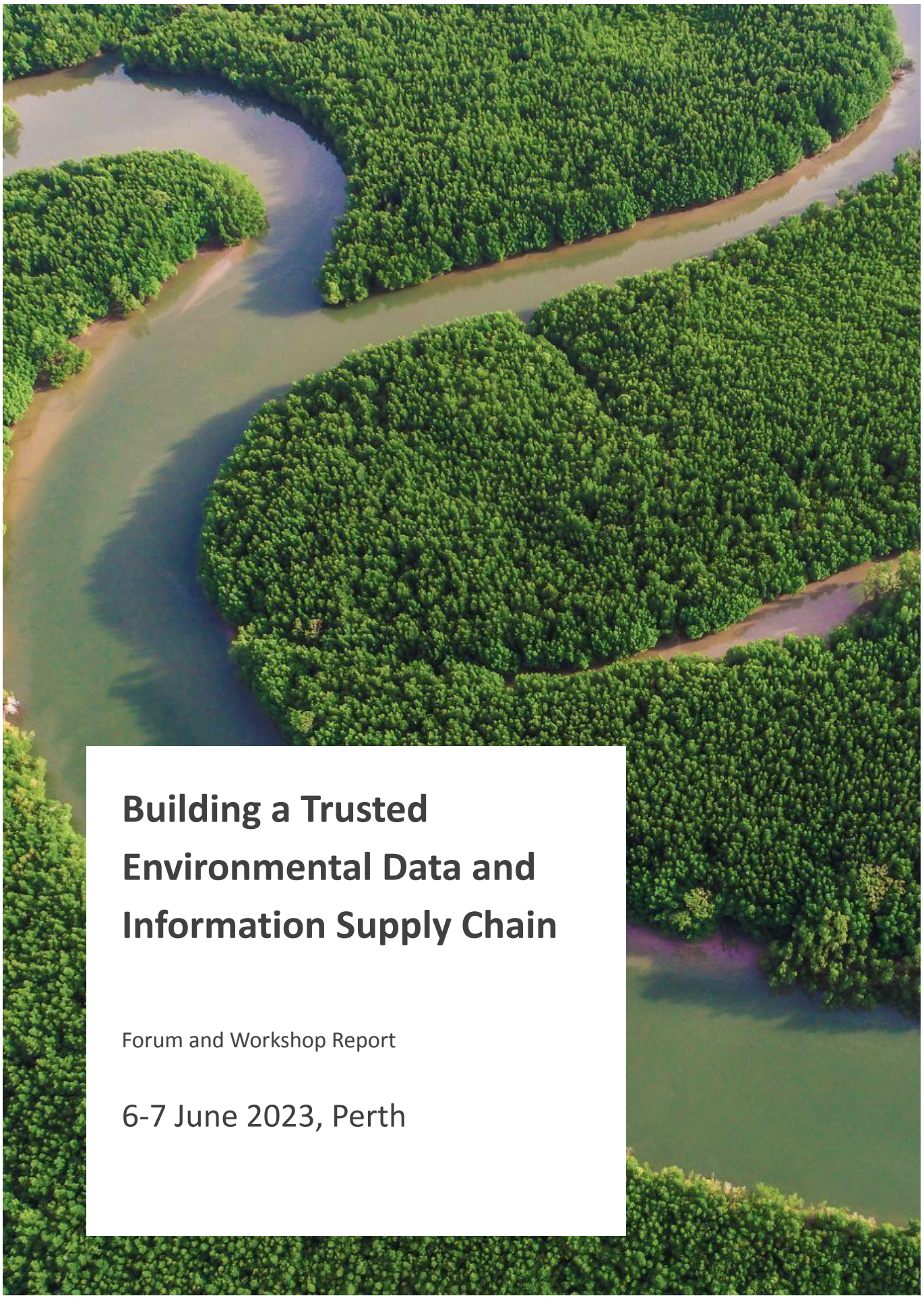

Building a Trusted Environmental Data and Information Supply Chain

Forum and Workshop Report

6-7 June 2023, Perth

WESTERN AUSTRALIAN MARINE SCIENCE INSTITUTION
The Western Australian Biodiversity Science Institute
Australian Research Data Commons
NCRIS National Research Infrastructure for Australia An Australian Government Initiative
ARDC is enabled by NCRIS

Acknowledgments

Thank you to our many partners across industry, government, research and community for their contributions, and for the help of Greg Terrill, Kerry Levett, Chris Gentle and Hamish Holewa in particular for compiling this report.

Cover Image - themorningglory - 391783296 / adobestock.com

DOI: *10.5281/zenodo.8337061*.

Citation:

ARDC, WABSI, WAMSI. Building a Trusted Environmental Data and Information Supply Chain - Forum and Workshop Report. 2023. Available from doi.org/10.5281/zenodo.8337061.

Published - 28 September 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.

Contributors

WESTERN AUSTRALIAN MARINE SCIENCE INSTITUTION

wamsi.org.au

The Western Australian Biodiversity SCIENCE INSTITUTE

wabsi.org.au

Australian Research Data Commons

ardc.edu.au

The Planet Research Data Commons is supported by the Australian Research Data Commons (ARDC). The ARDC is enabled by NCRIS.

CONTENTS

1. Introduction

Trusted data and information supply chains are vital for research and evidence-based decision making. The *Independent Review of the Environment Protection and Biodiversity Conservation Act* highlighted the need for an effective 'supply chain' of environmental information. Data and information are delivered through a series of processes that convert raw data, through curation and analysis, into end products that can be used – by researchers to understand the environment, and by decision-makers to inform their decisions. As with more traditional supply chains, each function can be provided by a different party, and coordination is needed for an efficient chain that delivers the right products at the right time to the right customers. A trusted data and information supply chain means having an effective system of data-sharing agreements and information systems that can talk with each other.

A trusted data and information supply chain is a powerful concept. It can align the endeavours of many organisations to a common national cause including:

- Industry - rapid and reliable environmental data for analysis, reporting on nature positive outcomes and disclosing nature-related financial information
- Government - more rapid data assembly for trusted analysis, enabling robust, repeatable environmental decision making, including for cumulative environmental impacts in a region
- Community - managing and reporting on natural resource management and Country, and increased trust and environmental outcomes
- Research - strengthening the availability of data for research, highlighting priority research questions and providing transparent translation of research to decisions.

The Australian Research Data Commons (ARDC)[1], The Western Australian Biodiversity Science Institute (WABSI) and the Western Australian Marine Science Institution (WAMSI) and partners are currently leading an approach to establish environmental information supply chains. To help advance this work, the ARDC, WABSI, WAMSI organised three events to explore the building of trusted environmental data and information supply chains at the University Club, University of Western Australia, Perth, 6-7 June 2023.

The events brought together senior representatives from large businesses and industry, senior regulators and policy leaders from the Australian and Western Australian Governments, and national research institutions, research infrastructure providers and the Australian Government Department of Education.

Over 80 stakeholders attended the events. The organisations represented can be found in Appendix 1.

[1] As part of the Planet Research Data Commons, which provides national-scale data infrastructure for earth and environmental research decision-making. Learn more about the Planet RDC.

1.1. Objectives

Three events were organised, with the following objectives:

Pre-briefing

- An opt-in briefing to provide an introduction to ARDC Planet RDC, the SAFE 2.0 guide, the Shared Environmental Analytics Facility (SEAF) and work underway in Cockburn Sound and Pilbara.

High level forum

- To elicit needs and capabilities from stakeholders in industry, government and research sectors;
- The high level forum explored two topics:
    - Trusted Environmental Data and Information Supply Chains: why they are critical for the government, industry and research sectors
    - What do we need to do to enable Trusted Environmental Data and Information Supply Chains by 2026?

Workshop

Intended to

- Agree on an inclusive process to define and map an environmental information supply chain, by October 2023
- Scope collaborative industry-research-government approaches to building a supply chain in a Western Australian region (for example, the Pilbara).

Format

- Workshopped current constraints and opportunities to operationalise robust, repeatable and sustainable environmental data and information supply chains
- Used the SAFE 2.0 framework (Culture, Collect, Curate, Integrate, Analyse ‘layers’) to map the current state of environmental data and information supply chains
- Assessed the maturity of the elements, and what can be done to lift the capability.

The list of speakers can be found in Appendix 2. Further detail on the agendas is included in Appendix 3.

2. Forum: the challenge

Speakers identified four core themes motivating regulators and driving government, industry and the community:

- the need for deeper ecosystem knowledge
- the importance of evaluating cumulative impacts
- the need to analyse environmental performance and outcomes
- the challenges facing robust and timely decision making.

Each of these themes is underpinned by reliable joined-up trusted data – a supply chain of environmental information. This does not currently exist.

The challenge for the workshop was to identify what is needed to create a trusted supply chain of environmental information to underpin decisions by regulators, industry, governments and the community. In short, to bring environmental data up to the standards and timeliness of financial and economic data.

3. Workshop: themes and feedback

The workshop explored the challenges and opportunities of the capabilities that make up the 'links' of the supply chain. The workshop was structured around the recently revised *Shared Analytic Framework for the Environment (SAFE 2.0)* Guide, which provides a common language to describe the capabilities and their interdependencies.

SAFE is a means to understand the complexity of the environmental data and analytics landscape. SAFE depicts the capabilities 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.

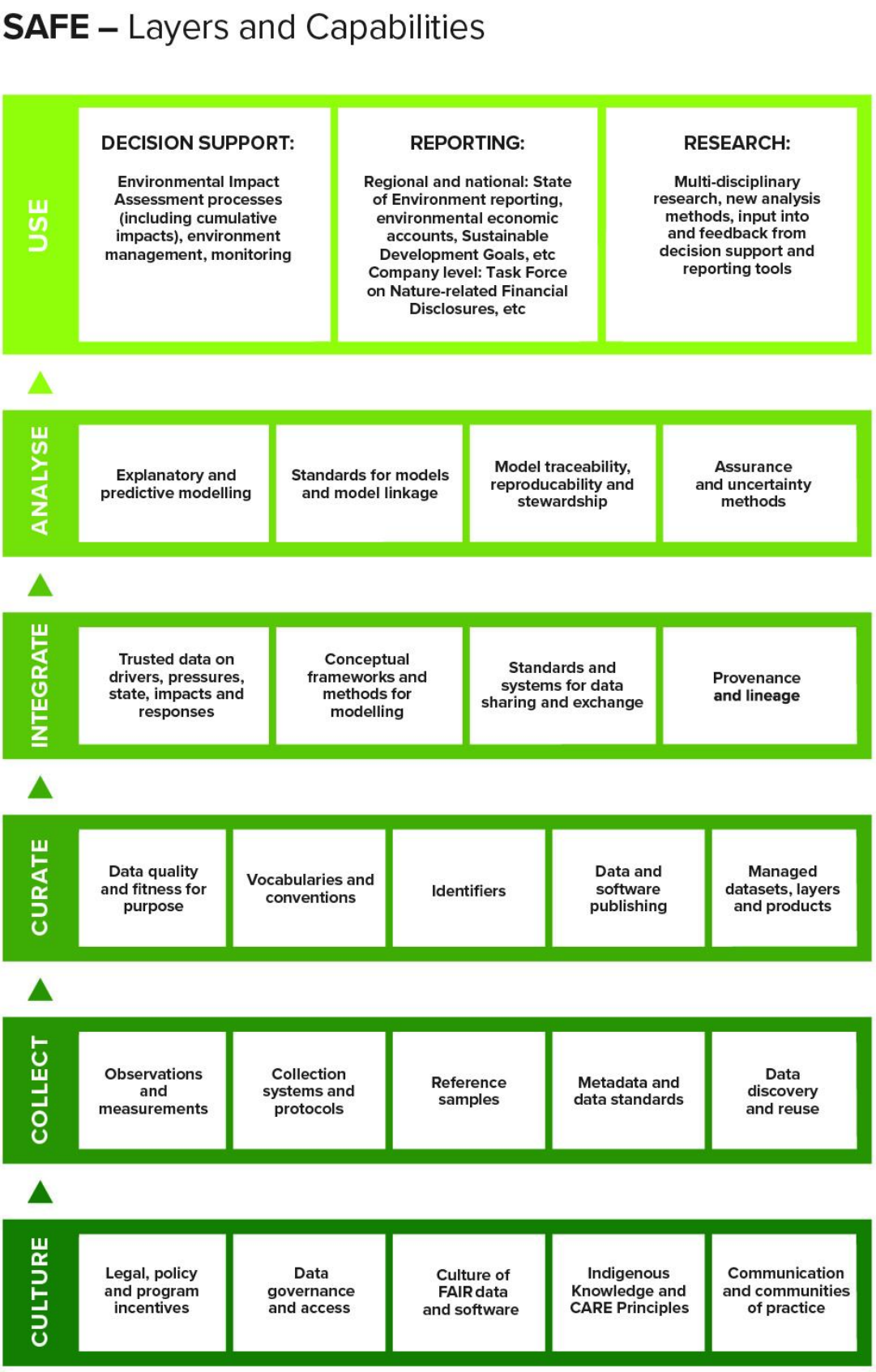

Figure 1. The SAFE matrix. SAFE has five layers, each of which consist of key capabilities (the boxes) that support the environmental information supply chain. All capabilities and layers interconnect and add value to each other. The layers and boxes are conceptual, and in many cases a particular organisation will provide services across more than one layer.

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. It is agnostic towards discipline and domain.

The SAFE framework provides a common vocabulary to show how the many components of the environmental data landscape work together. It helps 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 national supply chain of environmental information.

The workshop assessed the maturity of various elements of the environmental information supply chain by working through the five layers of the SAFE framework (Figure 1).

Participants provided their assessment of the maturity of the capabilities in each layer, and to comment on the key challenges and opportunities for improving each layer, using the online Mentimeter tool.

3.1 Culture

Participants provided an overall assessment of the current maturity of the five capabilities within the Culture layer (Figure 2).

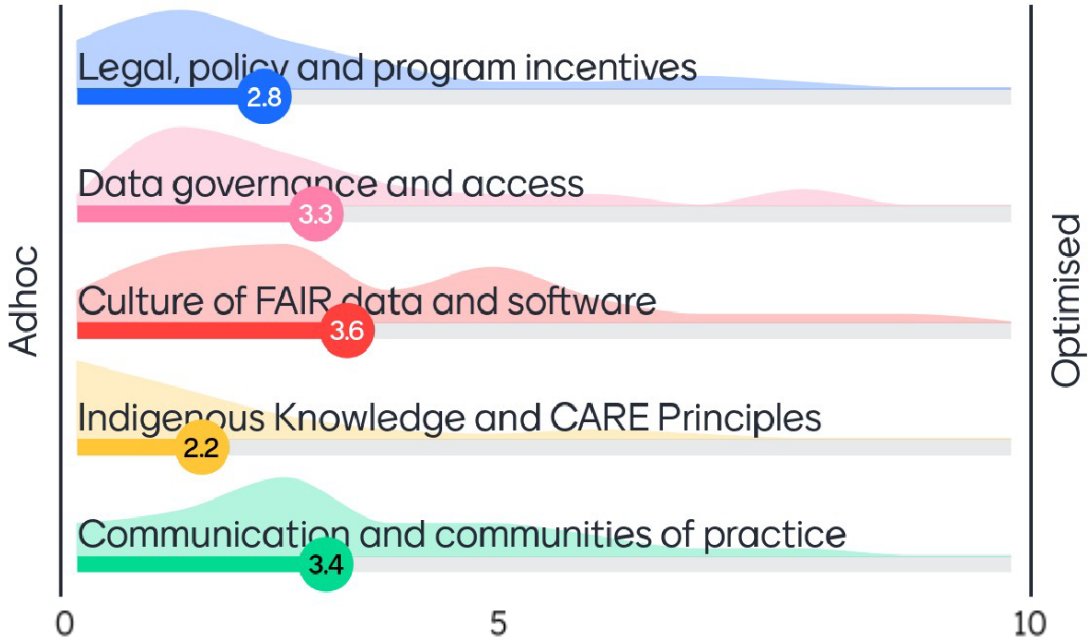

Figure 2. Assessment of the maturity of the capabilities within the Culture layer.

Participants were asked to comment upon the key challenges that inhibit actions in Culture. Selected responses include:

- From an industry perspective, if it's not legislated, data is not shared. ESG [environmental, social and corporate governance] is driving performance metrics but is not mandated. We want government to say what the data is, what standard is required and where it should go
- Need to have mechanisms to protect business from the risk of sharing data, versus the benefit to society and science as a whole
- Appropriate understanding of sensitivities and protection of Indigenous Knowledge
- Need trust elements for systems, and be able to easily recognise systems and institutions that meet minimal standards of data governance and processes, including service level agreements etc.

Participants were asked to comment upon the key actions for improving Culture. Selected responses include:

- Better inclusion of end-users within governance and legislative structures so that sharing can occur in a safe manner
- At a national political level implementing policy providing very good and rewarding incentives to make data open
- Implement data sourcing and context traceability
- Legislative imperatives and commercial values that incentivise data is curated and shared
- Undertake a design function to determine what a trusted backbone is and to build this in to projects that we are undertaking
- Easier and more transparent processes to protecting sensitive data
- Traditional owners data sharing improvement needs to build trust through listening and understanding concerns, priorities and respecting those.

3.2 Collect

Participants provided an overall assessment of the current maturity of existing the five capabilities within the Collect layer (Figure 3).

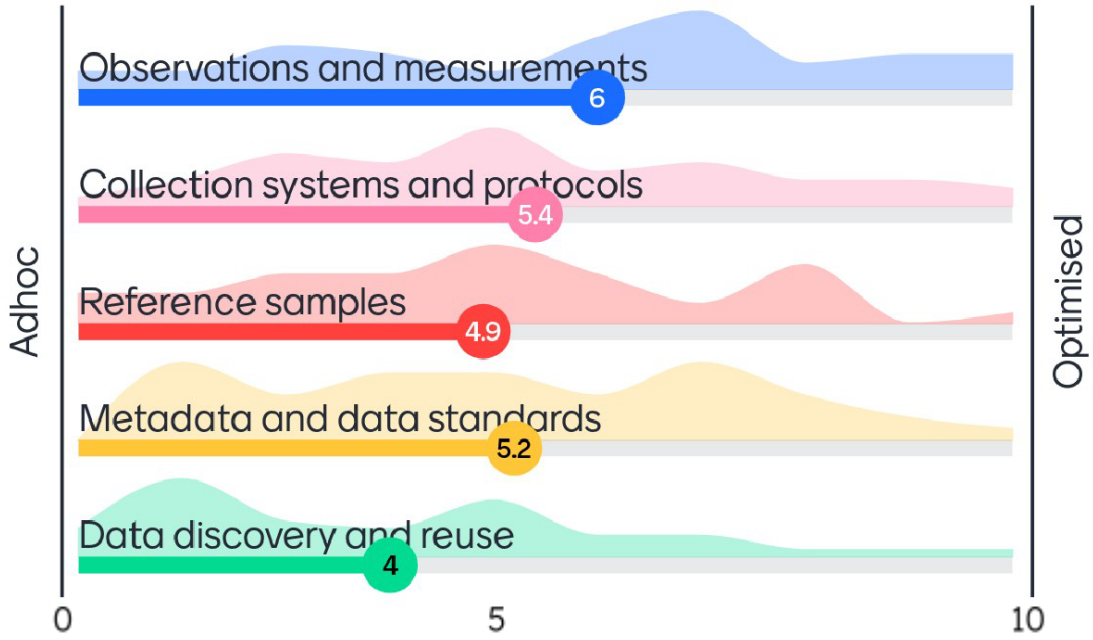

Figure 3. Assessment of the maturity of the capabilities within the Collect layer.

Participants were asked to comment upon the key challenges that inhibit actions in Collect. Selected responses include:

- Lack of flexibility and alignment in collection systems and data storage
- Absences of widely understood and agreed collection protocols, metadata and data standards
- No mandate to do trend and baseline analysis which means data is collected for 5 years then discarded
- No body that defines priority data areas/ gaps in data across regions, biodiversity/ ecology
- A lot of data are now published according to Open Data principles but are not readily discoverable or reuseable
- Metadata, metadata, metadata
- No facilities to provide active storage and curation of long tail data - where do I put data that is sitting under my desk?

Participants were asked to comment upon the key actions for improving Collect. Selected responses include:

- National standards
- Identifying and standardising essential metadata fields
- Grants awarded with clear requirements for data collection / contribution
- Consistency to standards. Standards are always changing. Perhaps introduce governing body on standards in Australia
- Cataloguing and exposing vocabularies so good metadata is collected
- Improve traceability (objectives of data collection, design of acquisition, where is it used further down the chain)
- Ensure collectors understand the pathway for appropriate and enduring (de-risked) re-use of the data they collect
- Proper uncertainty propagation due to lack of data, parameter and conceptual model uncertainty quantification.

3.3 Curate

Participants provided an overall assessment of the current maturity of existing the five capabilities within the Curate layer (Figure 4).

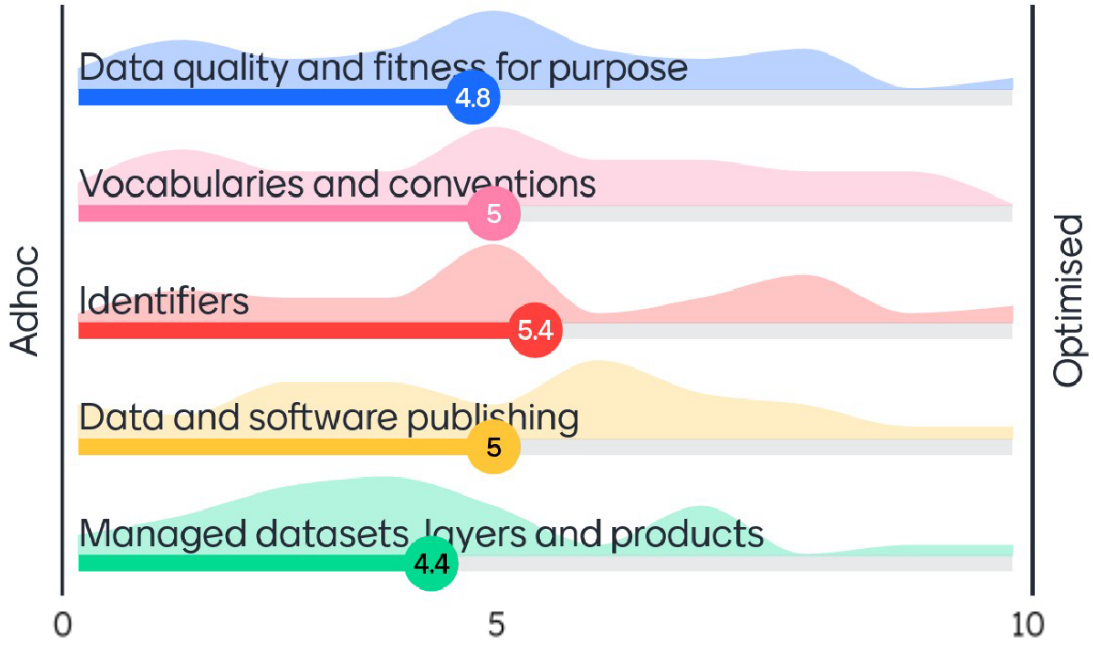

Figure 4. Assessment of the maturity of the capabilities within the Curate layer.

Participants were asked to comment upon the key challenges that inhibit actions in Curate. Selected responses include:

- Lack of widely known or widely used vocabularies and conventions outside research sector

- Lack of common metadata and standards makes it very difficult to curate data across sectors or disciplines.
- Lack of a continual improvement process to progressive improve the standards of curation by scientists.
- Understanding end-user requirements for managed datasets, layers and products.
- Lack of QA standards for curation.
- Sheer VOLUME of data being collected digitally
- Organisational reluctance to invest in data curation as a function in its own right as opposed to a means to a specific (project-focussed) end.
- "Fitness of purpose" is generally project-specific - we lack any overarching framework that defines this.
- Find a way to get data on balance sheets - make it an asset - raw and products
- incentives/reward for quality curated data.

Participants were asked to comment upon the key actions for improving Curate. Selected responses include:

- Exploit version control for precise attribution of software version used in analysis
- Define and agree on an overarching framework and standards, so that project specific work can be curated to be consistent with this
- make Data Stewardship an integral part of project management not just an afterthought
- Build and share automation tools to curators to scale their impact/increase accuracy and impact
- Building the cost of curation into research grants and making it a requirement of the grant
- Automate curation where possible
- Place a value on data as an organisational asset so that curation is a routinely funded activity
- Funding agencies require collated data to be provided as part of all project deliverables.

3.4 Integrate

Participants provided an overall assessment of the current maturity of existing the four capabilities within the Integrate layer (Figure 5).

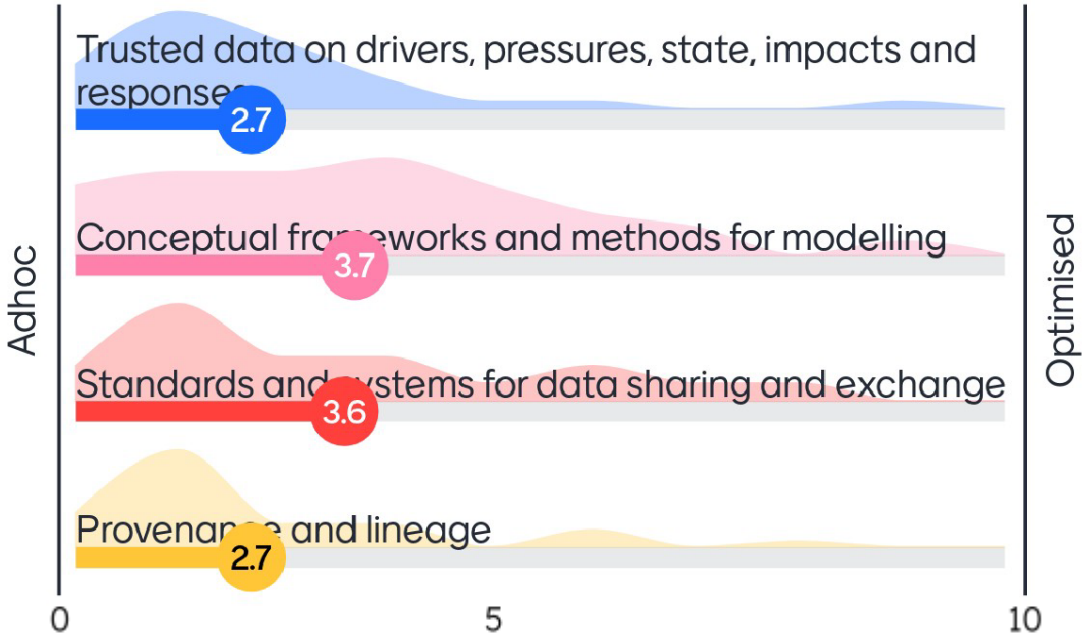

Figure 5. Assessment of the maturity of the capabilities within the Integrate layer.

Participants were asked to comment upon the key challenges that inhibit actions in Integrate. Selected responses include:

- Lack of focus on driver, pressure and response data in the culture, collect and curate layers, where biological data is the focus
- Compounding of failures in all the preceding layers (culture, collect, curate), particularly the absence of standards
- Having data ready and machine readable ready for use
- Not having translation pathways well defined with government and researchers
- Lack of information on relationships between state, drivers and responses
- Accessible, useable data with adequate metadata to inform its appropriate use and the confidence levels of derived products
- Fundamental culture barrier to sharing. Sharing diminishes individual relevance
- Integration happens for a purpose. The mechanisms and knowledge about how that happens is not shared. The silos of excellence are integrated but what is the meta integration that happens after that
- Scientific publication process inhibits the early release of new data, understandings and findings.

Participants were asked to comment upon the key actions for improving Integrate. Selected responses include:

- Make sure we get the lower layers right (culture, collect, curate)
- More knowledge broker roles connecting people between industry, research & government
- Formalise and widely adopt standard approaches for describing the level of confidence of any data or derived (integrated) products
- Establish common platform(s) for integration
- Need a single point of entry to all environmental data that is available to new methods.
- need to understand what are the problems that users have regarding integration. ie a system to capture the integration needs of users
- Forums to work with end-user to better define questions
- Integration layer should 'pull' the right products up, so the layers below are responding to the demand, rather than 'pushing' products upwards that may not meet the need.

3.5 Analyse

Participants provided an overall assessment of the current maturity of existing the four capabilities within the Analyse layer (Figure 6).

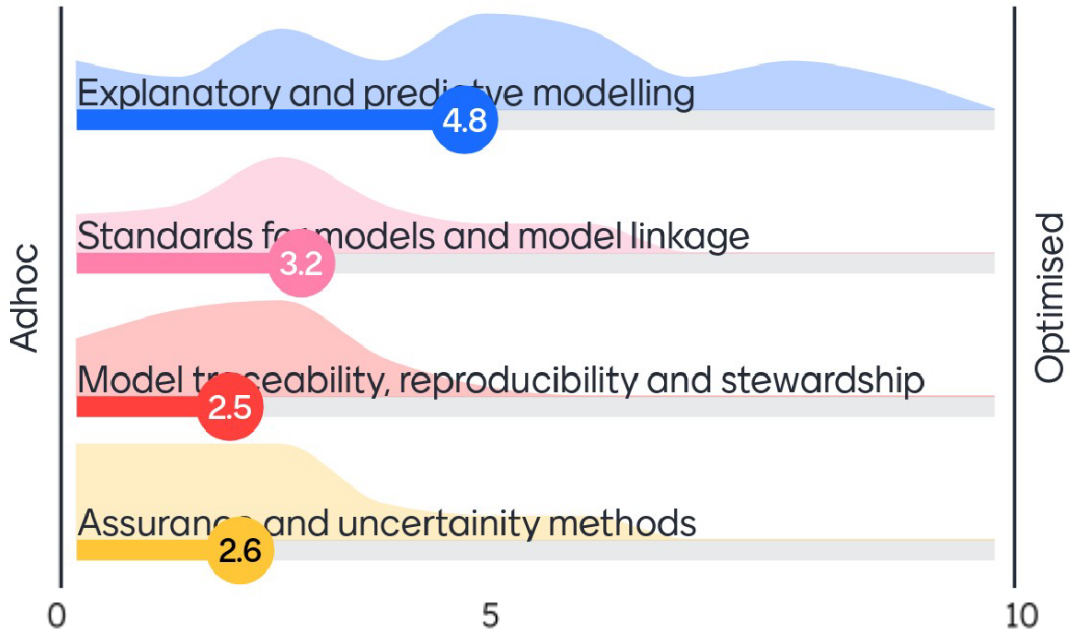

Figure 6. Assessment of the maturity of the capabilities within the Analyse layer.

Participants were asked to comment upon the key challenges that inhibit actions in Analyse. Selected responses include:

- Model version control
- There are many models - how do we ascertain quality and best practice
- Models are siloed and their outputs are not networked - what is the interoperability between them?
- Proper uncertainty quantification due to lack of data, parameter and conceptual model uncertainty quantification
- Compounding uncertainties - challenges in how to compound and how to communicate
- Different sources of uncertainty are not always considered and cannot be propagated
- User-proofing models where appropriate to reduce risk of inappropriate use, in tandem with good documentation.

Participants were asked to comment upon the key actions for improving Analyse. Selected responses include:

- Support development of ensemble modelling systems that are easy to use
- An integrated approach for uncertainty quantification across the different layers to enable proper uncertainty propagation in (predictive) model ensembles
- Communicating that uncertainty can be a decision product/output alongside the model, that can help identify gaps etc.
- Analytics and the associated statistics
- Peer review built in
- Quantified uncertainties with all model prediction and forecasts
- User-proofing models by design where appropriate and practical, along with providing robust documentation
- Include all assumptions and short falls in the models and their tabling. Including degree to which Traditional Custodians have been included.

4. Key Recommendations

Key recommendations for next steps include leading an inclusive process to define and map an environmental data and information supply chain with a focus on the Pilbara region. This process would involve relevant stakeholders across research, government, industry, traditional owners and non-government organisations that have an interest in the Pilbara region.

This work will form the basis of an integrative approach that includes trusted data, information and models to understand and predict environmental trends and cumulative impact across a region.

Outcomes from this process will be used to inform a repeatable method of defining and developing environmental data and information supply chains for other regions of interest.

It is recommended that a taskforce of interested stakeholders is established to build on the Pilbara *SEAF feasibility study* and *SAFE guide* by:

1. Conducting a collaborative project to refine cross-sector (industry, research, research infrastructure, government) approaches to building an environmental data and information supply chain that allows for future trend and cumulative impact analysis in the Pilbara region of Western Australia (including Products, Data and analytics, Platform, Science, and Management)
2. Identify mechanisms required for the environmental data and information supply chain to operate effectively including mechanisms for accrediting the trustworthiness of providers of data and services (e.g *CoreTrustSeal*), common data governance frameworks, data sharing protocols, licences, and for operationalising agreements (e.g. service-level agreements) between organisations for providing enduring data and services
3. Identify means to reduce the cost of data integration, including data being FAIR from the point of creation and throughout the environmental data and information supply chain
4. Develop an investable plan that delivers enduring environmental data and information supply chain capability for the Pilbara region.

Appendix 1: Organisations Represented

AARNet
AECOM
Australian Research Data Commons (ARDC)
Atlas of Living Australia (ALA)
Aurecon
AuScope
Australian Access Federation (AAF)
Australian Plant Phenomics Facility (APPF)
Australian Urban Research Infrastructure Network (AURIN)
AWS
BHP
Bioplatforms Australia (BPA)
Chamber of Minerals and Energy
CRCTime
CSIRO
Department of Biodiversity, Conservation and Attractions (DBCA) WA
Department of Climate Change, Energy, the Environment and Water (DCCEEW)
Deloitte
Department of Education
Department for Environment and Water (DEW) SA
Department of Jobs, Tourism, Science and Innovation (DJTSI) WA
Department of Primary Industries and Regional Development (DPIRD) WA
Department of Water and Environmental Regulation (DWER) WA
Edith Cowan University
Western Australian Environmental Protection Authority (EPA)
Eratos
FrontierSI
Google Australia
Human Centred Design Group
Institute for Marine and Antarctic Studies (IMAS) / University of Tasmania

AARNet
Integrated Marine Observing System (IMOS)
James Cook University
Microsoft
Minerals Council of Australia
National Sea Simulator (SeaSim)
NRM WA
Pawsey
Rangelands NRM
Rio Tinto
Sentient Hubs
Terrestrial Ecosystem Research Network (TERN)
Taskforce on Nature-related Financial Disclosures (TNFD)
University of Western Australia
WA Chief Scientist
WA Museum
Western Australian Biodiversity Science Institute (WABSI)
Western Australian Marine Science Institution (WAMSI)
Westport

Appendix 2: List of speakers

- Dr Simon Barry, CSIRO
- Emma Burns, Microsoft
- Jane Coram, Environment Information Australia
- Dr Ian Cresswell, WABSI and WAMSI Board Member
- James d'Ath, Taskforce on Nature-related Financial Disclosures
- Professor Allan Dale, James Cook University
- Chris Gentle, WABSI
- Rosie Hicks, ARDC
- Hamish Holewa, ARDC
- Dr Beryl Morris, TERN
- Professor Owen Nevin, WABSI
- Juanita Pettit, Australian Government Department of Climate Change, Energy, the Environment and Water
- Gavin Price, Minerals Council of Australia
- Dr Tim Rawling, AuScope
- Dr Tom Remenyi, Eratos
- Simon Taylor, WA Department of Water and Environmental Regulation
- Professor Matthew Tonts, Chair of WA Environmental Protection Agency
- Dr Luke Twomey, WAMSI
- Dr Andre Zerger, Atlas of Living Australia

Appendix 3: Workshop Agendas

Tuesday 6 June 2023 - Forum and Reception

Time	Session
1030-1100	**Morning tea on arrival**
1100-1245	**Pre-forum briefing** Introduction to Planet RDC & WA initiatives - *SAFE 2.0* guide, Shared Environmental Analytics Facility *(SEAF*), Cockburn Sound, and Pilbara
1245-1400	**Lunch**
1400-1410	**FORUM introduction**
1410-1550	**Trusted Environmental Data and Information Supply Chains: why they are critical for the government, industry and research sectors**
1550-1610	**Afternoon tea**
1610-1655	**What do we need to do to enable Trusted Environmental Data and Information Supply Chains by 2026?** 45 mins Q&A with expert panel
1655-1700	**Closing remarks**
1700-1900	**Reception**

Day 2 - Wednesday 7 June 2023 - Workshop

Time	Session
0900-0915	**Introduction**
0915-1000	**Opening remarks**
1000-1020	**Overview of workshop task** - Workshopping current constraints and opportunities to operationalise robust, repeatable and sustainable environmental data and information supply chains - Using the SAFE 2.0 framework (Culture, Collect, Curate, Integrate, Analyse 'layers') to map the current state of Environmental Data and Information Supply Chains - Assessing the maturity of the elements, and what can be done to lift the capability
1020-1100	**CULTURE**
1100-1130	**Morning tea**
1130-1210	**COLLECT**
1210-1250	**CURATE**
1250-1340	**Lunch**
1340-1420	**INTEGRATE**
1420-1500	**ANALYSE**
1500-1530	**Afternoon tea**
1530-1600	**Summary of findings and next steps in collaborative approaches to building trusted environmental data and information supply chains**
1600	**Close**