companies to manage nature-related risk and opportunities. LEAP has four phases:

- **Locate** your interface with nature
- **Evaluate** your dependencies and impacts
- **Assess** your risks and opportunities
- **Prepare** to respond to nature-related risks and opportunities and report

LEAP requires companies to access data on environmental matters related to their footprint and activities, including in a regional context. The challenge for companies is:

- navigating the many repositories of data and sources of analytic tools to find those most appropriate for their needs;
- the comprehensiveness, ease of integration and scalability of existing data.[6]

The navigation challenge can be mitigated by mapping data and tools through a simple framework such as SAFE, while the quality and integration issues highlight the value of national or regional chains of environmental information.

[6] Currently, financial institutions and companies don't have the information they need to understand how nature impacts the organisation's immediate financial performance, or the longer-term financial risks that may arise from how the organisation, positively or negatively, impacts nature. Accordingly, TNFD identifies nature-related frameworks, tools, data sources and other guidance throughout the phases of the LEAP approach.

The need for better environmental information supply chains has been noted in many reviews.[7] Environmental managers and policymakers need trusted data supply chains and tools that enable them to make data-driven decisions about land and sea management.

Researchers need access to a wealth of government and industry-held environmental data in an interoperable, machine-understandable form. Cross-sector digital integration at scale is needed to enable researchers to easily discover, access and combine data and natively link to networked modelling and analytics platforms, to answer multi-disciplinary research questions on adapting to climate change, saving threatened species, and reversing ecosystem deterioration.

[7] Multiple reviews as well as national science and environmental policies have emphasised the point - including the State of the Environment Report 2021 (https://soe.dcceew.gov.au/), the Royal Commission into National Natural Disaster Arrangements (the Bushfires Royal Commission, https://naturaldisaster.royalcommission.gov.au/); national science and environmental policies include the National Reconstruction Fund priority areas such as Renewables and low emissions technologies, value-add in agriculture, forestry, and fisheries sector, and Enabling Capabilities (https://www.industry.gov.au/news/national-reconstruction-fund-diversifying-and-transforming-australias-industry-and-economy), as well as the Nature Positive Plan 2022 (https://www.dcceew.gov.au/sites/default/files/documents/nature-positive-plan.docx), National Climate Resilience and Adaptation Strategy 2021-2025 (https://www.dcceew.gov.au/climate-change/policy/adaptation/strategy/ncras-2021-25), Critical Minerals Strategy 2022 (https://www.industry.gov.au/sites/default/files/2022-09/2022-critical-minerals-strategy_0.pdf)

The SAFE matrix – Overview

SAFE has five layers, each of which describes key capabilities that support the environmental information supply chain.

Each layer has several core components (Figure 3), and all layers interconnect and add value to each other. The layers and boxes are conceptual, and in many cases a particular organisation will provide services across more than one layer. Organisations are not individually identified in the framework, although illustrative examples are provided later in the document.

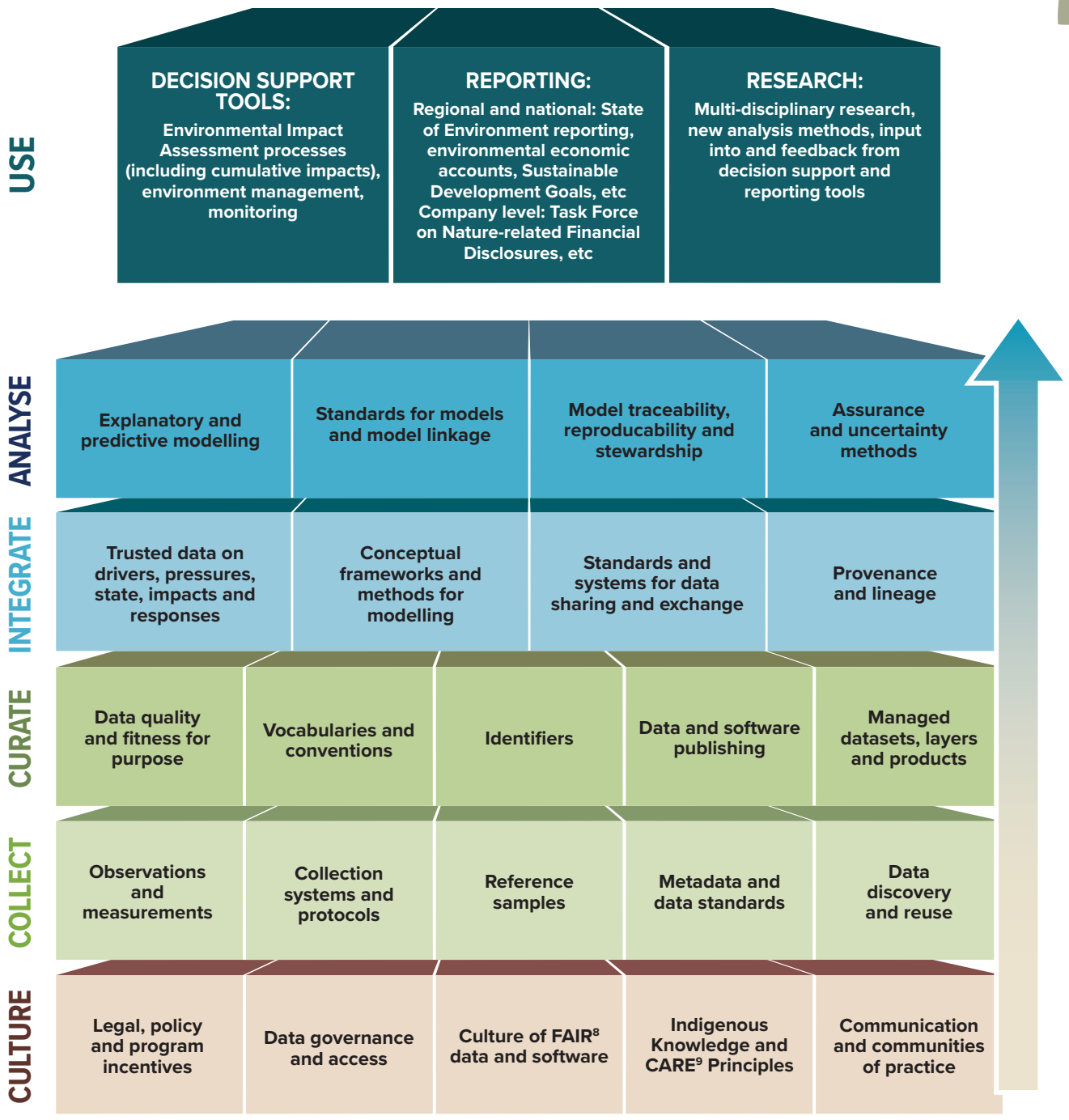


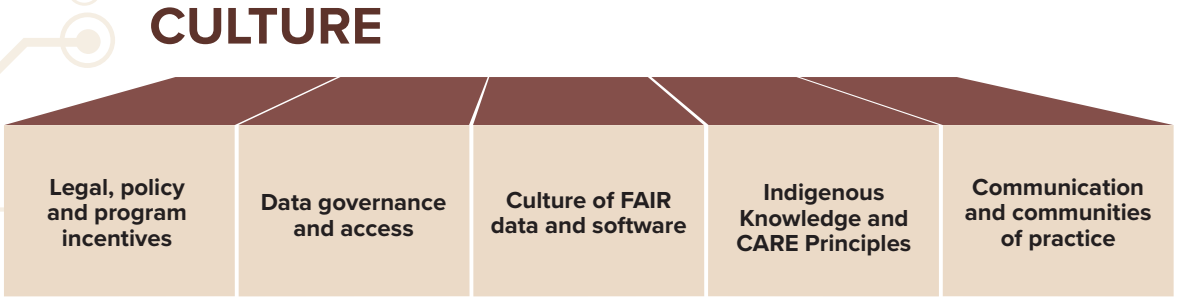
FIGURE 3: SAFE – Layers and capabilities

[8] FAIR – Findable, Accessible, Interoperable and Reusable

[9] CARE – Collective Benefits, Authority to Control, Responsibility, Ethics

The SAFE matrix – Detail

The following section outlines each layer and capability in more detail, providing some illustrative examples.



The Culture layer comprises the fundamental approaches and capabilities needed to enable all elements of SAFE to operate and to interact effectively.

Legal, policy and program incentives

Australia has some world-leading organisations for data collection and curation, for integration and for modelling. While sharing is widespread, a culture of sharing is not universal. There remain concerns that sharing data may lead to unwelcome scrutiny, lack of data control and confidentiality breaches. While policy, legislative and financial frameworks often include requirements to share, these do not fully align across an organisational landscape with many players. At the same time, there are also legal requirements to protect some aspects of data, including privacy, IP and contractual agreements.

Public policies, legislation and funding initiatives need to reinforce accessibility, interoperability and reuse of data and information products, as well as the application of standards and licensing to support interoperability and reuse. This applies to the system, as well as within individual programs, capabilities and projects. For example, all monitoring contracts and permits should require contractors to comply with requirements for public open data.

Data governance and access

SAFE encompasses a large range of capabilities and organisations, both public and private.

No single governance approach exists to cover all aspects; a network of governance approaches is therefore needed. Governance, as a supporting capability, will apply within each layer of SAFE (e.g. various naming standards and definitions related to data), and across the SAFE layers (e.g. routine means for data and data products to link to models, as well as for models to link to each other). Specific data governance considerations include ethics, privacy, access protocols and consent.

Ensuring that sensitive data (e.g. Indigenous data and knowledge, threatened species data and commercial-in-confidence information) can be appropriately and securely shared, managed and analysed is a critical requirement to support decision-making and reporting for environmental assessments, as well as furthering research and management. This requires both secure technical systems and strong governance.

Data sharing often takes place according to legal agreements between a custodian and a recipient. These can be effective means to manage the risk of unwanted release, though may also be time-consuming to finalise.

> ***Systems that track lineage and provide an audit log of actions can also facilitate data sharing through mitigating risk and providing reassurance to the data custodian that data have been used appropriately.***

The challenges extend beyond legal frameworks to include culture.

Culture of FAIR data and software

Data are expensive to collect and curate, and this cost should be incurred only once, while the benefits are derived many times. A commonly used framework for enabling effective data and software sharing is the FAIR[10] principles — that is, data and software should be Findable, Accessible, Interoperable and Reusable - encompassing aspects of standards, identifiers, vocabularies and licensing.

Making data FAIR requires a combination of technical and cultural changes, including the use of standardised metadata and data formats, data management practices, and a culture of data sharing and reuse.

Some environmental data is readily findable, though much is fragmented over many domains and repositories and there is no common discovery mechanism. Data is held in dispersed repositories and existing discovery mechanisms are largely structured by domain or by data creator. The importance of public metadata is critical to enabling data to be found.

> ***Data and software should be as openly accessible as possible, and be stored in a stable and secure location. Access controls should be limited, particularly where data has been gathered using public funding.***

Data sharing agreements and authentication and authorisation protocols may be needed to protect sensitive data or knowledge products.

Interoperability requires participants to adopt agreed international approaches and standards so that data, models or knowledge products can readily be used by other systems and people. This is particularly important when integrating elements from a range of sources, as is necessary in Australia's current fragmented environmental information landscape.

Reuse is a transparency and efficiency principle. Enabling reuse requires both cultural and technical effort, and is dependent upon agreed standards and frameworks at all levels. To make data reusable, it should be documented clearly, including on how it can be cited and reused.

FAIR implementation profiles can be used to catalyse convergence between capabilities and stakeholders, as they clearly articulate the FAIR implementation choices made by a community of practice for each of the FAIR Principles. These have been implemented across many environmental domains in the WorldFAIR project and were used by ENVRI-FAIR to integrate datasets over many environmental data types.

[10] https://www.go-fair.org/fair-principles/

Indigenous knowledge and CARE principles

Irreplaceable knowledge is held, and continually developed, by Australia's First Nations people. Where appropriate, some of this has been integrated into management practice, for instance through Indigenous ranger and Caring for Country programs. Best practice has Indigenous engagement and knowledge built into management approaches (Figure 4), based upon a clear sense of value to the Indigenous people involved, as well as long term ongoing engagement.

The Indigenous Data Network (IDN)[11] was established to support and coordinate the governance of Indigenous data for Aboriginal and Torres Strait Islander peoples and empower Aboriginal and Torres Strait Islander communities to decide their own local data priorities. The IDN works to engage with and leverage internationally leading developments in the data sciences to maximise the optimal collection, access and use of data resources for community empowerment.[12]

The *Our Knowledge Our Way* guidelines are best practice guidelines for working with Indigenous knowledge in land and sea management, developed under the Australian Government's National Environmental Science Program.[13] The *Our Knowledge Our Way* guidelines give voice to Indigenous land and sea managers who have found good ways to strengthen their knowledge and build partnerships for knowledge sharing in caring for Country.

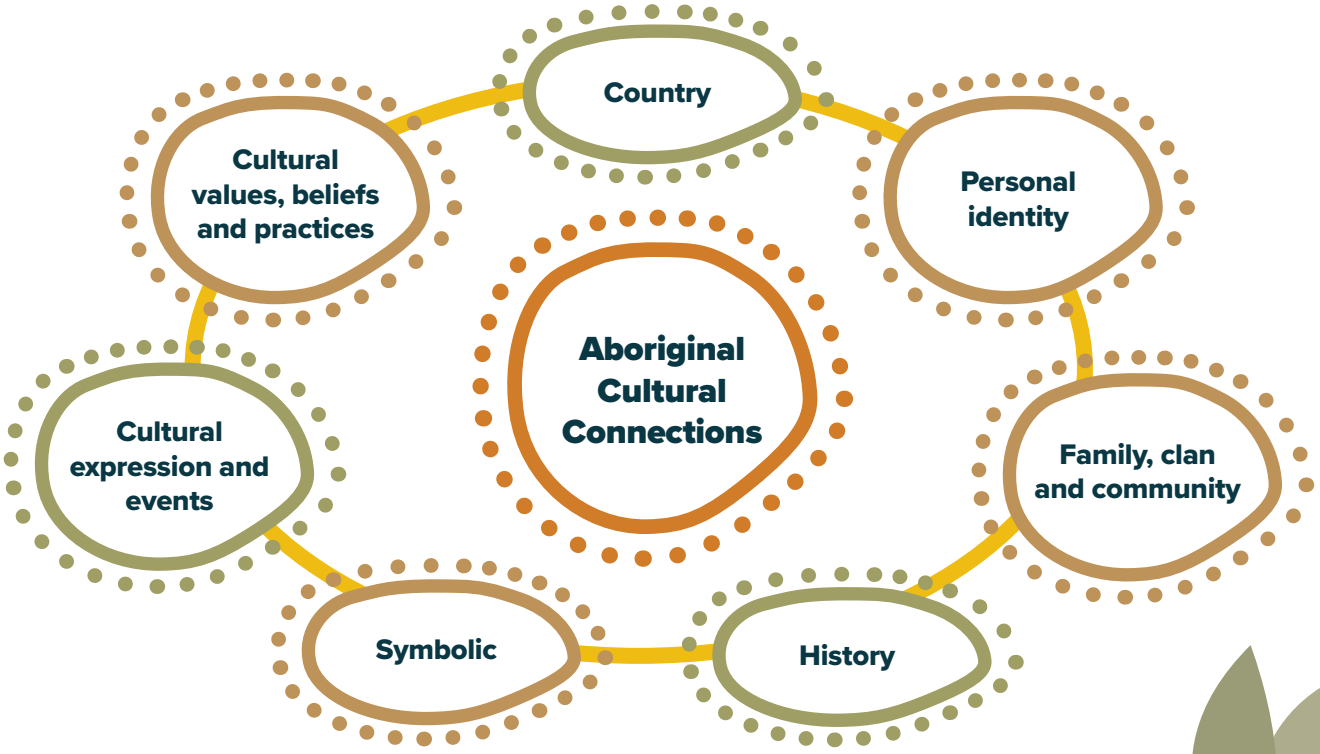


FIGURE 4: Considerations when engaging with First Nations communities[14]

[11] https://mspgh.unimelb.edu.au/centres-institutes/centre-for-health-equity/research-group/indigenous-data-network

[12] https://mspgh.unimelb.edu.au/centres-institutes/centre-for-health-equity/research-group/indigenous-data-network#about-us

[13] https://www.csiro.au/en/research/indigenous-science/Indigenous-knowledge/Our-Knowledge-Our-Way

[14] https://vpsc.vic.gov.au/html-resources/aboriginal-cultural-capability-toolkit/aboriginal-culture-history/

While the FAIR principles focus on the data and its management, the CARE principles for Indigenous Data Governance support the consideration of people and purpose within the process of data sharing, ensuring collective benefit and self-determination are considered.[15]

The **CARE Principles** are a tool to help understand what needs to be considered in association with the management of Indigenous components of research:

- **Collective Benefits:** Ensuring there are benefits to Indigenous people from the collection and analysis of data.
- **Authority to Control:** Indigenous people have the authority to control the access to the data in accordance with their values.
- **Responsibility:** If working with Indigenous data, you are responsible for sharing how the data are used to support Indigenous people.
- **Ethics:** The collection of data should minimise harm to Indigenous people, and bring about benefits from the collection and analysis of the data.

The CARE principles are interlinked and can be seen as an approach to support Indigenous people's self-determination and collective benefit. Further information on the application of the CARE principles to biodiversity information is in the Appendix.

Traditional Knowledge and Biocultural Notices and Labels

Traditional Knowledge and Biocultural Notices and Labels are developed by Indigenous communities and local organisations to allow the communities to express local and specific conditions for the sharing of data that adheres to existing community rules, governance and protocols. Traditional Knowledge and Biocultural Notices and Labels are visible digital identifiers that are applied to data and analysis to recognise cultural considerations and responsibilities for material.

[15] https://static1.squarespace.com/static/5d3799de845604000199cd24/t/6397b363b502ff481fce6baf/1670886246948/CARE%2BPrinciples_One%2BPagers%2BFINAL_Oct_17_2019.pdf

A Biocultural Notice recognises the 'rights of Indigenous peoples to define the use of information, collections, data and digital sequence information generated from the biodiversity and genetic resources associated with their traditional lands, waters, and territories'[16]. Biocultural Labels define community expectations about appropriate use of biocultural collections and data and focus on accurate provenance, transparency and integrity in research engagements with Indigenous communities. Biocultural Labels ensure Indigenous people are represented in the metadata and create opportunities for future researchers to connect and support appropriate benefit sharing.

Traditional Knowledge Notices 'should be used to recognise that place-based knowledge carries accompanying cultural rights and responsibilities and that appropriate permissions may need to be sought for future use of this material'[17]. Traditional Knowledge Labels support the inclusion of local protocols for access and use to cultural heritage that is digitally circulating outside community contexts.

> ***Traditional Knowledge Labels identify and clarify community-specific rules and responsibilities regarding access and future use of traditional knowledge.***

This includes sacred and/or ceremonial material, material that has gender restrictions, seasonal conditions of use and/or materials specifically designed for outreach purposes.

Skills and communities of practice

Aside from domain expertise in earth and environmental sciences, a number of core skills will be needed to manage a national supply chain of information, including:

- data governance, curation and analysis;
- information management: the management of digital information, including file organisation, storage, and retrieval;
- supply chain management, to monitor the flow of information through the supply chain and identify opportunities for improvement; and
- IT, platform, cloud, High Performance Computing (HPC), security and access controls.

National and international communities of practice exist within individual capabilities and layers, as well as across layers, and are important in sharing and converging on best practices. Some communities are mature (e.g. for many data capabilities), while others are evolving (e.g. to assure decision support tools, or multi domain model development and integration). There is scope for communities of practice in many areas including data governance and management, digital transformation, open data, and the ethics of information access and use.

[16] https://localcontexts.org/notices/aboutnotices/

[17] https://localcontexts.org/notices/aboutnotices/

Further sources: Culture

Legal, policy and program incentives:

- Taskforce on Nature-related Financial Disclosures (TNFD) – tnfd.global/
- EPBC Act – www.environment.gov.au/epbc
- WA Environment Protection Act – www.legislation.wa.gov.au/legislation/statutes.nsf/main_mrtitle_304_homepage.html
- National Environmental Science Program – www.environment.gov.au/science/nesp
- Australian Privacy Principles – www.oaic.gov.au/privacy/australian-privacy-principles

Data governance and access:

- Sensitive Data – ardc.edu.au/resource/sensitive-data/
- Data Sharing Agreements Guidelines – ardc.edu.au/resource/data-sharing-agreement-development-guidelines/
- Data Sharing Policy Guidelines – ardc.edu.au/resource/data-sharing-policy-development-guidelines/
- Information Security Manual – www.cyber.gov.au/acsc/view-all-content/ism

Culture of FAIR data and software:

- FAIR principles – ardc.edu.au/resource/fair-data/
- Sharing software – ardc.edu.au/resource-hub/working-with-research-software/
- FAIR Implementation Profiles – www.go-fair.org/how-to-go-fair/fair-implementation-profile
- WorldFAIR project – worldfair-project.eu/
- ENVRI-FAIR – envri.eu/home-envri-fair

CARE principles and Indigenous knowledge:

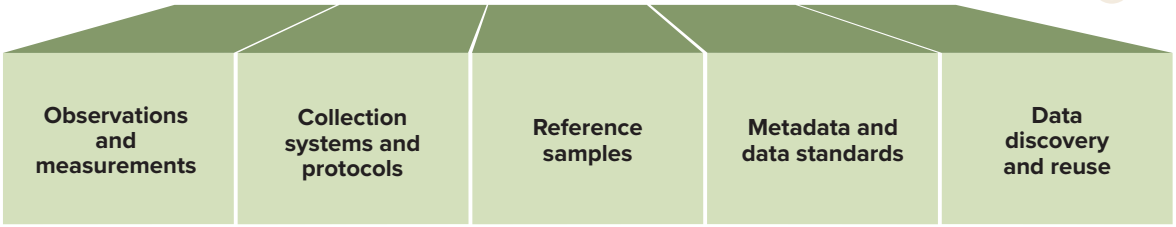
- CARE Principles for Indigenous Data Governance – www.gida-global.org/care
- Traditional Knowledge and Biocultural Notices and Labels – localcontexts.org/
- Indigenous Data Network – mspgh.unimelb.edu.au/centres-institutes/centre-for-health-equity/research-group/indigenous-data-network
- ARDC CARE Principles – ardc.edu.au/resource/the-care-principles/
- National Indigenous Australians Agency – www.niaa.gov.au/indigenous-affairs/environment
- CSIRO – www.csiro.au/en/Research/LWF/Areas/Pathways/Sustainable-Indigenous/Our-Knowledge-Our-Way
- Maiam nayri Wingara Aboriginal and Torres Strait Islander Data Sovereignty Collective – www.maiamnayriwingara.org
- Aboriginal cultural capability toolkit – vpsc.vic.gov.au/html-resources/aboriginal-cultural-capability-toolkit/aboriginal-culture-history/
- TNFD – framework.tnfd.global/additional-guidance/stakeholder-engagement/
- Task Force on Inequality-related Financial Disclosures (TIFD) – thetifd.org/

Skills and communities of practice:

- Research Data Alliance – rd-alliance.org/
- ARDC Communities – ardc.edu.au/get-involved/communities-and-groups/
- DTA communities of practice – www.dta.gov.au/help-and-advice/communities-practice

Capabilities for a national supply chain of environmental information

COLLECT



The Collect layer includes the capabilities to generate multiple types of data, from existing sources to new fieldwork observations and automated sensors.

Prior to the collection of new data, planning will normally involve assessment of whether similar data already exists. In Australia's fragmented environmental data landscape, this is not always easily determined.

Observations and measurements

This capability includes the primary data from observations made on the physical world. These are outputs can come from many providers, including in relation to:

- geospatial, satellite, drone and other remote earth observation technologies
- soils and geomorphology
- geology, geochemistry, geophysics
- hydrology
- atmospheric, including meteorological and climate
- marine and coastal, including ecology, biology, and genetics
- landscape and terrestrial, including ecology, biology, and genetics

Many of these observations and measurements are made in Australia. Some will be sourced internationally. There are rapidly emerging technologies for more automated collection and recognition of environmental characteristics, including environmental DNA sampling, automated species recognition software, and environmental sensors.

Collection systems and protocols

Data collection is often expensive and there are many ways to collect data, including traditional ecological field assessments, as well as newer technology involving field assessment mobile apps, and remote capture through cameras and other sensors. Systems and methods are rapidly evolving but the need to collect and curate data remains common – including interoperability across collection platforms, and comparability across time.

There are often trade-offs between survey methods, survey detail, and the range of subsequent uses. Different methodologies may cause a lack of comparability among datasets, in particular with national and regional datasets as well as with international protocols. In addition, it can be difficult to compare data collected at site level with that at larger scales.

Quantitative or sampling-event datasets typically derive from standard protocols for measuring and monitoring biodiversity such as vegetation transects, animal, bird and marine species censuses. There are also standard protocols for sampling soils and water and taking geological observations and measurements. Efforts are already underway for a national approach for marine data protocols.[18]

> ***Using standard protocols improves comparisons with data collected using the same protocols at different times and places, and allows for more accurate analysis of the data across time and space. Standard protocols should be used for all forms of collection, whether of specimens and samples, or by instruments.***

If standard protocols are not used, the methodology needs to be carefully considered and statistically designed to ensure it is fit for the required purpose.

[18] Establishing and supporting a national marine baselines and monitoring program: Advice from the Marine Baselines and Monitoring Working Group. https://www.marinescience.net.au/wp-content/uploads/2023/02/NMSC_TECH_REPORT_Marine_Baselines_FINAL.pdf

Reference samples

Reference samples are physical specimens or samples of biodiversity and geology that are used as a standard or a point of comparison for scientific research and analysis. These samples may include rocks, minerals, fossils, plants, and animal specimens that are curated and preserved for future reference. Reference samples include:

- specimens maintained in biological collections include the material samples on which new species are described – the type specimens – and also additional specimens that represent the variety and variability that support species identification. Collections are essential for taxonomic and systematic research, identification and naming; and
- collections of soils and geological materials such as drill core samples.

Reference samples may be housed in physical collections, such as museums, herbaria, national research infrastructures and other collection institutions. Many of these collections are now digitised.

Identifiers are needed here for both the samples that are collected and the 'feature of interest' that is sampled. Unique identifiers such as a barcode or unique number are often assigned to reference samples. This identifier should be linked to information about the sample, such as the sample source, date collected, and any other relevant information.

For the 'feature of interest' that is being sampled, it is important to use a unique identifier to keep track of the specific feature being analysed. This could be in the form of a gene ID, protein ID, metabolite ID, or other identifier that is specific to the feature being analysed. This identifier should be linked to information about the feature, such as its function, properties, and any relevant experimental conditions.

Using unique identifiers for both the samples and the features being analysed will help ensure that the data can be easily tracked, managed, and shared with others in a consistent and standardised manner.

Metadata and data standards

Metadata and data standards follow on from standard sampling protocols as the next step in providing interoperable and reusable data. Collecting data in a standardised form right from the start makes it easier to make data interoperable later.

Data standards or models are the rules by which data are recorded and described.

Metadata is information about a physical or digital object or dataset that describes characteristics such as content, format, location, temporal, quality, and access information. It also conveys how the data were created, the scale of the data, any cleaning, processing or validation processes applied to the data, and whether there are any restrictions that apply to the data. Metadata is an essential component of data quality and is important to enable decisions regarding fitness for further use. Metadata schemas specify a set of metadata concepts or terms, as well as their definitions and relationships.

There are many metadata schemas and data models in use, and best practice is to use an international or national standard if one exists for a domain. Standardising formats and meanings makes it easier to share, exchange, understand and use data. Metadata and data standards support data interoperability, processing and management, and are a key component of FAIR, particularly for machine readability.

Some standards (such as units of measurement) are of importance to many research domains, while other standards are domain-specific. And while some standards go through rigorous formal processes (such as ISO or W3C[19]), others may be conventions that are developed and adopted by a research community.

A research community may create a profile of a standard in order to better meet their needs. For example, they may develop a subset of a standard, or an extension to a standard. This enables a research community to maintain interoperability with the core of the standard, while also allowing for what may be specific to that community.

The ARDC recommends reuse of existing international and community-endorsed data and metadata standards, extending them where necessary.

Data discovery and reuse

Data discovery refers to the process of identifying relevant data sources for a specific use case, while data reuse involves leveraging existing data for new use cases or applications.

In an information supply chain, data discovery is typically the first step. It involves identifying relevant data sources, such as databases, data warehouses, or external data providers, and then extracting and preparing the data for analysis. Data discovery can be a time-consuming process, particularly when dealing with large and complex data sets, but it is critical for ensuring the accuracy and relevance of the information produced.

There are a number of international and Australian environmental sciences data catalogues where existing data can be discovered for reuse. For example, The Catalogue of Life is the most comprehensive and authoritative global index of species, holding information on the names, relationships and distributions of over 1.8 million species.

Data reuse involves leveraging existing data for new applications or use cases. This can include repurposing data that was previously used for a different business process, or combining data from multiple sources to gain new insights. Data reuse can significantly reduce the time and cost associated with data analysis, as well as increase the efficiency and effectiveness of the information supply chain.

In Australia, there are policies requiring data created using public funds to be made public.[20] While this is often the case, and there are some useful national and state level data aggregators, much research data can be difficult to find. A challenge for data access is the curation and management of longitudinal environmental data over decades as software, data systems and computational infrastructures evolve over time.

[19] https://www.iso.org/ and https://www.w3.org/

[20] https://www.arc.gov.au/about-arc/strategies/research-data-management

Capabilities for a national supply chain of environmental information

Further sources: Collect

Observations and measurements

- Australian Government authority on measurement – www.industry.gov.au/policies-andinitiatives/national-measurement-institute
- Geoscience Australia – www.ga.gov.au
- Bureau of Meteorology – www.bom.gov.au
- Atlas of Living Australia (ALA) – ala.org.au
- Terrestrial Ecosystem Research Network (TERN) – www.tern.org.au
- Integrated Marine Observing System (IMOS) – www.imos.org.au
- AuScope – www.auscope.org.au
- CSIRO – www.csiro.au, adaptnrm.csiro.au/biodiversity-options/ (and much more)
- Global Biodiversity Information Facility (GBIF) – www.gbif.org

Collection systems and protocols:

- Terrestrial Ecosystem Research Network (TERN) – www.tern.org.au
- Index of Biodiversity Surveys for Assessments (IBSA) – www.wa.gov.au/service/environment/environmental-impact-assessment/program-index-of-biodiversity-surveys-assessments
- Index of Marine Surveys for Assessments (IMSA) – www.epa.wa.gov.au/forms-templates/instructions-for-preparing-data-packages-for-the-index-of-marine-surveys-for-assessments-imsa
- National Soils Standards – ansis.net/standards/
- IMOS Community Practices and Protocols – repository.oceanbestpractices.org/handle/11329/556

Reference samples:

- Australasian Virtual Herbarium – avh.chah.org.au
- Australian Reference Genome Atlas – www.arga.org.au

- Geoscience Australia – www.ga.gov.au
- National Soil Archive – www.csiro.au/en/Do-business/Services/Enviro/Soil-archive and www.asris.csiro.au
- TERN Australia Soil and Herbarium Collection – www.tern.org.au/field-sample-library/
- National Core Virtual Library – www.auscope.org.au/nvcl

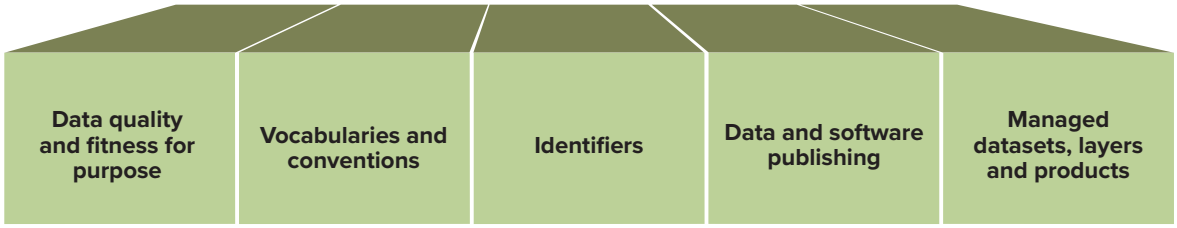
Metadata and data standards:

- Data standards – ardc.edu.au/resource/community-endorsed-data-standards/
- Good data practices – ardc.edu.au/resource/good-data-practices/
- Metadata – ardc.edu.au/resource/metadata/

Data discovery and reuse:

- Australian Ocean Data Network – www.portal.aodn.org.au
- Atlas of Living Australia – www.ala.org.au
- Data.gov.au – www.data.gov.au
- Research Data Australia – researchdata.edu.au/
- CSIRO Data Access Portal – data.csiro.au
- Global Biodiversity Information Facility – www.gbif.org
- Catalogue of Life – www.catalogueoflife.org
- Subject specific repositories – www.re3data.org
- Dryad – datadryad.org/stash
- Figshare – figshare.com/
- National Data Commissioner – www.datacommissioner.gov.au/
- National and state level data aggregators – data.gov.au/data/, www.data.vic.gov.au/, https://www.data.qld.gov.au/dataset, www.data.wa.gov.au/, data.sa.gov.au/, www.data.act.gov.au/, data.nsw.gov.au/

CURATE



The Curate level is the engine room where data are processed to make them fit for purpose, complete and FAIR. Data curation is an active and ongoing process that covers the full data lifecycle.

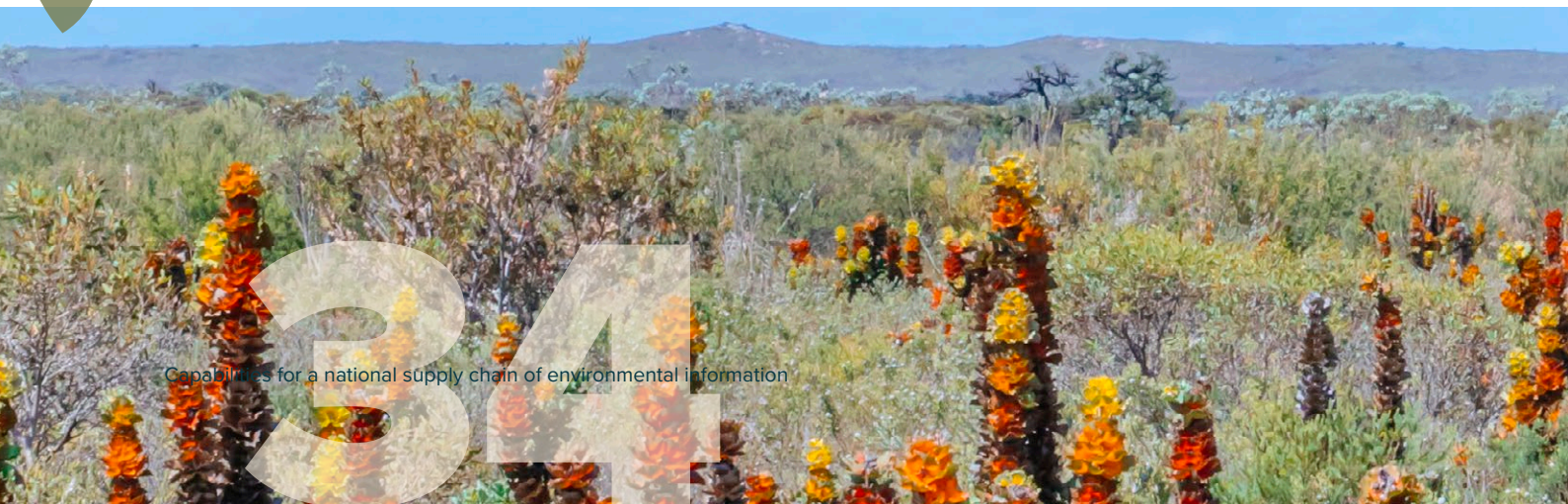
The result can be a dataset that differs in structure and form from the original data. Organising data in forms that can support analysis and modelling should increasingly take place through automated processing, based upon naming frameworks, structures and standards.

There are varying views of 'big data'. For some types of data there are powerful means emerging to automate curation, ingesting large amounts of data of different types, and enabling it to be used even if poorly structured or defined. Other types of data, including species observations data, have to date proved less amenable to such treatment and require manual, often expert, curation. Both can contribute. In some circumstances, standardised, long-term datasets are critical to validating and calibrating approaches that use large, unstructured datasets; and large unstructured datasets can sometimes extend conclusions that might be drawn from standardised, longer term datasets.

Data quality and fitness for purpose

In the case of environmental assessments, data contributes to investment planning, regulatory decisions and public trust, and these impose high requirements for data quality.

There is no simple agreed definition of data quality, and it differs depending upon whether the data are from samples, observations and measurements, or are statistical or derived from modelling. What is important is that the data are fit for purpose. Therefore, it is important to clearly articulate the data quality attributes in the dataset metadata, so that a user can make an informed judgement as to the suitability for their purpose. Clear information about data quality enables decisions to be made about how different forms of uncertainty can be propagated through analytics to provide the end user with overall estimates of uncertainty.



Data quality management is a process where protocols and methods are employed to ensure that data are properly collected, handled, processed, used, and maintained at all stages of the data lifecycle. Below is an example of a data management lifecycle.

There are international efforts underway to standardise the way quality information is expressed.

Data quality can be improved through annotations and corrections, using human and automated tools to correct and annotate individual data elements, so that annotations become visible to researchers who subsequently access the data. Annotations often take the form of metadata, whereas corrections modify the original data record. Annotations should be tied to the original data.

Annotations need to be transparent, and their provenance traceable. Tools have now been developed for online annotations and corrections to be associated with digital records.

Vocabularies and conventions

A vocabulary sets out the common language a discipline has agreed to use to refer to concepts of interest. For example, vocabularies, conventions and other knowledge organisation systems enable interoperability by ensuring that both machines and humans can interpret and use data arising from multiple sources. Agreed vocabularies are important to enable efficient collaboration to occur.

Vocabularies should be made available by an online vocabulary service (e.g. Research Vocabularies Australia, NERC Vocabularies), with a resolvable identifier for each concept. For example, a generic spreadsheet to capture tabular data can use controlled vocabularies within cells, with a reference to the vocabulary source included in the column header. This unambiguously defines values, for both humans and machines.

One of the current challenges for biodiversity data is that while there are broadly shared vocabularies, there are many exceptions. There are also important gaps, for example in relation to descriptions of pressures or threats to the environment.

Reference data is a form of metadata similar to a vocabulary, and is used to classify or categorise other data, e.g. units of measurement or calendar structures.

Identifiers

Persistent identifiers (PIDs) are a core component of a national information infrastructure and key to world-class research and innovation.

Identifiers are used in all computer-based systems to identify and retrieve datasets and software, and to connect data with related resources to enhance data discovery. Identifiers enable the tracking of important provenance information about data and the resulting models and decisions. By linking scientific concepts across systems, they enable interoperability, research innovation and efficiency.

There are various globally-unique persistent identifiers that can be used by government, research and industry, including:

- Digital Object Identifiers (DOIs) for data, software and workflows
- Open Researcher and Contributor ID (ORCiD) for people
- International Generic Sample Number (IGSN) for physical samples
- Research Organisation Registry (ROR) for organisations
- Research Activity Identifier (RAiD) for research projects

Data and software publishing

Providing long term access to data and software for assurance and reuse is a common problem with products developed by shorter term activities (e.g. surveys for environmental assessments, or research projects), or where incentives are weak. Too often, data and model resources disappear, go offline or change protocols, making any systems built on them unreliable and costly to maintain.

Publishing data and software is a way of providing long-term access by depositing them in a trusted data or software repository, and providing an appropriate licence and descriptive metadata to aid their reuse. Assigning a persistent identifier such as a digital object identifier (DOI) is important to ensure the longer term stability of references to particular datasets or software versions.

Australia has a fragmented data repository ecosystem, composed of institutional (i.e. University, CSIRO), government (e.g. data.gov.au), generalist (e.g. Figshare, Dryad, Zenodo), and domain specific (e.g. TERN and AODN portals) portals and repositories, and it can be difficult to know where to publish data, and for others to discover it. There are also subject specific repositories. Generalist and Institutional repositories do not offer domain specific QA/QC.

Minimally-processed large data streams that are often required by researchers require specialist storage and access, such as the European Copernicus satellite data available via NCI.

There are not yet comparable national repositories for model code and outputs, though public options such as GitHub and Zenodo are commonly used.

Managed datasets, layers and products

These terms refer to different levels of abstraction for data that can be used for different purposes within the information supply chain.

A managed dataset is a collection of related data that is stored and managed as a single entity. This can include data from various sources, such as databases, files, and external data providers. Managed datasets can be used for different purposes, such as data analysis, reporting, and machine learning.

A layer is a subset of a managed dataset that has been processed and transformed to meet specific analytical requirements. Layers can be used for different purposes, such as visualisation, spatial analysis, or machine learning.

In an information supply chain, managed datasets, layers, and products can be used to optimise data management, analysis, and reporting. Organising data into managed datasets can improve data quality and consistency, as well as improve data governance and security. Layers can provide additional context and meaning to the data, making it easier to analyse and visualise. Products can provide actionable insights that can be used to make informed business decisions.

Managed datasets, layers and products can be primary inputs for further analysis. The quality and the accuracy of the records in the data are an integral part in both the selection of the data and in preparing it for subsequent analysis. The method of analysis to be undertaken will determine the degree to which the data may need filtering. Data may not always cover the areas of concern and some form of modelling may be required to extrapolate into those areas where the data are inadequate.

There is often an extensive process involved in preparing datasets, layers and products for further use. It is critical that the approaches used to transform raw data into managed datasets, layers and products are documented with provenance metadata including persistent identifiers for transparency, verifiability, and attribution of the original data creators.

Further sources: Curate

Data quality and fitness for purpose:

- ARDC Good Data Practices – ardc.edu.au/resource/good-data-practices/
- Australian and New Zealand Data Quality Interest Group – sites.google.com/ardc.edu.au/australian-data-quality-ig/resources
- Atlas of Living Australia – Data Quality Project: www.ala.org.au/data-quality-project/
- ARDC – ardc.edu.au/resources/working-with-data/metadata/
- Australian Bureau of Statistics – www.abs.gov.au/websitedbs/D3310114.nsf/home/Quality:+The+ABS+Data+Quality+Framework
- Atlas of Living Australia – www.ala.org.au/blogs-news/annotations-alerts-about-newannotations-and-annotations-of-interest/

Identifiers:

- Identifiers – ardc.edu.au/resource/citation-and-identifiers/

Vocabularies and conventions:

- Australian Biological Resources Study (ABRS) – www.environment.gov.au/science/abrs
- Taxonomy Australia – www.taxonomyaustralia.org.au/about-taxonomy-australia
- Research Vocabularies Australia – ardc.edu.au/services/research-vocabularies-australia/
- Guide to Vocabularies – ardc.edu.au/resource/guide-to-vocabularies-and-research-data/
- Research Vocabularies Australia – vocabs.ardc.edu.au/
- NERC Vocabularies – vocab.nerc.ac.uk/

Data and software publishing:

Guide to Choosing a Data Repository – ardc.edu.au/resource/guide-to-choosing-a-data-repository/

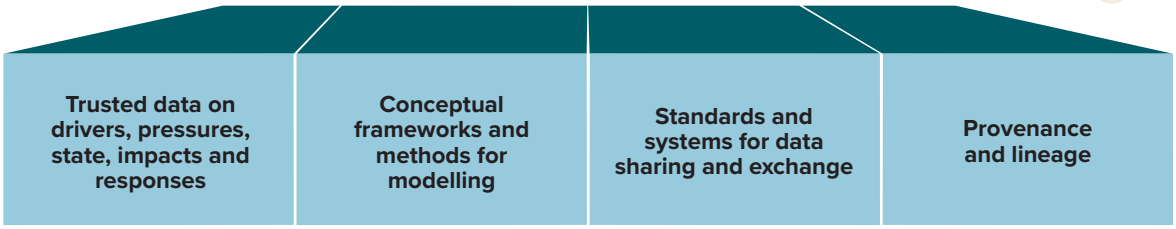
Trusted data repositories – ardc.edu.au/resource/trust-principles/

Registry of research data repositories – www.re3data.org/

Managed data sets, layers and products:

- Geoscience Australia – www.ga.gov.au
- Bureau of Meteorology – www.bom.gov.au
- Atlas of Living Australia – www.ala.org.au
- Terrestrial Ecosystem Research Network – www.tern.org.au
- Integrated Marine Observing System – www.imos.org.au
- AuScope – www.auscope.org.au
- ABARES – www.agriculture.gov.au/abares
- CSIRO – data.csiro.au
- Global Biodiversity Information Facility (GBIF) – www.gbif.org
- Research Data Australia – researchdata.edu.au/
- TNFD includes links to global data sets – https://framework.tnfd.global/tools-platforms/

INTEGRATE



The Integration layer takes data and curated data products and links them to other data products in preparation for use in analytic and modelling tools. It also identifies the key characteristics necessary to ensure their continued integrity, and the scientific and technical basis for their integration.

Trusted data on drivers, pressures, state, impacts and responses

The drivers-pressures-state-impacts-responses (DPSIR) model is a framework frequently used in the environment domain to understand the relationship between human activities and their impact on the environment. The components cover:

- **Drivers:** underlying causes that influence the state of the environment, for example population growth, economic development, or climate change
- **Pressures:** direct impacts of drivers on the environment, such as pollution, deforestation, or overfishing
- **State:** current condition of the environment, including the health of ecosystems, species, and natural resources
- **Impacts:** effects of pressures on the environment, for example declines in biodiversity, reduced water quality, or increased greenhouse gas emissions
- **Responses:** actions taken to address environmental challenges, such as conservation efforts, policy changes, or technological innovations.

The DPSIR model helps to identify the causal chain of events that lead to environmental problems and provides a framework for designing solutions. The DPSIR model is used in a wide range of environmental contexts, from local to global scales, and can be applied to different ecosystems and sectors including land use, water management and energy production.

Trusted data on the elements of the framework is important to provide accurate and reliable information for researchers, policymakers and other stakeholders to identify problems, research and evaluate solutions and assess progress toward goals.

Conceptual frameworks and methods for modelling

Conceptual frameworks and methods for modelling help organise complex information, identify relationships between variables and predict outcomes.

Conceptual frameworks such as the DPSIR model and the adaptive management cycle are used to understand the complex interactions between human activities and natural systems. These frameworks provide a systematic approach to identifying environmental problems, evaluating potential solutions, and monitoring progress.

Methods for modelling include statistical models, which analyse data and identify patterns and relationships between variables; simulation models, which use computer-based algorithms to predict the behaviour of complex systems; and optimisation models, which use algorithms to identify the best solutions to problems.

These approaches come together differently in different domains:

- **Climate change:** conceptual frameworks such as the carbon cycle and climate system models are used to understand the interactions between the atmosphere, oceans, and land surface. Modelling methods such as global circulation models are used to predict the impacts of human activities on the climate.
- **Biodiversity conservation:** there are many frameworks, including the ecosystem services framework and the adaptive management cycle, which are used to guide the conservation and management of ecosystems and biodiversity. Modelling methods such as population models and species distribution models help understand the dynamics of species populations and forecast the impacts of management interventions.

The Australian Ecosystem Models Framework, captures knowledge of ecosystem dynamics in a set of dynamic ecosystem models which describe the dynamic characteristics and drivers of Australian ecosystems. The models have the potential to provide an architecture for natural resource management prioritisation, including environmental assessments, as well as monitoring and evaluation.

- **Water resources management:** frameworks such as the water cycle and the watershed management approach are used to guide the management and conservation of water resources. Hydrological models are used to predict the availability and quality of water resources and to evaluate the impacts of different management interventions.
- **Energy and sustainability:** frameworks include sustainable development and the triple bottom line approach. Life cycle analysis and energy system models may help evaluate the environmental and social impacts of different energy sources and identify sustainable energy solutions.

The effectiveness of environmental conceptual frameworks and modelling methods depends on the quality of data and models used and the validity of the underlying assumptions. While there are maturing physical-process models (such as the ACCESS-NRI modelling framework for climate and weather, or the G-ADOPT[21] framework for geophysical modelling), conceptual models of the biosphere – from the molecular level to whole ecosystems – remain immature overall. There is considerable scientific work needed to build conceptual models to improve understanding of biological systems and integrate that knowledge into other models, from geology to economics.

[21] https://g-adopt.github.io/

Standards and systems for data sharing and exchange

Standards and systems for data sharing and exchange are protocols and guidelines that facilitate the sharing of data among different organisations and systems. These standards and systems enable data to be shared and used securely and efficiently and help promote transparency and interoperability for research and decision-making.

Examples of standards and systems for data sharing and exchange include:

- **Application Programming Interfaces:** a standardised way for different systems to exchange data and communicate with each other. APIs can be used to exchange data between different software applications or to integrate data from multiple sources
- **Open data standards:** a common format for sharing data and metadata between different systems. There are many such standards, including:
  - **Sensor Observation Service (SOS):** an open standard that defines a standard interface for requesting and receiving sensor data from different sources, widely used in environmental monitoring and management.
  - **Open Geospatial Consortium (OGC):** the OGC is a global organisation that develops open standards for geospatial data and technologies. OGC has developed several standards for environmental data, including the Web Feature Service (WFS), which provides a standard interface for accessing and sharing feature data, and the Web Coverage Service (WCS), which provides a standard interface for accessing and sharing raster data.
  - **Darwin Core:** a biodiversity data standard developed by Biodiversity Information Standards (TDWG) that provides a standardised format for sharing data on species occurrence and distribution.
- **Metadata Standards:** Metadata standards provide a common framework for describing and organising data. These standards help ensure that data can be easily understood and used by different users and systems. Examples of metadata standards include:
  - **Ecological Metadata Language (EML):** a metadata standard that provides a common format for describing ecological datasets.
  - **Climate and Forecast Metadata Conventions (CF):** a metadata standard developed by the climate research community that provides a common format for describing climate datasets.
- **Data Sharing Policies:** Data sharing policies provide guidelines and rules for how data should be shared and used. These policies can be developed by individual organisations or by governments and international organisations to ensure that data is shared in a FAIR, transparent, and secure manner.
- **Interoperability Standards:** Interoperability standards ensure that different systems can work together seamlessly. These standards define common protocols and interfaces that allow different systems to exchange data and communicate with each other.

Provenance and lineage

The ability to track unit level data from the point of creation through curation and to integration into data products and systems for exchange, to be analysed and modelled, is critical to building confidence in decision-support tools. This is often referred to as recording provenance or lineage.

Recording provenance information aids analysis of results based upon dependencies upon particular data or other inputs, as well as error-detection, auditing and compliance investigation. It also helps track intellectual property and enable attribution of the original creators of data assets.

Assurance of integration processes is needed to ensure that the inputs to be analysed and modelled are fit for purpose. Capturing provenance and lineage may require considerable metadata documentation, including data processing and transformations. It is often challenging to manually track which data or data products contributed to model results; analysis and modelling systems should automatically track and record this information in a machine-readable form.

> ***Persistent identifiers are important to ensure the longer-term stability of references to specific data or knowledge products or model versions.***

Further sources: Integrate

Trusted data on drivers, pressures, state, impacts and responses:

- DCCEEW Find Environmental Data – www.environment.gov.au/fed/catalog/main/home.page
- TNFD includes links to global data sets – framework.tnfd.global/tools-platforms/
- CSIRO – www.csiro.au
- Geoscience Australia – www.ga.gov.au
- Bureau of Meteorology – www.bom.gov.au
- Atlas of Living Australia (ALA) – www.ala.org.au
- Terrestrial Ecosystem Research Network (TERN) – www.tern.org.au
- Integrated Marine Observing System (IMOS) – www.imos.org.au
- AuScope – www.auscope.org.au
- AURIN – aurin.org.au/

Conceptual frameworks and methods for modelling:

- DPSIR – www.eea.europa.eu/themes/sustainability-transitions/state-of-the-environment-reporting/information-and-knowledge-for-a; https://www.fao.org/land-water/land/land-governance/land-resources-planning-toolbox/category/details/en/c/1026561/; https://archive.epa.gov/ged/tutorial/web/pdf/dpsir_module_2.pdf
- Pressure state response model – www.epa.wa.gov.au/state-environment-reporting
- Australia Ecosystem Model Frameworks – research.csiro.au/biodiversity-knowledge/projects/models-framework/
- Causal pathways – www.bioregionalassessments.gov.au/

Standards and systems for data sharing and exchange:

Australian Research Data Commons – ardc.edu.au/services/research-data-australia/, ardc.edu.au/resource/geospatial-data-and-metadata/

National Data Commissioner – www.datacommissioner.gov.au/resources/draft-data-sharingagreement-template

Design Standards for Whole of Australian Government Application Programming Interfaces – api.gov.au

Sensor Observation Service – www.ogc.org/standard/sos/

Darwin Core – dwc.tdwg.org/

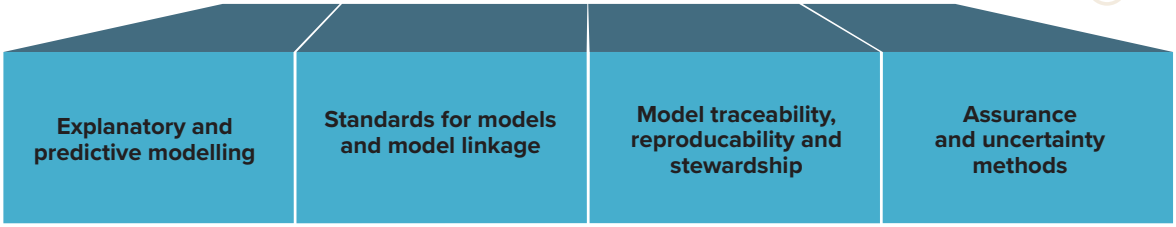
Geoscience Australia profile of the ISO 19115:2014 Geographic Information metadata standard – www.ga.gov.au/data-pubs/datastandards

Australian Government Locator Service (AGLS) Metadata Standard – www.naa.gov.au/node/264

Provenance and lineage:

Data provenance – ardc.edu.au/resource/data-provenance/

ANALYSE



The Analysis layer identifies the analytic and modelling capabilities that underpin research outcomes, reporting and decision support tools.

Explanatory and predictive modelling

Explanatory and predictive modelling (Figure 5) is used to support research, reporting and decisions[22]:

- **Explanatory modelling:** used for analysing large, complex datasets to test hypotheses, using statistical or machine learning techniques, to gain insights into data and identify causal relationships between observed variables. It can help identify potential areas for further research or analysis, and developed models often then provide the foundation for predictive modelling.
- **Predictive modelling:** uses existing models derived either through explanatory modelling, or through expert-based or process-based deductive modelling, to make predictions about a variable of interest for which complete observations are not available. It can help monitor and report system change through repeated observation of readily measurable driver variables, and through use of these as inputs to predictive modelling of less-readily-measurable response variables. Predictive modelling can be retrospective, predicting aspects of the system that are not directly observed, or future-oriented, offering forecasts or exploring scenarios.

[22] **Ferrier S**, Jetz W, Scharlemann J (2017) Biodiversity modelling as part of an observation system. Pp 239-257 in M Walters and RJ Scholes, eds. *The GEO Handbook on Biodiversity Observation Networks*. Springer International Publishing. https://link.springer.com/chapter/10.1007/978-3-319-27288-7_10

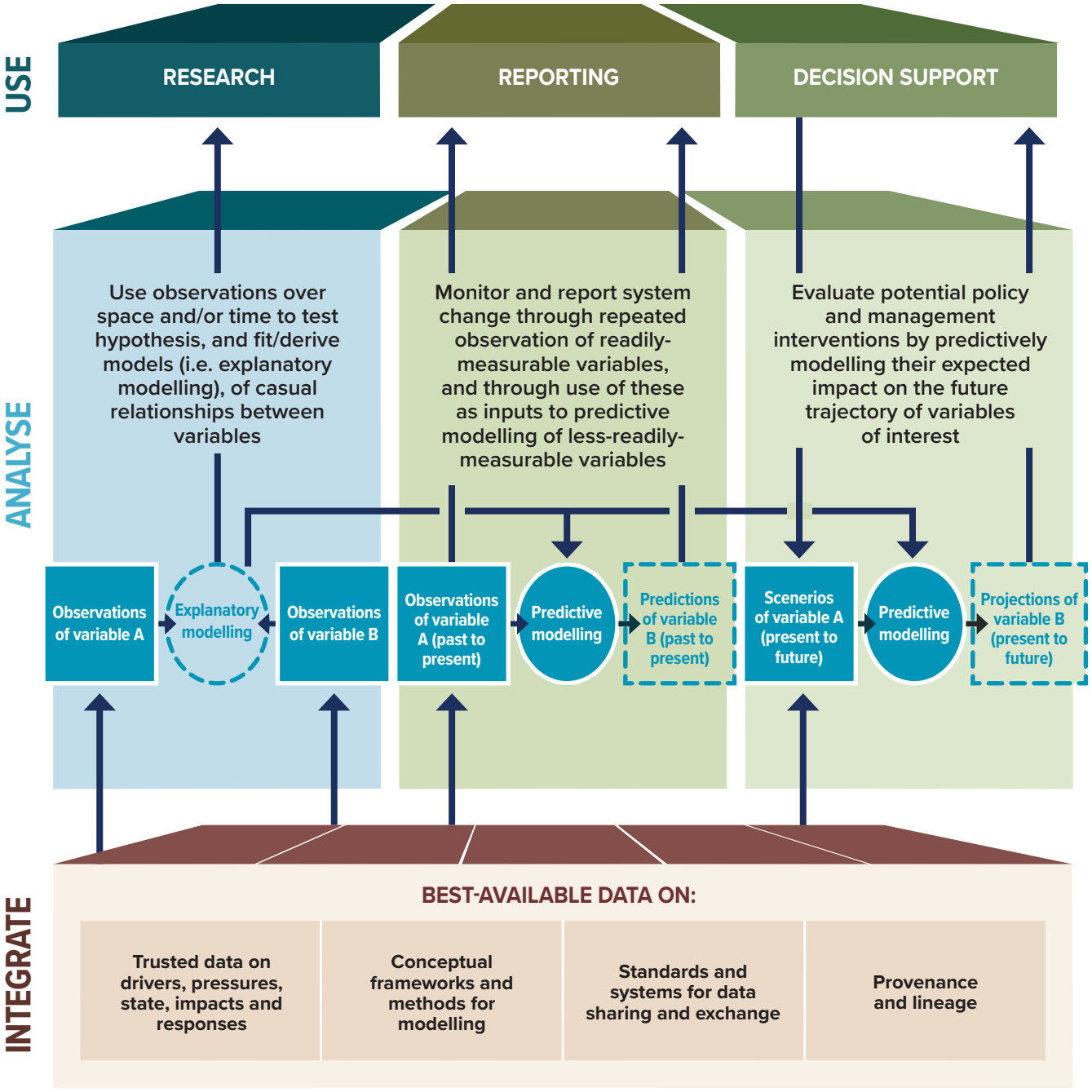


FIGURE 5: Explanatory and predictive modelling techniques

Domain specific models for each of these techniques are available for environmental characteristics including climate, land surface, species, hydrology and habitat extent and condition. There are many such models available, peer reviewed and curated, often with reliable data sources, though particularly at local scales data may need to be supplemented.

Cross-domain modelling can take place through the loose coupling of specialised models. This has the advantage that the specific strengths of each model are retained, though limited information is generally exchanged between coupled models, and often in only one direction, with an accompanying lack of feedback between the modelled components. There is a risk of inconsistencies in representations of the same phenomenon in the different models.

A challenge is to make models dynamic, able to readily update based on new flows of data.

The modelling of cumulative impact remains a challenge, as it is often not additive but multifactorial with feedback loops and relationships between different factors that are not always well understood.

Standards for models and model linkage

Standards for environmental models and model linkage are important for ensuring that environmental data and models are interoperable, accessible, and transparent. Standards for models and model linkage can help assess the reliability of model results, ensure transparency and consistency in the translation of scientific results into decision support tools, and focus on where improvements might be most needed in the underlying science.

The scale, or extent and resolution of models, can differ. Ecological patterns and processes change at different scales, and ecosystems have different features and structures that influence inter-relationships between interacting species. The scale-dependence of these relationships is not always apparent because of variations in methodological reliability as well as data availability and accuracy. Resolution Is often adjusted to enable models to run on cheaper and more accessible infrastructures (e.g., laptops, on-premise servers, local clouds).

Integrated assessment models embed different model representations of the system in a consistent manner. The inclusion of feedback and interaction between the different modules is generally stronger and there is more likely to be consistent representation of variables across the different modules. Such models have inherent complexity, which reduces the applicability and transparency of the models.

There are no overall standards for ecosystem models, though there are standards for elements of ecosystem modelling. There are however means to select models and tools for analysis, and standardised QA/QC procedures for risk characterisation and peer review.

Model traceability, reproducibility and stewardship

How models are specified can have considerable impact upon the results that they produce. It can be difficult to gain visibility of model parameters, or the impact of any choices and changes. Platforms are available that offer models with default parameters, which can be altered while creating an auditable record of change.

Stewardship is a concept commonly applied to the activities that preserve and improve the information content, accessibility, and usability of data and metadata. The same concept is important for models. Stewardship activities are a critical support for assurance and reuse, as well as long-term preservation.

A recent survey by the scientific journal Nature found that 'more than 70% of researchers were unable to reproduce research by others, and 50% were not even able to reproduce their own results'.[23] Metadata standards are one response to this, as are standardised datasets, models and model parameters (together with customisation and DOIs). Complex models may need large numbers of DOIs, leading to challenges of complex data citations.[24]

> ***The ability to reproduce modelled results is central to public trust and assurance of any decision-support tools.***

Assurance and uncertainty methods

Assurance includes setting standards for best practices, using model-data and model-model inter-comparisons to provide robust and transparent evaluations of uncertainty and encouraging new research into methods of measuring and communicating uncertainty and its impact on decision-making. It includes QA/QC approaches.

Uncertainty in scenarios and models arises from a variety of sources, including insufficient or erroneous data used to construct and test models; lack of understanding or inadequate representation of underlying processes; and low predictability or random behaviour in a system. Biodiversity and ecosystem models currently available provide a range of options to assist policymakers in understanding relationships between drivers and impacts, and in evaluating interventions.

For knowledge products to be used by proponents to shape investment proposals, for regulators to make decisions, and for public trust, models must be both of high standards and known to be so.

An example of model assurance in practice is through the scientific oversight built into EcoCommons. An expert committee provides assurance over 100s of curated datasets, 17+ peer reviewed species models, default model parameters and more. While users can introduce new data and vary parameters, DOIs can be minted for all analytic results creating a permanent record of all data, model parameters, etc, and offering an audit and reproducibility trail.

[23] Feng, X., Park, D.S., Walker, C. et al. 'A checklist for maximizing reproducibility of ecological niche models'. Nat Ecol Evol 3, 1382–1395 (2019). https://doi.org/10.1038/s41559-019-0972-5

[24] https://rd-alliance.org/group/complex-citations-working-group/case-statement/complex-citations-working-group-case-statement

Further sources: Analyse

Explanatory and predictive modelling:

- OECD ENV-Linkages Model – www.oecd-ilibrary.org/environment-and-sustainable-development/an-overview-of-the-oecd-env-linkages-model_5jz2qck2b2vd-en
- US EPA – www.epa.gov/measurements-modeling/environmental-modeling

Standards for models and model linkage:

- Foundation Spatial Data Framework – fsdf.org.au

Model traceability, reproducibility and stewardship:

- Model reproducibility standard – Feng, X., Park, D.S., Walker, C. et al. 'A checklist for maximising reproducibility of ecological niche models'. Nat Ecol Evol 3, 1382–1395 (2019). doi.org/10.1038/s41559-019-0972-5

Assurance and uncertainty methods

- A number of reports discuss approaches to uncertainty in their field, eg. Great Barrier Reef Outlook Report 2019 – www.gbrmpa.gov.au/our-work/outlook-report-2019

Appendix: CARE principles and biodiversity information

Applying the CARE principles for the collection and management of biodiversity data involves the following considerations:

CARE Component	Considerations
Collective benefit	• How will the data be used – What benefits will it bring to the community and industry? • Clear understanding towards the data being collected/ researched – Purpose • Clear documentation of the data to be collected and how it will be analysed and translated – Noting any potential impacts on Indigenous communities – What is the research question? • Decision making processes – be inclusive

CARE Component	Considerations

Authority to control	• Who is collecting the data?
	• Who are the data custodians?
	  – Have they engaged with the Indigenous community?
	  – Are they aware of who to liaise with?
	• Who manages the data?
	• How is the data to be shared/accessed?
	  – Who controls this process?
	• Decision making on access and analysis – who and the process for this
	  – Collaborative approach

Responsibility	• Who holds the responsibility of the data being collected/analysed/shared from the research project perspective? • Who within the Aboriginal community needs to be referred to in relation to the collection/analysis/sharing of the data? • Who are the decision makers on both sides? • How will research be communicated? • Ability to describe how the data will support collective benefit and self-determination

Ethics	• Statements on the impacts that the data collection and analysis might have
	  – Ensuring no harm on Indigenous communities exists
	• Describe the potential benefits/risks the research might have on Aboriginal communities
	• Awareness towards community dynamics
	  – Cultural beliefs
	  – Laws and rules
	• Consider the impacts of the research on the whole data management lifecycle
	  – Apply this to the needs of Aboriginal communities

wabsi.org.au
wamsi.org.au
ardc.edu.au