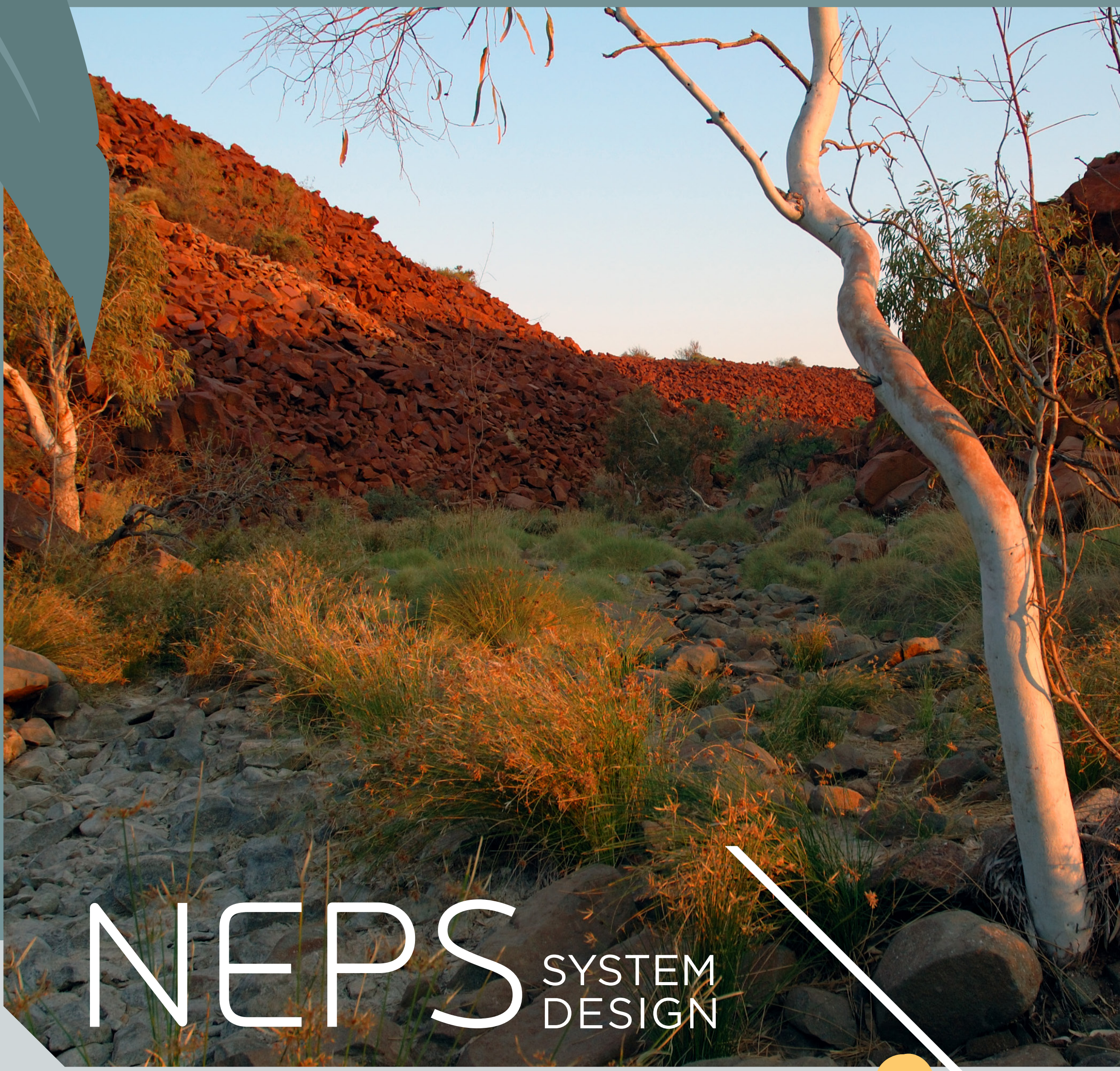
NOVEMBER 2019



NEPS SYSTEM DESIGN

CONTENTS

NEPS SYSTEM DESIGN

Executive Summary

The 2016 National Research Infrastructure Roadmap identified that investment is needed in infrastructure that integrates environmental observations with predictive modelling to improve environmental management and knowledge to benefit natural resource management in Australia. This collective capability was referred to as an integrated national environmental prediction system or NEPS. This Scoping Study retains the use of the term NEPS, acknowledging the name may change to best reflect the final agreement on infrastructure purpose and function.

This document is a product of the NEPS Scoping Study as proposed under the 2016 National Research Infrastructure (NRI) Roadmap[1] and included within the 2018 Research Infrastructure Investment Plan[2]. It expands upon the NEPS System Design Outline document submitted in August 2019.

NEPS is conceived as a networked or federated form of national research infrastructure, enabling integration of environmental observations with predictive modelling to produce data and information products and services to enhance Australian research capability and to support decision-makers and the economy with improved environmental knowledge and insight.

The System Design for NEPS needs to address the mechanisms required to support the necessary networking or federation across existing and future national research infrastructures (NRIs) and other significant national and State and Territory environmental data assets and to deliver the complementary capability required to facilitate integrated access and modelling based on these infrastructures and assets. NEPS will make significant use of existing capabilities and as far as possible reinforce existing NRIs to deliver components that are currently absent or insufficient.

Researchers and operational users of environmental data are expected to benefit from simplified access to rich cross-domain data on the Australian environment, with greater transparency, reproducibility and reuse for results. Contributing NRIs are expected to benefit through greater and more sophisticated use of their digital assets and from improved understanding of priorities for new data capture and enrichment.

The intent of the document is to establish a model for implementing NEPS as a synergistic addition to the landscape of Australian environmental research infrastructures. The design presented here seeks to address the following goals:

- To enhance the value and maximise reuse of existing environmental research infrastructure investments
- To facilitate research and outcomes reliant on cross-domain integration of environmental data and informatics capabilities
- To stimulate increased alignment between different research domains

1 https://docs.education.gov.au/documents/2016-national-research-infrastructure-roadmap

2 https://docs.education.gov.au/documents/research-infrastructure-investment-plan

- in planning and implementing capture and management of environmental observations and measurements
- To develop the capability of Australian researchers to model complex environmental systems and to predict their state across space and time
- To facilitate access to environmental data, models and computational workflows by research, government and industry
- To support future relevant technology developments such as digitisation of assets

The design is an inclusive framework that will inform remaining activities within the NEPS Scoping Study and support prioritisation of options for inclusion in the NEPS Investment Plan to be submitted in May 2020.

The design includes three complementary architectural viewpoints that together form the basis for implementing NEPS, focusing particularly on the first two (since the technical architecture must align closely with the infrastructures of the NEPS partners):

- **Social Architecture** – a shared model, addressing governance, sustainability, culture, incentives and funding, to support collaboration between NCRIS environmental capabilities and other key stakeholders (NEPS Partner Facilities) in delivering NEPS as a federated system and to support strategic and technical decision-making to guide the implementation and evolution of NEPS
- **Information Architecture** – a generalised framework for NEPS Partner Facilities to share, integrate and re-use observations and measurements and other digital assets from all environmental domains and to support cross-domain alignment and interoperability of essential variables, data and models
- **Technical Architecture** – the engineering approaches, standards and architectural frameworks that will guide the technical implementation of NEPS

This document draws on best-practice international models and Australian exemplars, assuring the widest possible interoperability with relevant environmental and geospatial data initiatives, in particular:

- The Reference Model for Open Distributed Processing (RM-ODP)[3], which establishes the architectural viewpoints for design of open distributed processing solutions. The focus in this document is on the RM-ODP Enterprise Viewpoint (as Social Architecture) and Information Viewpoint (as Social Architecture). The RM-ODP Computational, Engineering and Technical Viewpoints are here combined as a single Technical Architecture.
- The *Guidelines for the development of a Data Stewardship and Governance Framework for the Agricultural Research Federation* (AgReFed)[4], which applies the RM-ODP model in a context closely parallel to NEPS.

3 http://www.lcc.uma.es/~av/RM-ODP/index-original.html

4 https://www.agrefed.org.au/resources/EP192770.pdf

- The *Spatial Information Services Stack* (SISS)[5], which has served as the spatiotemporal data framework used by projects across multiple Australian Government agencies and research organisations and was adopted for the technology viewpoint in the context of the *National Environmental Information Infrastructure (NEII) Reference Architecture*[6].

This document serves two primary purposes. First, it offers an outline for the Department of Education (DoE) of the expected scope and direction for further work in developing the NEPS scoping study. Secondly, it serves as the basis for the next stage of community consultation to develop the necessary national consensus around a NEPS proposal. Annex A summarises findings from the first stage of stakeholder consultation in support of the NEPS scoping study. Annex B describes the planned activities and timeline for the remainder of the scoping study period.

5 https://www.seegrid.csiro.au/wiki/Siss/WebHome

6 http://www.neii.gov.au/system/files/filedepot/1/NEII%20Reference%20Architecture.pdf

Background

The 2016 National Research Infrastructure Roadmap identified that investment is needed in infrastructure that integrates environmental observations with predictive modelling to improve environmental management and knowledge to benefit natural resource management in Australia.

Informed by and in response to the 2016 Roadmap, the Australian Government released the 2018 Research Infrastructure Investment Plan that outlined future research infrastructure projects against current emerging Government priorities. The 2018 Investment Plan recommended the undertaking of a scoping study to articulate a value proposition and establishment plan for the proposed NEPS.

The Australian Government DoE has commissioned the NEPS Scoping Study to provide the technical assessments and requirements analysis for a NEPS, and to define implementation costs and timeframes to establish and manage a NEPS as national research infrastructure to meet researcher and operational user needs. The NEPS Scoping Study involves undertaking targeted consultations with key experts and stakeholders, including relevant areas of the existing National Collaborative Research Infrastructure Strategy (NCRIS) network.

The Australian Government has established an Expert Panel to conduct the NEPS Scoping Study. The Chair of the Panel is Professor Rob Vertessy and the members are Professor Bronwyn Harch, Dr Andrea Hinwood, Dr Adam Lewis, Dr Phil McFadden AO, Mr Warwick McDonald and Dr Steve Morton. The Terrestrial Ecosystem Research Network (TERN), an NCRIS-funded initiative headquartered at the University of Queensland, provides support to the Expert Panel.

The NEPS Scoping Study has two key objectives, namely to:

- obtain broad agreement from key stakeholders regarding the scope of a NEPS; and
- develop a detailed establishment plan, including identification of stakeholder co-investments and actions necessary to support the development and maintenance of a NEPS.

Basic principles for developing NEPS System Design

The outline presented here incorporates the thinking of the Expert Panel (EP) and the inputs received to date from stakeholder consultations. A key aspect of early discussions has been to clarify the scope under consideration for possible implementation as NEPS. The following principles are considered foundational for development of the System Design.

1. The primary focus for NEPS will be to deliver world-class research infrastructure. However, the work of environmental researchers and the expected outputs from NEPS have significant value for decision-making and regulatory uses. As a national capability, it is appropriate for these operational uses to be considered key use cases that the system should ideally support. Likewise, the capabilities foreseen for NEPS will offer significant benefits and possible applications for commercial users. These different channels present the opportunity for diversified long-term funding to support the costs of NEPS.
2. In the context of NEPS, "national" is considered to include capabilities that integrate data and meet the needs of researchers and consumers of environmental information at all scales from continent down to paddock, including the scale of States and Territories.
3. In the context of NEPS, "environment" is taken to encompass all biotic and abiotic factors that may vary over time and space, in both natural and artificial systems (cities, agriculture, etc.) and across terrestrial, freshwater and marine areas, with a focus on how these factors affect the ability of people, crops, livestock and biodiversity to survive and exploit these systems. Relevant factors include (but are not limited to) climate, geology, soils, biodiversity, land cover, land use, hydrology, topography, infrastructure, air quality and pollution, at all scales from global through continental to local.
4. "Prediction" is understood to include the ability to forecast future states of the Australian environment. However, delivering such capability will depend on NEPS having a prior and more general capability to offer modelled predictions of the state of environmental variables in the present or recent past. Such a capability, spanning multiple environmental domains and operating across a range of spatial and temporal scales would itself be a major asset for Australian researchers and for use in operational decisions and policy development. The scope for NEPS is accordingly considered to encompass this broader capability.
5. NEPS should build on, rather than duplicate, existing data services offered by NCRIS or other national capabilities, including TERN, ALA, IMOS, AuScope, AURIN, BOM, and GA. Each of these capabilities already addresses data standards and integration for one or more subdomains of environmental information. In most cases, these facilities also include modelling activities relating to the states and trends associated with these subdomains. In the same way, NEPS should as far as possible integrate the

resources and models offered by ARDC and NCI rather than developing new compute resources or standards.

6. NEPS should complement any other similar initiative that occurs over the present planning horizon, including the ACCESS Scoping Study. ACCESS is Australia's weather, climate and Earth System modelling system. It is the foundation research infrastructure for Australian weather and climate science in areas that carry significant implications for Australia, such as weather forecasting, seasonal prediction, decadal prediction, climate projection and extremes including cyclones, droughts, floods, bushfires and heatwaves.

   ACCESS-NRI has relevance to the NEPS activity in that its priorities include:

   - To establish Australia's international leadership in weather, climate and Earth System modelling and provide a coherent, planned and strategic focal point for the national research effort that addresses weather, climate, water, natural disaster and environmental change risk.
   - To optimally use Australian investment in high performance computing involving nationally significant reference data from national and international observational networks.

7. NEPS must show a clear value proposition with demonstrated benefits for the Australian research community and with collateral benefits for operational users and for Australia's positioning as an international leader in digital environmental research. It is impossible to determine the extent of these benefits in the absence of a straw-person design that can support concrete discussions around detailed scope and priorities and around possible modules that build on the core deliverables. This outline, to be followed by the System Design, will serve this purpose and support the next round of NEPS consultations.

8. Achieving the cross-domain goals of NEPS and maximising the reuse of digital assets shared or developed within NEPS is best supported by following international and national best practices for data management, description, licensing and provenance. In particular, NEPS should develop and support clear policies to ensure to the fullest extent possible that data are open and conform with the FAIR principles[7] and that Australia's environmental data assets are preserved in systems that as far as possible satisfy the CoreTrustSeal Data Repositories Requirements[8].

9. NEPS must support emerging innovation based on new technology, or through new business models or forms of economic or social organisation, including the rise in digitisation of assets and new ways of capturing data. Advancement of technology and computation capabilities has already fuelled the growth of using non-intrusive sensors (audio, image and video) to monitor biodiversity and other biological and non-biological environmental parameters. The NEPS design will anticipate novel sources of data, which will be processed in new ways to make them suitable for analysis and ready to use in environmental prediction modelling frameworks.

7 https://www.go-fair.org/fair-principles/

8 https://www.coretrustseal.org/why-certification/requirements/

Elements of NEPS System Design

Given the maturity and scale of Australia's existing environmental NRIs, it would be inefficient and unreasonable for NEPS to replicate the capabilities that they offer or to construct a data warehouse that spans all categories of Australian environmental data. NEPS must be implemented as a federated cross-domain solution that will rely on best-practice models for open distributed processing (as described in the RM-ODP framework).

Delivering an open distributed system requires a design that provides clarity and facilitates cooperation between multiple stakeholders. RM-ODP addresses this requirement by specifying five discrete but comprehensive viewpoints for exploring system design. The present document focuses on two of these. The Enterprise Viewpoint is here presented as a **Social Architecture** that addresses the mechanisms by which NEPS Partner Facilities will collaborate. The Information Viewpoint (here **Information Architecture**) provides a lifecycle perspective for cross-domain environmental data products and models. The remaining viewpoints (Computational, Engineering and Technical) are here addressed more briefly as a **Technical Architecture**, addressing the models and key standards that NEPS should adopt to integrate the capabilities of the existing NRIs.

Figure 1 represents the focus of and the inter-relationships between the three architectural levels.

The technical, information and social architectures for NEPS each need to be appropriate and supportive for the involvement of all relevant stakeholders, particularly the existing NCRIS environmental capabilities and governmental data infrastructures. The consultation activities under the NEPS scoping study aim to clarify expectations and develop an agreed architecture at each of these levels.

Although these three architecture levels are interdependent, the primary

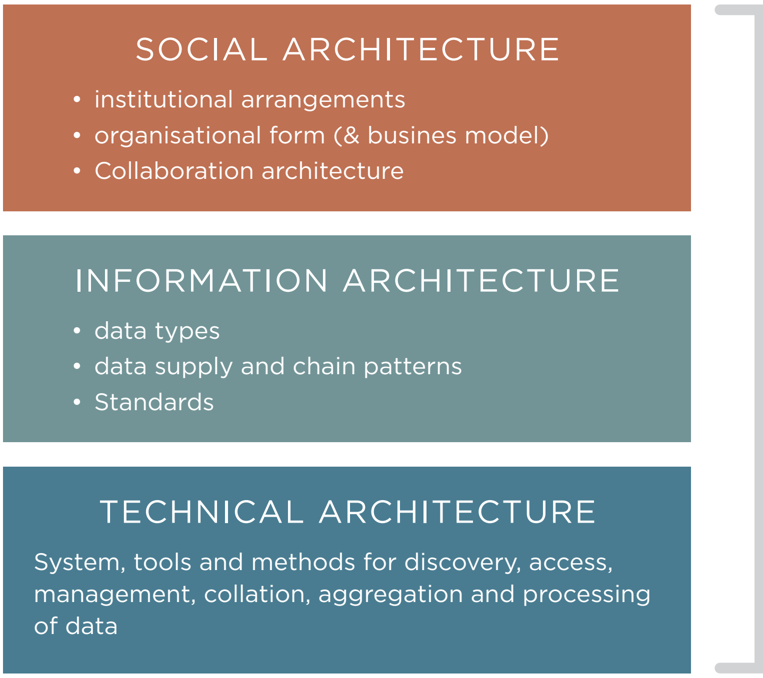


Figure 1
NEPS Architectural levels (adapted from a figure by Paul Box & David Lemon, CSIRO L&W)

focus for initial design must be on the information architecture. As a national research infrastructure, NEPS must facilitate access to information assets that expand the capabilities of the Australian research community and that reinforce the interconnectedness of the NCRIS landscape. This outline therefore focuses primarily on this level.

If the information architecture for NEPS can be defined and agreed and the benefits from this architecture can be shown, this will provide the foundation for detailed discussion both of the necessary social architecture to provide the delivery framework for NEPS and of the technical architecture that will make optimal use of the resources and strengths of existing infrastructures.

The following sections define in more detail the social architecture, provide the requirements for the NEPS information architecture and identify key areas of the associated technical architecture.

Social Architecture

Community structure

The Social Architecture relates to the organisational, business and social context within which NEPS must exist. It aims to clarify the purpose of the system and how it relates to the interests of its stakeholders. At the same time, it defines the roles with which these stakeholders will interact with NEPS and how they will collaborate to ensure the success of the system.

The success of NEPS is dependent on the alignment between the purpose and goals of the system and those of the partners collaborating to deliver it. In particular, it depends on openness and inclusivity that grants these partners a real stake in the course of NEPS and its outputs and enables them to operate as a **community**, i.e. a set of participants working together to achieve shared objectives. The behaviour of a community is shaped and guided by the definition of policies (e.g. related to data standards or decision-making) that define how those holding different roles interact to carry out processes within the community, as illustrated in Figure 2.

NEPS occupies a position at the intersection between multiple existing communities that support and deliver Australian environmental research infrastructures and multiple communities that make use of environmental data and are expected to benefit from NEPS deliverables. These different communities are not fully independent from one another and overlap significantly in participation. It is essential to consider how each of these existing communities will be represented within NEPS and how NEPS policies can use these relationships to deliver the greatest possible benefit. NEPS should also serve as the context in which strategies (e.g. decadal plans) spanning the whole environmental domain can be developed and implemented.

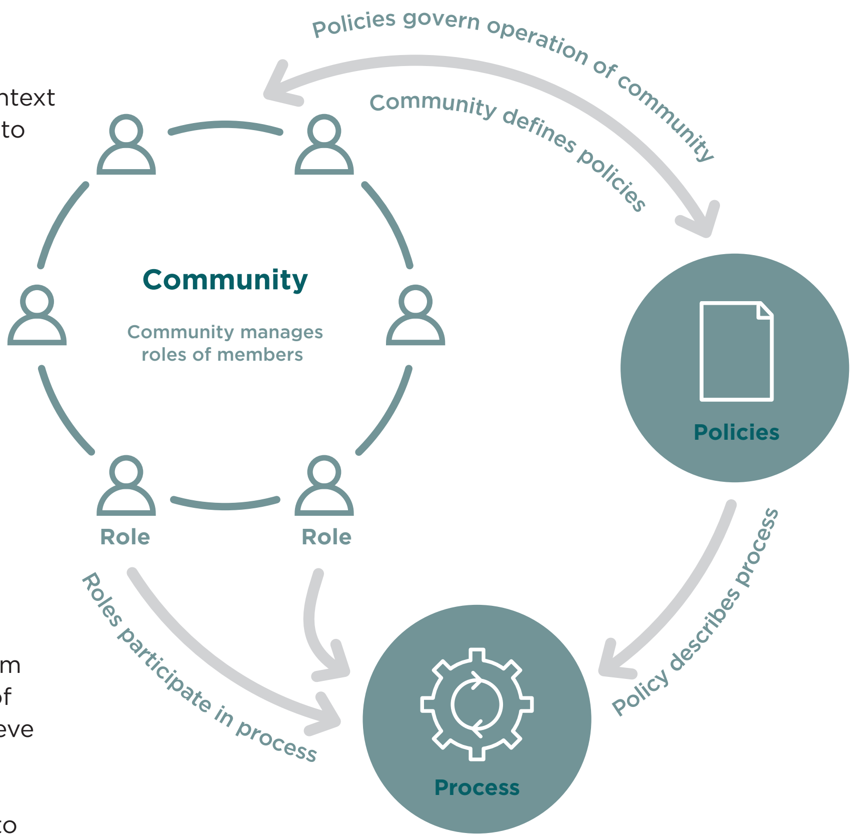


Figure 2
Roles, processes and policies in the context of a community (adapted from Box et al. 2019, Guidelines for the development of a Data Stewardship and Governance Framework for the Agricultural Research Federation (AgReFed))

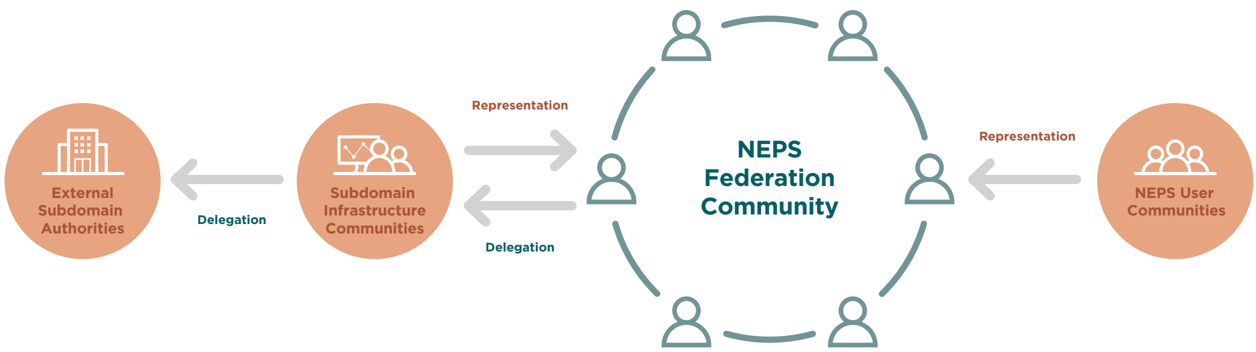


Figure 3 Enterprise view of communities involved in NEPS

Figure 3 is a high-level view of the key community relationships for NEPS.

Subdomain Infrastructure Communities are the major Australian environmental research infrastructures which exist today, including the NCRIS infrastructures (especially the capabilities with significant spatial capabilities, including ALA, AURIN, AuScope, IMOS and TERN) and other significant national infrastructures (particularly BOM and GA, but engagement is also sought with AgReFed, SmartSat CRC and others). These communities each serve already as coordination and integration points at the continental scale for major subcomponents of environmental information (climate, oceans, geology, soils, hydrology, terrestrial ecosystems, biodiversity, land use, agriculture and urban environments), referred to here as “subdomains”. Research communities and agencies at the Federal and State levels participate significantly in each of these Subdomain Infrastructure Communities. NEPS should not duplicate the work of the Subdomain Infrastructure Communities. Instead NEPS can rely on each of these as a hub for coordinating standards, data and models for its own set of environmental subdomains.

Each Subdomain Infrastructure Community should identify and participate in relevant international bodies that set standards, promote interoperability and coordinate large-scale programmes relating to its environmental subdomains. These bodies are represented here as **External Subdomain Authorities**. Each Subdomain Infrastructure Community effectively delegates some responsibility to External Subdomain Authorities when it commits to adopt and implement relevant standards. NEPS can leverage these relationships and rely on the expertise of each Subdomain Infrastructure Community to mediate relevant standards for each subdomain and to address any specific Australian requirements in the development of these standards. Hence NEPS itself will delegate responsibility to each Subdomain Infrastructure Community to handle these decisions. The modularity of this approach significantly reduces the risks associated with planning and delivering NEPS.

The **NEPS Federation Community** is a federation of the capabilities of the Subdomain Infrastructure Communities that will enable them to coordinate their efforts across different subdomains and deliver the additional elements that are necessary to deliver capabilities at the level of the Australian environment as a whole. The NEPS Federation Community will delegate responsibility to each Subdomain Infrastructure Community to coordinate and deliver standards, data and models for its subdomains. Additionally, each of the Subdomain Infrastructure Communities must be represented in both strategic and technical decision-making processes within the NEPS Federation Community.

Some aspects of standards adoption and development span multiple or all environmental subdomains. This is likely to be increasingly the case as more countries invest in environmental information and prediction. In these cases, the NEPS Federation Community may assume responsibility as the interface to relevant international standards-setting and coordination activities, which function as **External Domain Authorities** similar to the External Subdomain Authorities. Where appropriate this responsibility may still be mediated through established relationships with a Subdomain Infrastructure Community. By either approach, NEPS will delegate some responsibility to these External Subdomain Authorities.

NEPS User Communities will also be diverse, including research groups spanning all aspects of environmental science, government and regulatory users, industry stakeholders and others. As NEPS develops, it is important that the expertise of these communities, along with their needs, is represented in establishing the technical directions and priorities for the NEPS Federation Community. The NEPS Federation Community will require processes to recognise important NEPS User Communities and to give these a role in the steering of the NEPS infrastructure.

This System Design does not seek to define the full extent of the potential membership of the NEPS Federation Community. The focus here is on defining the model that will serve as a framework for environmental research infrastructures to participate as Subdomain Infrastructure Communities and the basic principles for representation by NEPS User Communities. The set of Subdomain Infrastructure Communities that will establish NEPS will be developed and refined during the remainder of the Scoping Study period and presented within the Investment Plan in 2020, along with a set of critical or high-impact use cases. The Investment Plan will also refine the categorisation of NEPS User Communities and their contributions and roles within the NEPS Federation Community. Formalisation of relationships with External Domain and Subdomain Authorities will take place as part of the operationalisation of NEPS.

Governance roles within NEPS Federation Community

Figure 4 represents the governance roles responsible for steering the NEPS Federation Community.

The **NEPS Council** is responsible for strategic and business decision making for NEPS. The **NEPS Technical Committee** reports to the council and provides technical decisions about common technology choices, data standards, service level criteria, etc. The NEPS Council will include representatives from each Subdomain Infrastructure Community. The NEPS Technical Committee will include representatives from Subdomain Infrastructure Communities and NEPS User Communities. Other parties may also be represented on these bodies in accordance with the policies defined once NEPS is established.

The NEPS Council and the NEPS Technical Committee together are responsible for establishing, maintaining and, if appropriate, terminating NEPS. They steer the federation and ensure that it operates in accordance with its mission and funding and meets the needs of the Australian research community and other stakeholders.

By establishing the policies and defining the processes within the NEPS Federation Community, these bodies oversee and enable the operational relationships and interactions between Subdomain

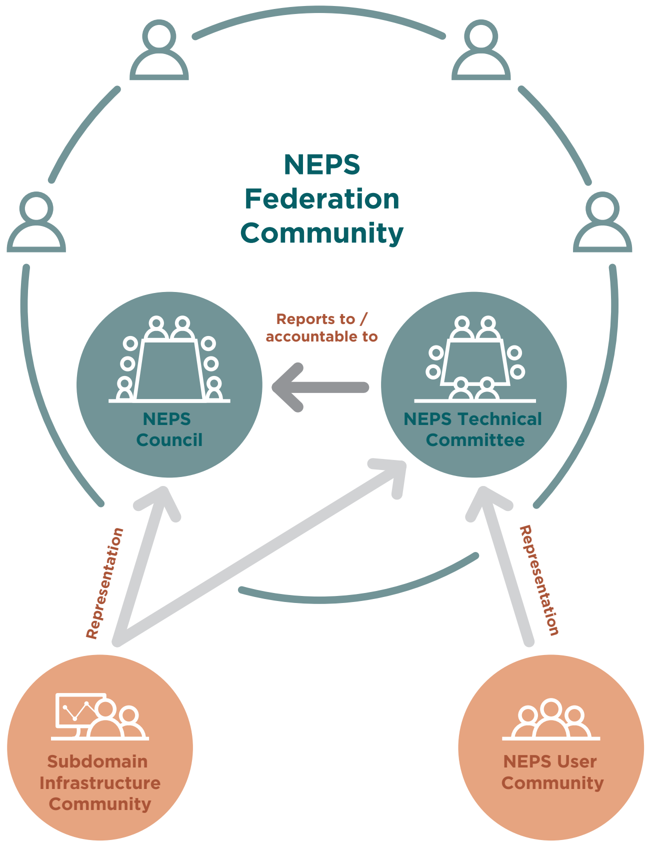


Figure 4 Roles within the NEPS Federation Community - NEPS Council and NEPS Technical Committee

Infrastructure Communities and NEPS User Communities to achieve community objectives.

Initial settings for the roles and accountabilities of the NEPS Council and NEPS Technical Committee are presented in *Annex C - NEPS Federation Community – roles and accountabilities*.

Operational roles within NEPS Federation Community

The NEPS Council and NEPS Technical Committee have responsibility for the decisions around policies and processes to steer NEPS. The NEPS processes established by these governance bodies will define the operational roles assumed by NEPS stakeholders as they interact to achieve community objectives.

The set of operational roles required by NEPS will be refined and is likely to expand over time. However, defined roles will be necessary and representatives for these roles should be nominated by each Subdomain Infrastructure Community to ensure alignment between data assets and data management activities from contributing infrastructures, particularly in the following areas:

1. Registration of data standards and best practices for management of data contributing to the subdomain
2. Registration of Essential Environmental Variables for the subdomain
3. Liaison with External Subdomain Authorities
4. Assurance that subdomain digital assets comply with NEPS FAIR Data Policy
5. Assurance that subdomain data publication and repositories comply with NEPS Trusted Repository Policy

These roles are essential to ensure the seamless operation of NEPS as a federation of subdomain infrastructures. Aspects of NEPS operations that relate to cross-disciplinary integration and modelling are under the oversight of the NEPS Technical Committee.

NEPS policies

Although NEPS policies will be developed in detail, partly within the Investment Plan and partly as the NEPS Infrastructure Community is implemented, two key elements can be identified as foundational policies to align the practices and approaches of Subdomain Infrastructure Communities both to meet NEPS objectives and to enhance contribution to broader research infrastructure needs. These two policies are 1) a NEPS FAIR Data Policy that defines the qualifying thresholds for data publishing in accordance with the FAIR principles and 2) a NEPS Trusted Repository Policy for sustainable management of digital assets in accordance with CoreTrustSeal requirements. Initial settings for these two policies are presented in the *Annex D - NEPS FAIR Data Policy – initial settings and Annex E - NEPS Trusted Repository Policy – initial settings*.

Alignment and synthesis

Since NEPS will be established as a broad cross-domain community, there will be an ongoing requirement for representatives from different environmental research fields to collaborate with one another and with other stakeholders to develop standards, projects and plans that will advance the interoperability and integration of their efforts.

Between 2009 and 2014, TERN operated the Australian Centre for Ecological Analysis and Synthesis (ACEAS)[9], which served effectively to enable scientists and managers in Australia and internationally to develop evidence-based environmental management strategies and policy at regional, state and continental scales. A similar Synthesis Centre should be a core component of NEPS, with the resources to convene technical workshops both to support the operation of the NEPS Federation Community (and especially the NEPS Technical Committee) and also to host merit-based selection and hosting of synthesis activities proposed through one or more Subdomain Infrastructure Communities and User Communities.

9 https://www.tern.org.au/Australian-Centre-for-Ecological-Analysis-and-Synthesis-ACEAS-pg17735.html

Information Architecture

A system of systems

The capabilities expected from NEPS centre on the ability to integrate, organise and analyse data from a wide range of sources to understand the patterns and dynamics of different components of the environment and the interactions between these components. Most environmental data sets represent attempts to measure aspects of the environment (a set of variables) at a particular time and place or to model these aspects at particular times and places (usually as continuous modelled surfaces).

An early stage in implementing NEPS will be a review of technical standards among the Subdomain Infrastructure Communities to develop a consistent and comprehensive reference architecture for discovery, access and use of environmental information from all subdomains. The NEII Reference Architecture will serve as a starting point for this exploration, although this will require extension to accommodate additional data types, particularly for non-spatiotemporal information. The NEPS Synthesis Centre will support the necessary discussions.

In recent years, there have been significant efforts globally to develop standards that can deliver an information architecture for major components of environmental data. The success of the Essential Climate Variables (ECVs) has driven other efforts to define and document sets of essential variables, e.g. Essential Ocean Variables (EOVs) and Essential Biodiversity Variables (EBVs). The Essential Environmental Measures for Australia program[10] was a national effort to explore needs at the national level for similar standardisation. This program was an effective tool for cross-disciplinary exploration and alignment between Australian environmental data holders, but there is no current mechanism to continue this work. This is another area in which a NEPS Synthesis Centre will be important, providing the framework and support mechanisms for discussions and reports.

In all cases, these efforts have sought to develop clear but flexible models to guide and prioritise *in situ* and remote-sensed earth observation activity, to structure efforts to integrate data from multiple sources, and to support modelling efforts around complex Earth systems. Such effort by multiple research communities is a prerequisite for delivering large-scale cross-domain integration of environmental data, for example as proposed for GEO's Global Earth Observing System of Systems (GEOSS)[11] and as required to assess baselines, monitor change and plan strategies in regard to international

10 https://doi.org/10.25919/5b7f04377bbfd

11 https://www.earthobservations.org/geoss.php

targets, including the Sustainable Development Goals (SDGs)[12], the Strategic Plan for Biodiversity[13] and the Sendai Framework for Disaster Risk Reduction[14].

All of these initiatives seek to characterise a dynamic system through repeated collection of standard observations and measurements in space and time and through modelling to remove errors and bias and to create consistent data products. The primary focus is usually on representing past and present states of the environment at the best possible scale and resolution, both spatial and temporal. However, the same variables and models are also fundamental for modelling future states of the environment under diverse scenarios.

It should be clear that, for any environmental variable, a continuum may exist from 1) primary observations and measurements using particular methods and devices, through 2) systematically normalised, aggregated and cleaned observations and measurements, to 3) modelled estimates of the most likely values for the variable across space and/or time. It is important to register and track data sets created at each stage in this process. Primary observations remain the evidence on which all subsequent modelling depends. Research infrastructures must document the provenance and interdependence of these digital assets and the models that are used to produce more derived assets.

Australia has made significant investments in data infrastructures that are each responsible for data standards, data aggregation and in many cases modelling for particular environmental subsystems, including climate, oceans, geology, soils, hydrology, terrestrial ecosystems, biodiversity, land use, agriculture and urban environments.

These investments are foundational for NEPS and simplify the task of delivering an integrated environmental information and prediction system. The appropriate model will be for each of these existing infrastructures to serve as a hub to stimulate standardised observing efforts and to aggregate, normalise and clean data representing a subsystem. NEPS will be a collaborative activity to provide cross-domain integration and services across all these subsystems and to facilitate interoperability and modelling that depends on enhancements to the deliverables from multiple hubs.

There have been associated advances in Australia in implementing virtual laboratories to deliver modelling and predictive tools in multiple environmental subdomains. Examples include Ecocloud[15], the AuScope Virtual Research Environment (AVRE)[16], the Marine Virtual Laboratory (MARVL)[17], the AURIN Portal[18], Collaborative Environment for Scholarly Research and Analysis (CoESRA)[19] and the Biodiversity and Climate Change Virtual Laboratory (BCCVL)[20]. Each of these includes models, components, expertise and experience in interface design that can contribute to the development of NEPS as a cross-domain predictive system.

As described under *Social Architecture* above, each of these existing infrastructures will operate as a Subdomain Infrastructure Community within NEPS, with responsibility for coordinating technical approaches, data management and modelling related to its information subdomain. The NEPS Federation Community will coordinate the alignment and integration of these communities and will be responsible for the additional components required to deliver the cross-domain capabilities of NEPS. The NEPS Synthesis Centre will support the explorations required to achieve this.

12 https://sustainabledevelopment.un.org/
13 https://www.cbd.int/sp/
14 https://www.unisdr.org/we/coordinate/sendai-framework
15 https://ecocloud.org.au/
16 https://www.auscope.org.au/avre
17 http://imos.org.au/facilities/aodn/imos-data-management/marvl/
18 https://aurin.org.au/resources/aurin-portal/
19 https://coesra.tern.org.au/
20 http://bccvl.org.au/

National Prediction System

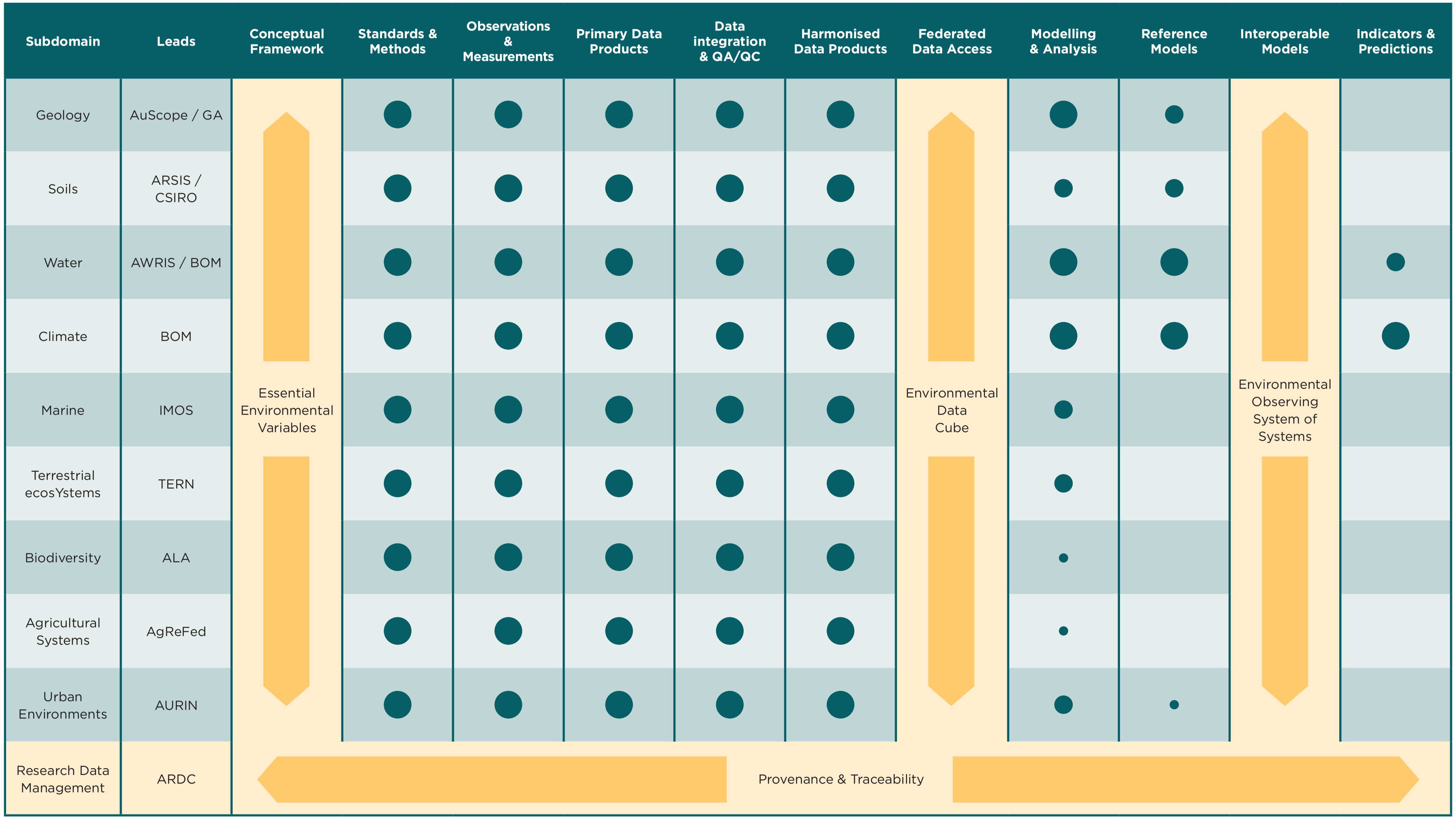


Figure 5 Australian research infrastructures as Subdomain Infrastructure Communities within NEPS

Current level of activity within lead initiatives for subdomain:

- Major Focus
- Significant activity
- Some activity
- Yellow represents additional coordination and investment required for the delivery of NEPS

Figure 5 provides a high-level perspective on the embedding of Subdomain Infrastructure Communities within NEPS and highlights the areas for which additional NEPS investments will be necessary.

Each row in the figure represents an information subdomain, associated with lead infrastructures expected to serve as Subdomain Infrastructure Communities within NEPS.

The green cells in each row are indicative of the current level of activity or investment by these lead infrastructures in different stages of the research data lifecycle, either directly as part of the operational activity of the infrastructure community or via contributions from the community members. All infrastructures invest in standardisation and research data management, but there is significant variation in the maturity of associated modelling capabilities and deliverables. This variation is a result of the different levels of complexity represented by each subdomain and of the degree to which societal requirements have driven effort to develop standard models and indicators (particularly in the context of climate and water). It should be noted that the levels of activity shown are purely indicative and do not necessarily reflect all relevant activity across the Australian research and environmental monitoring landscape.

The yellow cells and arrows represent the aspects for which additional cross-domain coordination and investments will be required to deliver NEPS. These aspects are the focus for the NEPS Federation Community, building on the expertise of each Subdomain Infrastructure Community and aligning with ARDC recommendations on best practice for research data management.

The elements represented by the columns are as follows:

- **Conceptual framework**
  The representation of environmental states required to underpin cross-domain environmental information management and modelling. Shared community understanding of these states and the associated variables is a prerequisite for delivering NEPS. This element corresponds to the former Australian Essential Environmental Measures program and should be taken up by NEPS as a responsibility for the NEPS Technical Committee. Each Subdomain Infrastructure Community will be responsible for monitoring and aligning with international efforts to establish Essential Environmental Variables relevant to its subdomain. However, NEPS will coordinate between subdomains to ensure clarity and to maximise interoperability.
- **Standards & methods**
  Subdomain Infrastructure Communities are best positioned to monitor and respond to the national and international context around standards development for their subdomains and around developments in sensors and observational tools and data representation. NEPS will delegate responsibility for this element to each Subdomain Infrastructure Community. However, NEPS will co-ordinate the harmonisation of vocabularies to standardise the representation of data and methods so that cross-disciplinary researchers can understand and use artefacts across the entire environmental space.
- **Observations & measurements**
  The streams of environmental data captured via sensors, field researchers, citizen science and other sources, including remote sensing. Not all subdomain infrastructures invest directly in collecting observations and measurements. A large proportion of all data is collected and shared with these infrastructures as contributions from data providers. Subdomain Infrastructure Communities are well positioned to serve as the hubs for coordinating these efforts for their subdomains. NEPS will delegate responsibility for this element to each Subdomain

Infrastructure Community but will coordinate use of vocabularies that describe the platforms, instruments, individuals, and locations associated with data collection to resolve ambiguity and harmonise different terms.

- **Primary data products**
  The representation of observations and measurements as published datasets deposited in appropriate subdomain repositories. NEPS will delegate responsibility for this element to each Subdomain Infrastructure Community.
- **Data integration and QA/QC**
  The validation and normalisation of primary datasets to integrate and standardise data in formats that meet the needs of user communities and that provide assurance of the quality or assessment of issues surrounding the data. These processes are specific to each subdomain. NEPS will delegate responsibility for this element to each Subdomain Infrastructure Community.
- **Harmonised data products**
  The delivery of quality-assessed aggregated data products suitable for use by most users, with provenance for primary data sources appropriately represented. Where appropriate, these harmonised data products are versioned reference data products that compartmentalise the complexity of observations and measurements and primary data products for most uses. NEPS will delegate responsibility for this element to each Subdomain Infrastructure Community but will promote efforts to deliver harmonised data products in support of community-recognised Essential Environmental Variables. A key outcome will be delivery of fit-for-purpose data products.
- **Federated data access**
  A fundamental capability required in support of cross-domain environmental data integration and access and to achieve the goals of NEPS for modelling and predicting environmental states. This includes standardised description of and access to a wide variety of digital assets, but most importantly to efficient infrastructure enabling access to harmonised spatiotemporal data products from the Subdomain Infrastructure Communities. Achieving this goal requires cross-domain standards for data representations and metadata and mechanisms to handle registration, interpolation and extrapolation of data across time and space. This element will be taken up by NEPS with governance overseen by the NEPS Council and technical models by the NEPS Technical Committee.

- **Modelling & analysis**
  Subdomain Infrastructure Communities invest to different degrees in developing models to represent the subsystems associated with their domains. NEPS will support the efforts of each of these communities to develop models to represent the environmental subsystems within their subdomain. In many cases, these models exploit data products from other domains as inputs. NEPS federated data access will simplify and standardise this usage and enable each Subdomain Infrastructure Community to benefit from the most current and best-validated representation of each variable. The Essential Environmental Variables included within the NEPS conceptual framework should aim to standardise the variables used in this way, thus supporting a modular approach to building and exploiting reference data products. NEPS will delegate responsibility for modelling activities to each Subdomain Infrastructure Community.

- **Reference models**
  In many cases, the products of the modelling and analysis activities of each Subdomain Infrastructure Community will themselves constitute valuable national reference data products. These should be integrated back into the NEPS federated data access for further reuse by Subdomain Infrastructure Communities, NEPS User Communities and elsewhere. In particular, they should be made accessible within NEPS to facilitate cross-domain model integration. NEPS will delegate responsibility for this element to each Subdomain Infrastructure Community but will promote efforts to deliver reference models for community-recognised Essential Environmental Variables.

- **Interoperable models:** Each Subdomain Infrastructure Community will focus on development of models specific to its subdomain (in many cases using harmonised data products from other subdomains). NEPS will address opportunities to develop truly cross-domain modelling and predictive capabilities. Developing a national predictive capability will require research and development both around workflows to integrate best-practice dynamic models for each subdomain (a system-of-systems approach to modelling) and around *de novo* development of cross-cutting models (including machine learning and dynamic state models). This element will be taken up by NEPS with governance overseen by the NEPS Council and technical models by the NEPS Technical Committee.

- **Indicators & predictions**
  NEPS is envisaged as a platform for delivering best estimates and predictions of the state of environmental variables. A key societal requirement is to exploit such estimates and predictions in evaluation of different future scenarios, based on agreed criteria (a value system for determining the merit or demerit associated with resulting states). These criteria, and the environmental-economical and other models used to evaluate them, are independent from the predictive outputs that NEPS will deliver. Subdomain Infrastructure Communities, along with relevant NEPS User Communities, are better positioned to provide the bridging capability from NEPS federated data access and interoperable models to derived indicators and environmental accounts. NEPS will delegate responsibility for these outputs to each Subdomain Infrastructure Community or NEPS User Community.

- **Provenance and traceability**
  Cutting across all aspects of the data and modelling lifecycle for all subdomains and for NEPS as a whole, there is a need for all stages from initial capture of observations

and measurements to deliver of indicators and predictions to be documented with machine-readable and human-accessible metadata to track provenance and support evaluation of the transformations and products. NEPS has a major requirement to develop and adopt cross-domain standards and best practices in this area. The NEPS FAIR Data Policy and NEPS Trusted Repository Policy play an important role in this respect. Additionally, NEPS will work closely with ARDC to ensure alignment with best practice across the whole Australian research environment and to facilitate reuse of NEPS developments elsewhere.

The following subsections will address in more detail the four areas requiring NEPS investments in cross-domain solutions.

Essential Environmental Variables

Developing a standardised conceptual framework for characterising environmental states is likely to be the single most significant activity for NEPS. This framework will build on international efforts to standardise Essential Variables for different subdomains, including the Essential Climate Variables (ECVs)[21], Essential Biodiversity Variables (EBVs)[22], Essential Ocean Variables (EOVs)[23], etc. GEOEssential[24] is an EU-funded consortium-based effort to align these activities. NEPS must participate in the development and alignment of essential variables, both at the level of environmental subdomains through the involvement of Subdomain Infrastructure Communities in relevant processes and through direct engagement at the cross-domain level. Relevant bodies will constitute External Domain and Subdomain Authorities in the model presented under *Social Architecture* above.

NEPS will serve as an interface between these diverse international efforts and the requirements of Australian research communities for standardised representation of environmental states. This will serve as a continuation of the focus of the former Australian Essential Environmental Measures Program. In

21 https://public.wmo.int/en/programmes/global-climate-observing-system/essential-climate-variables

22 https://geobon.org/ebvs/what-are-ebvs/

23 http://www.goosocean.org/index.php?option=com_content&view=article&id=14&Itemid=114

24 http://www.geoessential.eu/

accordance with the recommendations from the evaluation report[25] for the program, the term Essential Environmental Variables (EEVs) will be used to refer to the full set of environmental variables to be addressed.

EEVs have multiple roles within the NEPS framework. As NEPS evolves, the goal will be to align these roles to maximise synergistic benefits from the efforts of the whole environmental research community.

1. EEVs serve to define the goals for observations and measurements and to establish priorities for monitoring programs for new data capture and for digitisation of historical data assets
2. EEVs provide the focus for subdomain-level data integration and QA/QC, reducing the number of alternative representations that need to be maintained and facilitating review of processes and algorithms to aggregate data
3. EEVs provide the focus for delivery of national- and state-scale harmonised data products at best-possible spatial and temporal resolutions
4. EEVs ideally serve as the parameters for implementing modelled representations of environmental states – as NEPS matures, the cross-domain information needs for developing models for each subdomain should be formalised as reusable EEVs
5. EEVs are the focus for reference models to deliver the best possible estimates of environmental state
6. EEVs ideally serve as input parameters for environmental-economical models and for development of environmental indicators

The NEPS Technical Committee will have responsibility for coordinating development and registration of the set of EEVs required within NEPS and by NEPS User Communities. It will be the responsibility of individual Subdomain Infrastructure Communities to adopt or develop and to promote corresponding standards and best practices.

Federated data access

A key challenge for environmental data infrastructures is the immense range of scales (both spatial and temporal) and diversity of features and systems for which data are collected and organised. This challenge increases as data from different methods and sensors is combined and as models need to incorporate datasets from different domains and infrastructures. There is a fundamental and highly general need for tools that can reliably upscale and downscale data according to application needs and that can assess whether shifts in scale reduce information to unacceptable levels.

The integration of data for all environmental variables can be modelled as a space-time hypercube with a set of aggregated or modelled variables for each unit of space within a given time interval. Users need to the ability to access such a hypercube and process spatiotemporally integrated data both in gridded formats and in relation to spatial units of interest (shapes). Efforts such as the ECVs, EOVs and EBVs exploit this model of stackable spatiotemporal layers.

Developing NEPS as a generalised environmental information and prediction system will require the ability to deliver (whether centrally or via federation) such a hypercube with efficient capabilities to align, rescale, interpolate and extrapolate data and hence to offer estimates for a suite of environmental variables at any point in, or for any portion of, time and space (within appropriate limits).

Such a hypercube would have enormous value for many researchers and regulatory users and would be a valuable extension to existing NCRIS capabilities. It would be a prediction system in that it would predict the values of environmental variables for different times and places even in the absence of known measurements. It may be considered an extended version of the core that underlies many of the individual NCRIS or Federal capabilities.

Digital Earth Australia[26] is a significant Australian effort to deliver a gridded

25 https://doi.org/10.25919/5b7f04377bbfd

26 https://www.ga.gov.au/dea

environmental data cube. The existing virtual laboratories maintained by Subdomain Infrastructure Communities (including Ecocloud, AVRE, MARVL, AURIN Portal, CoESRA and BCCVL) also represent investments relevant to implementing and using this hypercube. Investing in reinforcing and extending these initiatives will help NEPS rapidly and sustainably to secure the required capability.

NEPS will also require access to non-spatial digital assets outside the scope included within the spatiotemporal data hypercube. An ideal modelling and prediction system for the environment should be able to exploit evidence from observed properties of the classes of entity within the environment, e.g. through access to field or lab measurements of thermal tolerance for a given species, life tables for populations of a species at well-studied sites, or the hydrological characteristics of different soil groups. Similarly, NEPS should be able to exploit evidence from field and lab studies on functional relationships between different variables in the system (response curves, etc.).

Non-spatiotemporal assets of relevance to NEPS include at least:

1. Vocabularies, ontologies, gazetteers and directories relating to environmental data
2. Attributes of entities (including species) that are independent of spatial locations but required to parameterise models
3. Process models, equations and other modular elements for modelling complex environmental systems

Some or all of these may be managed directly by one of the contributing infrastructures, but it is likely that NEPS will nevertheless need to include services that support consistent and integrated access to these resources. Such services will be developed in conjunction with ARDC to maximise broader interoperability around digital asset management within Australia.

It is not possible at this stage to determine the full scope of the digital assets required to implement NEPS. A modelling framework would assist with setting an agenda and priorities for collection of the traits that make most difference (and for phylogenetically informed use of default estimates for different taxa). Since the potential scope is enormous, NEPS should catalogue these assets based on actual need as part of defined solutions. Nevertheless, the need for access to diverse classes of non-spatiotemporal assets must be recognised from the beginning and addressed as part of the NEPS information architecture.

Across all of these assets, both spatiotemporal and non-spatiotemporal, the NEPS Federation Community will need to adopt or establish agreed metadata standards, directory services, brokering systems, etc.

Interoperable models

A major focus for NEPS will be to deliver efficient national capabilities from cross-domain environmental modelling using both dynamic state models and machine-learning approaches to exploit rich but often noisy data.

As already indicated, Subdomain Infrastructure Communities offer the knowledge and expertise specific to each domain and are best positioned to address noisy data and develop modelled representations of environmental states. In many cases, the outputs from these models will be new derived datasets which

can be published and accessed through the same spatiotemporal infrastructure as harmonised data products. These modelled data products, particularly when representing the predicted state of EEVs, will play an important role within NEPS and be significant reference datasets for wide reuse by Australian and international researchers and agencies.

However, the development of NEPS also offers scope for more complex initiatives to model the complexity of environmental systems. The availability of all harmonised and modelled data products through a spatiotemporal data cube will assist researchers in developing cross-domain models that may otherwise be unachievable, but more can be achieved and greater consistency and transparency can be delivered through a more consolidated approach to cross-domain modelling. By leveraging the efforts within each Subdomain Infrastructure Community to model associated environmental subsystems, NEPS has the opportunity to benefit from the modularisation of subdomain expertise in the context of federated dynamic state models.

NEPS will promote the adoption of EEVs as the parameters for communication and binding between dynamic models developed by each Subdomain Infrastructure Community. Non-EEV parameters required in such contexts are likely to be candidates for future adoption as EEVs.

Developing workflows and models that exploit subdomain models will add a further level of digital asset management to NEPS. Resulting models should be publicly accessible, well documented and sustainably managed. There is scope for automated regeneration of model outputs based on good and sufficient metadata to detect updates to input datasets and algorithms.

Provenance and traceability

It is essential that NEPS follow good contemporary practices for documenting, publishing, referencing and reusing both spatiotemporal and non-spatiotemporal assets. The same principles will apply across all these assets. ARDC (and more broadly RDA[27]) provide the leadership and models that NEPS should follow.

27 https://www.rd-alliance.org/

NEPS, along with the Subdomain Infrastructure Communities, must align where applicable with guidance offered by ANZLIC[28] in regard to spatial data standards and metadata. Domain-specific alignments (e.g. with the EOVs or EBVs) will be coordinated by the relevant Subdomain Infrastructure Community.

The FAIR Data Principles[29] directly or indirectly address the requirements for NEPS to maximise the discoverability and reuse of data assets, in particular the need for robust and stable identifier schemes, clear and comprehensive metadata, machine-readable licensing, and clear attribution and credit. NEPS will follow ARDC recommendations on preferred mechanisms to implement these principles and will maintain the NEPS FAIR Data Policy as the community contract for following these recommendations. As far as possible, data within the NEPS framework should also be licensed as freely as possible to maximise reuse. NEPS partners must evaluate the need for the system to support secure access for authorised users to sensitive or restricted data assets and to enable use of these within modelling tools. The extent and complexity of these needs may significantly influence the NEPS architecture at all levels.

The NEPS Council will be responsible for developing policies and overseeing processes to manage the lifecycle of reference models and data products delivered through the NEPS framework. Particularly in relation to the Essential Environmental Variables, users require clarity around the applicability and any limitations of such assets, around quality assurance processes, and around versioning in light of new or improved data and models. Formal mechanisms are required to support these expectations.

Clear tracing of provenance for all digital assets and consistent practice around citation and reuse will encourage and facilitate world-class cross-domain research. By serving as an integration point for all Australian environmental data, NEPS is expected to enhance the use and value of the contributing infrastructures.

28 https://www.anzlic.gov.au/

29 https://www.go-fair.org/fair-principles/

Technical Architecture

Service stack

Many of the components that are necessary to build NEPS already exist within current NCRIS capabilities or federal or state data infrastructures. The tools and systems offered by ARDC and NCI provide the technological foundations for additional NEPS developments. The infrastructure needs to be capable of scaling according to user need and to support flexible provisioning and release of resources within an on-demand computing infrastructure.

NEPS will as far as possible simply adopt and adapt the products of these other infrastructures. The scope for the NEPS technical architecture will therefore be defined by identifying the gaps between these infrastructures. Some elements identified this way may best be delivered by extending an existing infrastructure. Others may best be delivered as new independent NEPS investments. Determining which approach is best will depend on the social architecture, particularly the governance and funding models, adopted by NEPS.

NEPS will wherever applicable follow the Spatial Infrastructure Services Stack (SISS)[30] architecture (modified as illustrated in Figure 6), which is already in wide use by many Australian geospatial science and governmental initiatives. The model has here been slightly expanded to reflect the full range of digital assets, including non-spatial assets, that must be accessible to NEPS, and to accommodate the diverse international vocabularies standards in use within Subdomain Infrastructure Communities. The layering in the model facilitates separation of concerns between different stakeholders and infrastructures.

The **resource layer** primarily comprises the digital assets managed by each Subdomain Infrastructure Community, including environmental data, vocabularies and ontologies, and reusable services and

30 https://www.seegrid.csiro.au/wiki/Siss/WebHome

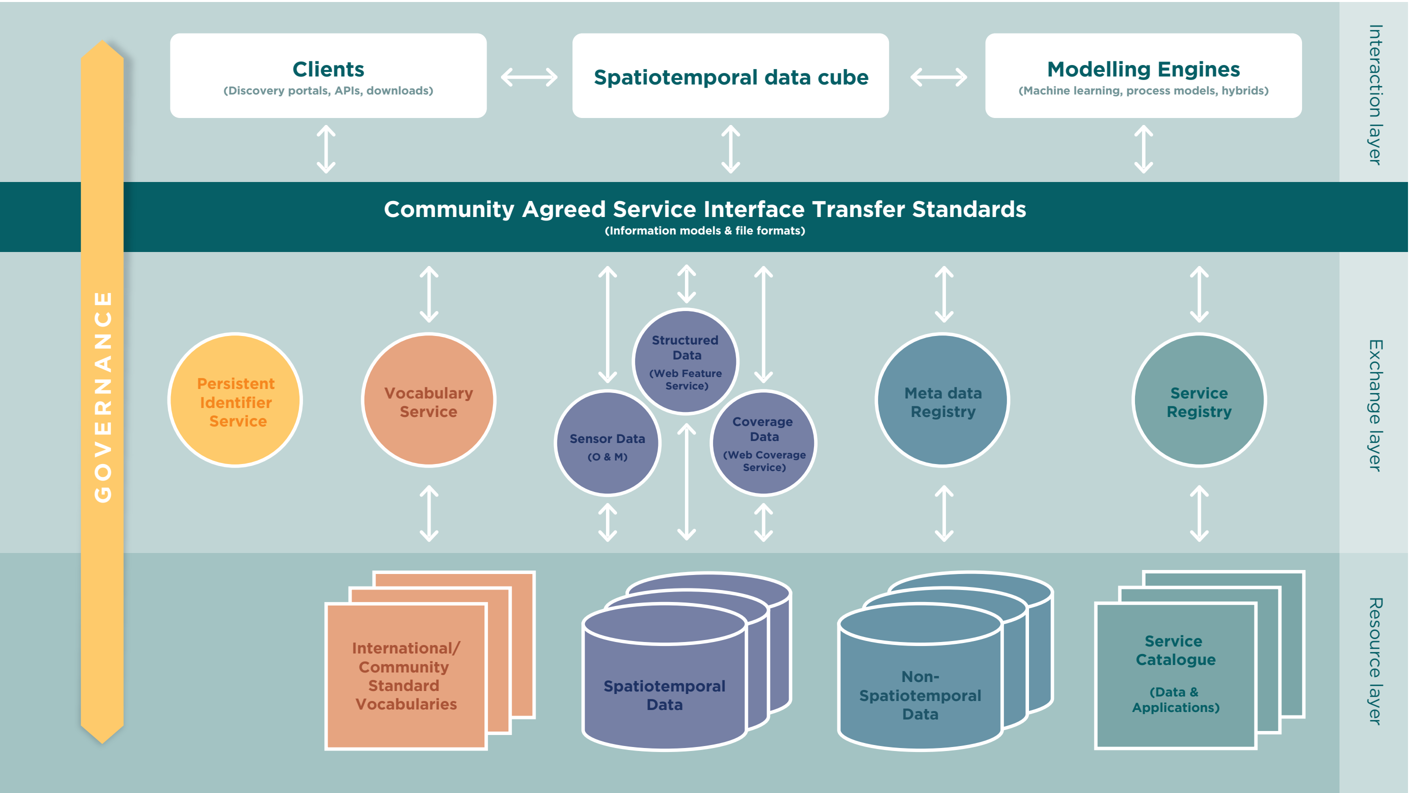


Figure 6 Infrastructure stack (adapted from AuScope Spatial Infrastructure Services Stack (SISS) documentation, https://www.seegrid.csiro.au/wiki/Siss/WebHome)

modular components. In most cases, the Subdomain Infrastructure Community itself operates as a federation of distributed assets. Provided that the community's practices satisfy the NEPS FAIR Data Policy and the NEPS Trusted Repository Policy, such assets are likely to be readily accessed within NEPS.

The assets managed in the resource layer are each exposed through the **exchange layer** via catalogues, registries, standards and services appropriate to the different data classes under consideration. Australia has significant expertise in this area. Adoption of specific standards, technologies and implementations within the exchange layer will follow a cross-domain review by the NEPS Technical Committee of practices and capabilities within the subdomains. The exchange layer brokers data from different sources and in different formats to deliver a simplified and standardised access platform for clients.

The NEPS **interaction layer** includes three inter-related categories of application that exploit this platform.

First, there is a general requirement to support diverse clients, including portals and services operated by Subdomain Infrastructure Communities. These clients offer diverse paths to support the needs of different user communities for accessing and visualising environmental data. NEPS will simplify and expand the capabilities of these clients by fostering cross-domain standardisation and consistent access to resources from all communities. NEPS will engage with NEPS User Communities to assess requirements for cross-domain portals, both for research users and for regulatory and industry applications. Regulatory and industry uses are especially likely to require a range of highly standardised modelling flows and clear reporting of provenance and confidence.

Secondly, NEPS requires the implementation of a spatiotemporal data cube that supports efficient alignment between data organised at differing temporal and spatial scales and can transparently apply a range of interpolation and extrapolation strategies. This capability will require processing capabilities outside the scope of the exchange layer and can be regarded as a specialised client with wide application for users, including many of the clients already specified.

Thirdly, the interaction layer will include modelling engines to support analysis and prediction based on data and software elements accessed directly through the exchange layer or via the spatiotemporal data cube. This capability may also be used directly by the other applications in the interaction layer.

The structure presented for the interaction layer is highly simplified. The interdependencies between different applications indicate the possibility for a more modular component architecture, but this should be developed following early review by the NEPS Technical Committee of requirements and priorities within the NEPS Federation Community.

Deployment of services

NEPS is intended to be a pragmatic and lightweight investment that leverages and reinforces existing research infrastructure capabilities within Australia.

Wherever possible, NEPS will rely on existing services offered by Subdomain Infrastructure Partners, working through the NEPS Federation Community to coordinate any necessary adaptation of these services.

Where new services are required, the NEPS Council must make a judgement regarding the appropriate host to deploy the service. The service should be hosted by a Subdomain Infrastructure Community whenever it is well aligned with the mission and the capabilities of the community and can be accommodated sustainably within its work programme and infrastructure. In other situations, and for components that belong strictly to NEPS rather than a Subdomain Infrastructure Community, the NEPS Council should develop policies to select an appropriate host for the service (in compliance with the NEPS FAIR Data Policy and NEPS Trusted Repository Policy).

Annex A NEPS Scoping Study Stage 1 Consultation

Stage 1 of the consultation process, undertaken in Q1 of 2019, involved the Expert Panel consulting with an initial representative group of stakeholders to refine the Panel's consultation approach.

Both potential end-users and potential providers of environmental and supporting information were consulted. Table 1 shows the individuals and groups that were consulted during Stage 1.

Stage 1 Consultation Organisation/Individual	Contact Point	NEPS Expert Panel Interviewer/s
Threatened Species Research Hub (NESP)	Brendan Wintle	Steve Morton
WA Environmental Protection Authority	Tom Hatton	Rob Vertessy & Warwick McDonald
Northern Australian Research Hub	Michael Douglas	Steve Morton
Centre for Air Pollution, Energy and Health Research	Guy Marks	Andrea Hinwood
Commissioner for Environmental Sustainability Victoria	Gillian Sparkes	Andrea Hinwood
Environmental Health Grouping of national health departments & agencies	Sophie Dwyer	Andrea Hinwood
NRM Regions Australia	Kate Andrews	Adam Lewis & Warwick McDonald
National Earth and Environmental Science Facilities Forum	Tim Rawling	Adam Lewis & Warwick McDonald

Table 1
NEPS Stage 1 Consultation list

Stage 1 consultation questions for potential users of a NEPS

The focus of our engagement with potential users will be to ascertain where the demand for and expectation of a NEPS might lie. In governments, industries and communities, decisions are made regularly regarding the way we use, protect and remediate our natural and managed environments. We seek to identify where better environmental, social and economic outcomes can result from the stronger use of information about past, present and future state(s) in decision making. Questions we will test with this group in the Stage 1 consultation include:

Purposes

1. How often do you make decisions that affect environmental outcomes, or that are dependent on the state of the environment, and what are the consequences of those decisions?
2. What are the most important environmental management or policy questions you are challenged by that urgently demand better information?
3. What, if any, expectations would you have of a NEPS? What would your priorities be for developing a NEPS? Is there 'low hanging fruit' for a NEPS?

Information Priorities

4. In making those decisions, what environmental variables are of most interest to you and how satisfied are you with the present availability of information on them?
5. Thinking about the information you need most or would be most beneficial to your decision making, what proportion is historic data, very recent data, short-term forecasts or long-term predictions?
6. At what level of spatial and temporal detail (granularity) do you require that information?

Use and Adoption

7. What are the factors that lead you to trust and use that information?
8. Thinking about a trusted environmental information source that you currently use in decision making, tell us about its positive and negative attributes.
9. What could be done to improve the connection between end-users and providers of environmental information?

Consultation

10. Thinking about the NEPS consultation strategy, what engagement modalities do you favour?
11. Who are the other important potential end-users of environmental information that we should consult?

Stage 1 consultation questions for potential creators of a NEPS

The key goal of our engagement with potential creators will be to ascertain where the state-of-the-art environmental science, monitoring, prediction and informatics lies, with a view to identifying the challenges and opportunities of developing a system, the potential roles of the providers, and areas where more effort will be required. Our level of scientific understanding in environmental processes is strong in some areas and weak in others. Also, the availability of environmental data of different kinds varies significantly.

Recent technological developments are rapidly changing our ability to capture data and improve understanding through modelling to support better decision making. Questions we will test with this group in the Stage 1 consultation include:

Priority Applications

1. Reflecting on your experiences with end-users of environmental information, where do you think they need more support? Do you see areas of particular expectation from end-users for a prediction system?
2. In what areas do you feel that we have useful environmental information that is yet to be harnessed in decision making?
3. In what areas do you see a need to build information?

Information Base and Infrastructure

4. Thinking about trusted existing environmental information sources being used in decision making, what do you see as their strengths and weaknesses?
5. What new data sources, analytic methods or tools do you see arriving in the next five years that are potential game changers for environmental management?
6. What environmental information infrastructure do you have (or plan to have) that would be necessary to implement the NEPS?

Use and Adoption

7. What are the issues involved in getting end-users to trust and use that information?
8. What more needs to be done to evolve that information into a state that satisfies end-user requirements?

Consultation

9. Thinking about the NEPS consultation strategy, what engagement modalities do you favour?
10. Who are the other important potential providers of environmental information that we should consult?

Key findings from Stage 1 consultations

1. Users of environmental intelligence were very keen to have access to better information to support their decision making but found it challenging to define their requirements. We concluded that deep dialogue is necessary to elicit user requirements.
2. Researchers who create tools and data sets for environmental analysis are eager to increase their impact by getting closer to users of environmental intelligence but struggle to identify points of entry. The large number of users and their segmentation across jurisdictions makes it hard for the research community to engage in a meaningful way. They see contributing to the building and maintenance of national research infrastructure as an ideal way to engage with user communities.
3. The users and creators of environmental intelligence who we consulted highlighted three promising user communities that we have deemed worthy of closer examination.
    a. Environmental Regulators regularly make highly consequential decisions that affect the state of our natural environment. Their decisions are subject to high levels of scrutiny, often expected within tight timelines and need to give regard to cumulative effects. Comprehensive, timely and reliable environmental intelligence is thus a highly sought-after commodity for environmental regulators.
    b. State of Environment (SoE) reports play a vital role in informing the Australian public about the current state and trend in environmental conditions across our country. SoE reporting is undertaken in all Australian jurisdictions but on different schedules and based on different methods. Several SoE reporting teams have highlighted the difficulties they face in getting access to quality data sets and methods and are seeking help from the science community.
    c. Australia's 56 regional Natural Resource Management bodies have a huge task; to deliver positive environmental outcomes relating to water, soils, vegetation, biodiversity, climate change and other important facets of environmental management and so their information needs are extremely broad. Whilst their environmental management remit is wide, their capability is modest and they rely on a range of institutions to provide the vital environmental intelligence they need to perform their mission.

Annex B NEPS Scoping Study Stage 2 Consultation

Stage 2 of the consultation process commences in towards the end of Q3 of 2019 and will be conducted over the rest of the year. The Expert Panel's Discussion paper, which is available on the NEPS web site (www.science.uq.edu.au/neps), will be widely promoted through a wide range of mainstream and social media. The NEPS web page will invite participation in the consultation process via a number of mechanisms (see *Consultation Methods and Tools*). The Expert Panel will consult with a wide range of stakeholders with a potential interest in the NEPS. Both potential end-users and potential providers will be consulted.

Stakeholder Identification/Categories

Consultations will be held with key Australian informants, being either individuals or small groups representing particular organisations, who have an interest in the development of the NEPS, particularly including representatives from:

- relevant Commonwealth, State and Territory government departments;
- current NRI staff; and
- researchers at universities.

Key domain conferences will also be leveraged to gain an understanding of the views of the broader ecological science community.

At a minimum the study should engage with:

- Atlas of Living Australia;
- AuScope;
- Australian Research Council (ARC);
- ARC Centres of Excellence;
- Australian Research Data Commons (ARDC) – including the National Research Data Cloud;
- Australian Urban Research Infrastructure Network (AURIN);
- Australia's Tier 1 High Performance Computing (HPC) facilities – including National Computational Infrastructure (NCI) and Pawsey Supercomputing Centre;
- Bureau of Meteorology (BoM);
- Commonwealth Scientific and Industrial Research Organisation (CSIRO);
- Department of the Environment and Energy, including groups associated with its environmental economic accounting agenda;
- Department of Agriculture and Water Resources;
- Ecosystem Science Council;
- Integrated Marine Observing System (IMOS);
- National Environmental Science Program; and
- Terrestrial Ecosystem Research Network (TERN).

Discussions, but not overseas travel, will occur with relevant international stakeholders, such as overseas environmental and climate monitoring facilities and any institutions offering environmental prediction services in other countries.

Consultation Methods and Tools

The NEPS Expert Panel will use a range of consultation methods and tools as appropriate, including written submissions, surveys, interviews, roundtables and workshops.

Method of consultation	Possible Advantages	Possible Disadvantages
Written submissions	Opportunity for in-depth information exchange, including detailed use cases Input for those who may not attend workshops/roundtables	Reviewing submissions can be labour intensive, time consuming Hard to target specific stakeholders
Telephone interviews	Input from those who may not attend a public meeting Higher response rate than mailed surveys	More expensive and labour intensive than mailed surveys.
Face-to-face interviews	Opportunity for in-depth information exchange in non-threatening forum Opportunity to gain feedback from all stakeholders Can be used to evaluate potential citizen committee members	Time consuming Expensive
E-surveys	Input from those who may not attend a workshop or roundtable Provides a mechanism for extending a mailing list Provides information from a cross section of the community not only activists Statistics are valuable in determining a decision	Response rate can be low To get statistically valid results, can be labour intensive and expensive Level of detail may be limited
Meetings with existing groups	In-depth information exchange in a non-threatening forum	May be too selective and can leave out important groups
Workshops/ Roundtables	Promotes community involvement Excellent for discussions Maximizes feedback obtained from participants Capitalises on existing networks Enhances credibility Fosters public ownership in solving the problem	Can be difficult to control information flow Can build false expectations Can be expensive and time consuming (professional facilitators)

*adapted from IAP2 Toolbox

Consultation communication objectives

- To ensure that the ecosystem science community and broader NEPS stakeholders are engaged throughout the various consultation and planning activities of the NEPS Scoping Study.
- To engage the relevant sections of the Australian ecosystem science and management communities for producing the NEPS Investment Plan.
- To produce the highest quality consultation summaries and resultant NEPS Investment plan to address why a NEPS is necessary, what it is to produce, where to focus effort, and who should be involved in building and maintenance.
- To positively position the NEPS Investment Plan as an exciting and unique opportunity for the ecosystem science community to set the foundations for building and delivering a sustainable and effective environmental prediction system for Australia.

Consultation Questions

The NEPS Expert Panel will use a pre-defined set of questions in all consultations with stakeholders. These questions have been refined based on feedback from the Expert Panel following Stage 1 consultations. In the case of written submissions, Instructions on the NEPS web site will encourage those making a submission to address the Discussion paper and System Design Document.

Example text and consultation questions to capture use cases

As part of the NEPS Scoping Study, we wish to understand current and expected needs for access to integrated information on aspects of the Australian environment, past, present and predicted.

In this context, "environment" is taken to encompass all biotic and abiotic factors that may vary over time and space, both in natural and artificial systems (cities, agriculture, etc.) and across terrestrial, freshwater and marine areas, with a focus on how these factors affect the ability of people, crops, livestock and biodiversity to survive and exploit these systems. Relevant factors include (but are not limited to) climate, geology, soils, biodiversity, land cover, land use, hydrology, topography, infrastructure, air quality and pollution, at all scales from global through continental to local.

In order to clarify requirements and to understand the need across Australia for a National Environmental Prediction System, we ask you to describe your information needs in support of your research or applied decisions.

These questions are to help with development of broad requirements and needs for environmental prediction. We do not expect all needs to be addressed within an initial proposal for a NEPS, but this information will help to clarify the ideal ultimate scope for such a system.

Please use the following questions to describe a typical inquiry (research question or decision) that you might carry out based on one or more environmental factors. If your research or work includes multiple classes of inquiry with different characteristics and needs, we welcome separate descriptions of each. We also welcome descriptions of situations where your information needs relating to the Australian environment are difficult or impossible to address with existing resources.

1. Briefly describe the context and purpose for the inquiry
2. Define the core question to be addressed in the inquiry
3. Explain the spatial (continent, state/territory, local, site, 20km^2, 1m^2, etc.) and temporal (decade, year, week, day, hour, etc.) scales at which this question needs to be addressed
4. Explain the importance or significance of being able to address the question, including financial aspects, future expenditure of effort, regulatory or conservation outcomes, role within a larger research agenda, etc.
5. Identify any time-critical aspects around addressing the question (e.g. within 5 working days)
6. Are you able properly to address this question today?
   a. If so, please describe how you source and use information as inputs to address the question
   b. Otherwise, please identify gaps or constraints around existing information and tools that limit capacity in this regard, and describe any processes you currently follow to approximate an answer
7. Please list categories and sources (where applicable) of environmental information that you use, or need, to address the question

Here include a table to document information types – needs some more refinement, but should probably include at least the following elements:

- Picklist of broad categories (climate, biodiversity, land use/cover, marine, freshwater, air, infrastructure)
- Specific data inputs (mean annual rainfall, vegetation class, human population per km^2, species distributions, CO_2 concentration ppm, etc.)
- Spatial precision required
- Temporal precision required
- Availability (Existing satisfactory, Existing deficient, Lacking, …)
- Notes on specific requirements for use of these data

Consultation Timeline

Activity	Time (start date)
Stage 1 Consultation	February to March 2019
Planning began for Stage 2 Consultation	February 2019
NEPS Scoping Study Expert Panel meeting to discuss Stage 1 Consultation outcomes and Stage 2 Consultation strategy	March 2019
Stage 2 Consultation (see table below for more detail)	August 2019 to March 2020
Expert Panel to collate and analyse consultation responses and develop the first draft of the Investment Plan.	January to March 2020
Further consultation (at this time the Expert Panel may want to seek advice/guidance from DoE regarding the content, detail and proposed direction).	April 2020
Redraft of Investment Plan based on further consultation	May 2020
Approval of the Investment Plan by DoE	May 2020

Method of consultation	Consultation Period
Written submissions	Q2 to Q4 2019
Telephone interviews	Q3 & Q4 2019
Face-to-face interviews	Q3 & Q4 2019
E-surveys	Q3 & Q4 2019
Meetings with existing groups	Q3 & Q4 2019 – additional reviews in Q1 2020
Workshops/Roundtables	Q1 2020

All forms of consultation will be completed by 1 May 2020 to allow sufficient time for the Expert Panel to collate and analyse responses and develop the draft Investment Plan. The draft Investment Plan is to be submitted to DoE by 30 May 2020.

Annex C NEPS Federation Community – roles and accountabilities

	NEPS Council	NEPS Technical Committee
Mandate	Authority initially established through funding agreement to establish NEPS	Authorised and provided with delegated decision rights by the NEPS Council
Main functions	1. Establishment, and decisions about the overall direction and operation of NEPS 2. Establishment and oversight of technical level authority and advisory structures, including the NEPS Technical Committee 3. Oversight of the operations of NEPS 4. Representing NEPS and communicating with those outside NEPS	1. Making decisions about technical decision domains, including technical prioritisation 2. Providing recommendations to the NEPS Council in relation to agreed decision domains 3. Oversight of the alignment processes between Subdomain Infrastructure Communities
Decision domains	Social architecture – agreements and policies regarding: 1. The purpose and objectives of NEPS Federation Community 2. The strategic direction and (business) operation of NEPS Federation Community 3. Collective business processes that NEPS Federation Community will use 4. Deployment and maintenance procedures for NEPS Federation Community 5. Establishment of roles within NEPS Federation Community 6. Engagement with and recognition of new Subdomain Infrastructure Communities 7. Recognition of NEPS User Communities (with eligibility to designate representatives for NEPS Technical Committee) 8. Engage with and approve recognition of external domain authorities 9. Governance and quality assurance for models and data products 10. Exit strategy articulating approaches to exiting and terminating NEPS	Information architecture – agreements regarding: 1. Determining what constitutes acceptable data for sharing through NEPS 2. Authorising the publication of data through NEPS 3. Determining acceptable levels of FAIRness for publication 4. Determining acceptable levels of NEPS Trusted Repository requirements 5. Identifying common information models (data structures) that will be supported by the community 6. Determining agreed semantics – provider specific and agreed community vocabularies and ontologies 7. Determining agreed Essential Environmental Variables within NEPS Technical architecture – agreements regarding: 1. The design and deployment of common infrastructure elements such as resource discovery mechanisms 2. The computational interfaces to be supported by each Subdomain Infrastructure Community 3. The supported end-user experience (portal) 4. Where components (services) are deployed and using what technology

Decision rights	Decide on all decision domains specified above, as well as endorse technical architecture recommendations from Technical Committee	Initial settings: Decide on all decision domains relating to information architecture and make recommendations on all decision domains relating to technical architecture (for endorsement by NEPS Council)
Representation	One representative designated by each contributing Subdomain Infrastructure Community Observers may be included (subject to criteria and terms to be specified in the council ToRs)	One representative designated by each contributing Subdomain Infrastructure Community One representative designated by each NEPS User Community (user communities as designated by NEPS Council) Other representatives may be invited to contribute (subject to criteria and terms to be specified in the council ToRs)
Decision-making processes	• One seat for each Subdomain Infrastructure Community • Other seats according to ToRs at establishment • Consensus-based decision making, with vote or chair decides if no consensus reached	• One seat for each Subdomain Infrastructure Community • One seat for each recognised NEPS User Community • Other seats according to ToRs at establishment • Decision-making according to process to be agreed by NEPS Council

Annex D NEPS FAIR Data Policy – initial settings

Adapted from Box et al. 2019, Guidelines for the development of a Data Stewardship and Governance Framework for the Agricultural Research Federation (AgReFed).

NEPS will require a data policy to establish clarity for Subdomain Infrastructure Communities around application of the FAIR Data Principles. The following table outlines a series of increasingly rigorous alignments with each of the FAIR principles. The cells marked yellow are suggested as acceptable levels for FAIR compliance within NEPS, with the brightest shade of yellow as the target level for each principle.

	Principle	> > > Increasingly FAIR > > >				
	FINDABLE					
Q1	The data product has been assigned (an) identifier(s)	No identifier	Local identifier	Web address (URL)	Globally unique, citable and persistent identifier (e.g. DOI, PURL, or Handle)	
Q2	The data product identifier is included in all metadata records/ files describing the data	No	Yes			
Q3	The data product is described by a metadata record	The data is not described	Brief title and description	Brief title and description, and multiple other fields filled out, albeit briefly.	Comprehensively (including all AgReFed required fields*) using a formal machine-readable metadata schema.	
Q4	The data product is described by a metadata record that is indexed in a searchable registry or repository	The data is not described in any registry or repository	Local institutional repository	Generalist public repository	Generalist public repository	Data is in one place but discoverable through several places (i.e. other registries, RDA, Google Data Search)
	ACCESSIBLE					
Q5	How accessible is the data? The access method(s) must be explicitly stated in the metadata record, e.g. if any authentication is needed, or there are any restrictions to access.	No metadata record	Access to metadata only	Unspecified access conditions e.g. "contact the data custodian to discuss access"	Embargoed access after a specified date; or A deidentified version of the data is publicly accessible	Fully accessible public, or to persons who meet and follow explicitly stated conditions and processes, e.g. ethics approval for sensitive data
Q6	Data are available for reuse via a standardised communication protocol, such as file download over https, or a web service.	No access to data	By individual arrangement	File download from online location	Non-standard web service (e.g. OpenAPI/Swagger/ informal API)	Standard web service API (e.g. OGC)
Q7	The repository/registry agrees to maintain the persistence of the metadata record, even if the data product is no longer available.	No (or not applicable, if no metadata record exists)	Unsure	Yes		

	INTEROPERABLE					
Q8	The data products are available in (an) open (file) format(s)	Data are mostly available only in a proprietary format	Data are available in an open format	Data are available in an open, documented, widely-used standard format (i.e. NetCDF, CSV, JSON, XML, etc)		
Q9	The data is machine readable (see Glossary for definition)	The data are unstructured	The data are structured and machine-readable (i.e. csv, JSON, XML, RDF, database files, etc)			
Q10	The data are semantically interoperable, because they use standard, accessible ontologies and/or vocabularies to describe the data elements/variables.	Data elements are not described (i.e. fields or objects are labelled with codes or not at all)	Data elements are described (so that a human user can correctly interpret the data), but no standards have been used in the description	Recognised standards have been used in the description of data elements, but no published vocabularies with resolvable URIs	Published vocabularies using resolvable global identifiers linking to explanations are used, so that the data can be read and understood by machines as well as humans.	
Q11	The relationships to other data and resources (e.g. related data-sets, services, publications, grants, etc) are described in the meta-data or data, to provide context around the data.	There are no links to other metadata or data	The metadata record includes URI links to related metadata, data and defini-tions	Qualified links to other resources are recorded in a machine readable format, e.g. a linked data format such as RDF		
	REUSABLE					
Q12	Machine-readable data licens-es are assigned to each data product, and are stated in the metadata record.	No license is applied	Non-standard license applied, without a license deed URL encoded in a ma-chine-readable format (e.g. RDF/XML) in the metadata record	Standard license applied (e.g. Creative Commons), without a license deed URL encoded in a machine-read-able format (e.g. RDF/XML) in the metadata record	Non-standard license applied, WITH the license deed URL encoded in a ma-chine-readable format (e.g. RDF/XML) in the metadata record	Standard license applied (e.g. Creative Commons), WITH the license deed URL encoded in a machine-read-able format (e.g. RDF/XML) in the metadata record
Q13	The provenance of the data prod-uct is described in the metadata, i.e. project objectives, data gen-eration/collection (including from external sources) and processing workflows.	No provenance information is recorded	Partially recorded	Comprehensively recorded in a text format (i.e. TXT or PDF)	Comprehensively record-ed in a machine readable format (i.e. in metadata record's schema or PROV, or in RDF, JSON, NetCDF, XML, etc)	
Q14	The preferred citation for the data product is provided in metadata record	No	Citation does not include identifier	Citation includes identifier		

Annex E NEPS Trusted Repository Policy – initial settings

Adapted from Box et al. 2019, Guidelines for the development of a Data Stewardship and Governance Framework for the Agricultural Research Federation (AgReFed).

In a distributed information system such as NEPS, the repositories managed by data providers, in particular those of the Subdomain Infrastructure Communities, need to be reliable and trustworthy. The CoreTrustSeal certification provides a process whereby custodians can measure repository compliance levels against sixteen identified characteristics of trustworthy repositories, the Core Trustworthy Data Repositories Requirements (hereafter referred to as the CoreTrustSeal requirements):

- R1. Mission/Scope
- **R2. Licenses**
- **R3. Continuity of Access**
- **R4. Confidentiality/Ethics**
- R5. Organisational Infrastructure
- R6. Expert Guidance
- R7. Digital Object Management
- R8. Appraisal
- R9. Documented Storage Procedures
- R10. Preservation Plan
- **R11. Data Quality**
- R12. Workflows
- **R13. Data Discovery and Identification**
- **R14. Data Reuse**
- **R15. Technical infrastructure**
- **R16. Security**

Dara providers wishing to share data via the NEPS will conduct an assessment of the repository through which the data will be made persistently available (NEPS Trusted Repository Self-assessment).

The NEPS uses a simplified process based on the CoreTrustSeal requirements. Instead of five levels of compliance for each of the requirements, NEPS assesses whether each requirement has been implemented or not.

The Subdomain Infrastructure Communities will undertake NEPS Trusted Repository Self-assessments of their repositories. The Federation Data Steward (or other agreed community role) will review and validate the assessment against the agreed NEPS Trusted Repository Policy.

The NEPS Trusted Repository Policy set by the NEPS (through the Federation Technical Committee) will be used to determine the acceptable level of trustworthiness that will be need to be met for sharing of data through the NEPS.

Initially it is proposed that for NEPS the repository meet requirements relating to Licences (R2), Continuity of Access (R3), Confidentiality/Ethics (R4), Data Quality (R11), Data Discovery and Identification (R13), Data Reuse (R14), Technical Infrastructure (R15), and Security (R16). These requirements may change over time.

NEPS SYSTEM DESIGN

National Environmental
Prediction System Scoping Study

15 November 2019