A National Environmental Information Infrastructure

A. Zerger, Horswell, R., Woolf, A., Percival, D. and Millard, J.

Bureau of Meteorology, Canberra ACT 2600
Email: a.zerger@bom.gov.au

Abstract: Australia invests significant resources in environmental data acquisition, management and publication. Although data are abundant, users are typically hampered by an inability to discover, access and re-use the data. The National Environmental Information Infrastructure (NEII) activity will improve the effectiveness and efficiency of discovering, accessing and re-using environmental data beyond its primary purpose. The NEII is envisioned as a federation of environmental data nodes adopting common standards to create a data infrastructure with an initial focus on nationally significant environmental data. Its primary focus is on the discovery and re-use of national environmental data that is already well-managed, but that currently has limited application beyond its original business purpose. In common with spatial data infrastructures, the NEII encompasses common data models, exchange formats and standard network protocols along with centralised catalogues of uniform metadata descriptions. It also includes standardised models for describing environmental measurements, monitoring sites and methods used to observe the environment to address the unique requirements of environmental information. These architectural elements are more fully described in the NEII Reference Architecture (Bureau of Meteorology, 2014b, www.neii.gov.au).

The Bureau of Meteorology is leading the development of the NEII including provision of core coordination and integration infrastructure, as well as the governance and collaboration framework for its development and operation. However its ongoing success is dependent on establishing enduring partnerships with major national environmental information organisations. The NEII Programme is framed around five focus areas including (a) Communication, (b) Engagement, (c) Policies and frameworks, (d) Data management, and (e) ICT build. This paper provides a more detailed overview of one component of the Data Management area regarding the development of the NEII Conformance Framework.

Conformance means following established guidelines, specifications and standards or working towards them. A conformance framework is a structured set of guidelines that detail the levels at which a participant provides services that comply with the NEII architecture. The NEII Conformance Framework has been developed to (a) enable partners to set priorities against specific expectations that will jointly lead to developing a sustainable federated environmental information system; (b) enable users of the NEII to rapidly assess the suitability of data for their business needs; and (c) provide a common measurement system to monitor progress in the development of the NEII. The framework adopts a capability maturity model approach to describe the steady improvement required for NEII nodes to progress from a basic data release (experimental data service with limited operational support) to an enduring federated system providing environmental information. The design of the model is informed by other maturity-based approaches for open data such as the Open Data Institute's certification approach (Open Data Institute 2015). The paper also presents a worked example of the conformance framework against the Bureau of Meteorology's Australian Hydrologic Geospatial Fabric NEII data services.

The paper concludes with an overview of major learnings from the programme to-date. These relate primarily to reducing the cost of participation in NEII by data custodians given the challenge of delivering data using NEII standards; and developing approaches to better support users to use the portfolio of NEII data given some of the methods of data delivery may be new to non-technical users. The NEII programme has prioritised both these and developed work packages to achieve improvements.

Keywords: *NEII, federated information system, OGC, Open Data Institute, conformance, capability maturity model*

1. INTRODUCTION

Governments, industry and the community need access to comprehensive, trusted and timely environmental information to inform assessments and decisions about how development of Australia's natural resources might proceed and what impact development and use might have on our environmental assets. The National Plan for Environmental Information (NPEI) initiative was established in 2010 to improve the quality and accessibility of Australian environmental information. A core activity under the initiative is the development of a National Environmental Information Infrastructure (NEII) to improve discovery, access and re-use of environmental data. This paper provides a more detailed overview of the NEII introducing the broader policy context, purpose and an overview of how the programme is structured. It then pays particular attention to the NEII Conformance Framework that has been developed to (a) enable partners to set achievable priorities for how to participate in the NEII, (b) where an assessment against the conformation framework has been completed to enable users to assess data suitability for their business needs, and (c) to provide the NEII programme a common measurement system to monitor progress regarding the development of the NEII.

2. WHY IS AN NEII NECESSARY?

Australia invests significant resources towards environmental data acquisition, management and publication through the activities of many organisations. At the commonwealth level this includes the breadth of weather, climate and water data managed by the Bureau of Meteorology; earth science and national mapping data from Geoscience Australia, CSIROs research data; and other data from agencies including the Department of Agriculture and the Department of the Environment. State and Territory agencies are also major producers and custodians of environmental data including, for example, soil mapping, air quality and ecological observations, and land use mapping. These data are often typically aggregated at a national scale to produce compendium data products.

Investment in environmental data acquisition also occurs through other national programs, for example: research and research infrastructure investment supported by the National Collaborative Research Infrastructure Strategy (NCRIS), the National Environmental Science Program (NESP); and the Cooperative Research Centres (CRC) program; by catchment management authorities through developed natural resource management grants; and the activities of the research sector including in particular CSIRO and universities.

Although environmental data are abundant, potential users are typically hampered by an inability to discover, access and re-use the data. It often exists only within individual agencies to support internal business requirements or within individual environmental domains. Consequently, enabling discovery, access, and re-use across domains (e.g., air, land, oceans, and water) remains challenging. Where data can be found, gaining access can introduce new challenges because not all data custodians are equipped for data provision and outreach services; data may utilise proprietary file formats; or large data volumes make delivering data difficult. Finally, the absence of standards introduces a challenge when users attempt to re-use data and lack the domain-specific understanding to make informed decisions around its quality and fitness-for-purpose.

3. BROADER POLICY CONTEXT

Development of the NEII recognises the importance of improving transparency and access to government data and information. Its objectives thus align with those of the international Open Government Partnership and Australia's activities in this space including the recent Declaration of Open Government; data.gov.au; Gov 2.0; The Australian Government's Big Data Strategy; and principles championed through the Office of the Australian Information Commissioner's Principles on Open Public Sector Information. The national and international focus on data and information transparency will play a defining role in the organisational, political and technical implementation aspects of the NEII.

4. WHAT IS A NATIONAL ENVIRONMENTAL INFORMATION INFRASTRUCTURE

The NEII is a programme to improve the effectiveness and efficiency of discovering, accessing and using environmental data. Its primary focus is on national environmental data that is already well-managed, but that currently has limited application beyond its original business purpose. It is envisioned as a federation of environmental data nodes adopting standards for interoperability and thus data will continue to be retained at-source with the organisations that capture, curate and support the data with appropriate subject-matter expertise. In common with spatial data infrastructures (SDI), the NEII encompasses common data models, exchange formats and standard network protocols along with centralised catalogues of uniform metadata descriptions. It also includes standardised models for describing environmental measurements, and monitoring sites and methods used to observe the environment. These later components differentiate the NEII

from related infrastructure activities such as spatial data infrastructures. The key architectural elements are more fully described in the NEII Reference Architecture (Bureau of Meteorology, 2014).

5. NEII PROGRAMME – OVERVIEW

The Bureau of Meteorology is leading the development of the NEII including providing the core coordination and integration infrastructure as well as a governance and collaboration framework for its development and operation. However its ongoing benefit will only be realised through partner participation. The NEII Programme is framed around five focus areas including (a) Communication, (b) Engagement, (c) Policies and frameworks, (d) Data management, and (e) ICT build. The ICT build focus area seeks to realise the architectural vision presented in the Reference Architecture (Bureau of Meteorology 2014) while other focus areas are primarily in place to support this objective. Further information regarding the NEII focus areas and related deliverables can be found at the NEII web and collaboration space (neii.gov.au). The following sections provide a more detailed overview of one component of the Data Management focus area, the NEII Conformance Framework.

5.1. NEII conformance framework

Conformance means following established guidelines, specifications, and standards or working towards them. A conformance framework is a structured set of guidelines that detail the levels at which a participant provides services that comply with the NEII architecture. A conformance framework is necessary to:

- enable partners to set priorities against specific expectations that will jointly lead to developing a sustainable federated environmental information system;

- enable users of the NEII to rapidly assess the suitability of data for their business needs; and

- provide a common measurement system to monitor progress in the development of the NEII

The framework expands on the NEII Reference Architecture (Bureau of Meteorology 2014), which uses the Reference Model for Open Distributed Processing (RM-ODP) to guide how information and computational components are integrated to produce robust and sustainable NEII nodes. RM-ODP factors a federated system description into five complementary viewpoints:

- **Enterprise** defines the purpose, scope and policies of the system, and can describe the business processes and how to meet them.

- **Information** describes the semantics of information and information processing within the system, and can describe the structure and content type of data.

- **Computational** decomposition of data and information into operation, behavior, flow and interaction.

- **Engineering** describes the system infrastructure and mechanisms supporting data federation.

- **Technology** focuses on the technology choices selected to realise the system.

The framework primarily addresses the computational and engineering viewpoints of conformance with the enterprise providing context to data users. The framework uses a capability maturity model approach to describe the steady improvement required for NEII nodes to progress from a basic data release (experimental data service with limited operational support) to an enduring federated system providing robust environmental data. The design of the model is informed by other maturity-based approaches for open data such as the Open Data Institute's certification approach (Open Data Institute 2015). It is organised around NEII's discovery, access and re-use concept and is described in further detail below and summarised in tabular format in Table 1.

5.1.1. Discovery

The discovery element of the maturity model focuses on delivering metadata through metadata standards to describe information resources in the NEII (for example, geographies, gridded data, and observations). Metadata describe an information resource according to standard characteristics, such as its title, abstract, lineage, access and data availability and the custodian information. Table 1 shows the progression of metadata from non-standardised online resources such as PDF documents (entry level) to a network of catalogues adopting user-agreed standards (advanced level).

5.1.2. Access

The access element of the framework focuses on delivering environmental data through machine-readable web data services. Beyond the first step of data discovery, most users ultimately need access to environmental data. Web data services enable software applications and users to access data in federated environments. Through the adoption of standards (for example, Open Geospatial Consortium), an NEII node provides web data service access to key environmental data, including:

- geographic features, such as hydrological catchment boundaries, roads, administrative boundaries;

- environmental observations, such as time-series data of rainfall observations at a site; and

- gridded data, such as satellite imagery, and digital elevation models.

5.1.3. Re-use

The re-use element of the maturity model is essential to providing value beyond the intended purpose of the data. It includes:

- licensing;

- operational robustness and support; and

- information elements, or the amount of formalism in the data.

From a licensing perspective, open access licensing frameworks are central to maximising re-use and minimising the transaction cost to both users and providers. The conformance framework aligns with the principles championed by the Australian Government Open Access and Licencing Framework (AusGOAL) and the Australian Government Principles on Open Public Sector Information (Office of the Australian Government Environmental Information Commissioner 2015). This relates specifically to the recommended use of Creative Commons licencing as a default position. The operational element is a defining element of the NEII, and emphasises data delivery through operational infrastructure in partner organisations. It addresses both technical and support aspects, such as: the extent of testing and quality assurance in place to support a web data service; whether a support team is available to users; and the extent of user documentation supporting adoption.

The information element also aligns with the information modelling component detailed in the NEII Reference Architecture (Bureau of Meteorology 2014, page 17), which provides users with information about the logical structure and semantic content of data to make them more useful beyond their intended purpose. In essence the information element explains how much explicit formalism is provided around the structure of data, and is addressed in further detail in the NEII Information Modelling Discussion Paper (Bureau of Meteorology 2015a).

Table 1. Conformance maturity model

	Discovery	Access			Re-use		
		Metadata	**Geographies**	**Gridded Data**	**Licensing**	**Operational**	**Information**
Entry	Data details are available online	Basic interface deployment	Basic interface deployment (for example, WMS, WFS)	Basic graphic rendering (for example, WMS)	Licence conditions explicit to users but not through an open licence	Data services experimental; no data support function in place	Information is unstructured
Base	Metadata are structured according to a standard (for example, ISO19115, Dublin Core) and are visible through a catalogue	Mandatory interface deployment	Partial mandatory interface deployment	TBC		Data services operational; data and data service documentation exists describing method of data creation	Information is structured
Foundation	Metadata are visible through domain-specific catalogue	Basic deployment for discovery and publication	Complete mandatory interface deployment	TBC	Creative Commons by Attribution	Data documentation exists describing method of data creation; services considered operational (test, scalable, redundancy in place)	Information has basic governance
Advanced	Metadata records are visible through multiple catalogues	Complete deployment for discovery, publication and harvesting	Partial optional interface deployment	TBC		Data support function in place; explicit service or operating level agreements exist	Information can be directly traced to business requirements
Mature	Domain independent discovery (for example, Google)	Complete deployment for discovery, publication, harvesting and remote update	Complete interface deployment	TBC		Tools in place to monitor data service reliability	Information process is thoroughly governed

5.1. NEII conformance – worked example

To illustrate the application of the framework Table 2 provides an example against recently released NEII data services. The Bureau of Meteorology has released the Australian Hydrologic Geospatial Fabric (Geofabric) as operational NEII web data services. The Geofabric data products are available as simple web services through Web Map Services (WMS) and Web Feature Services (WFS) on operational-grade Bureau of Meteorology infrastructure. The collation of Geofabric information follows a formal and informal process from data creation and aggregation to public delivery. An overall maturity score can also be created by

calculating the mean of six elements (on a five-point scale for each maturity level) to provide an ordinal ranking of node conformance, and ultimately overall progress of the NEII programme. At an overall maturity level the Australian Hydrologic Geospatial Fabric is considered as advanced as shown in Table 2.

Table 2. Worked example: NEII Conformance Framework applied against the GeoFabric

Name	Australian Hydrologic Geospatial Fabric		
Abstract	The Geofabric is a digital database of surface and groundwater features that registers the spatial relationship between important hydrologic features. It comprises six product datasets: hydrology reporting regions; hydrology reporting catchments; surface catchments; surface network; surface cartography; and groundwater cartography.		
	Element	**Maturity**	**Description**
Discovery	Discovery	Advanced	Geofabric structured metadata are discoverable through the Bureau of Meteorology NEII catalogue search interface at www.neii.gov.au\catalogue. Metadata records are available at a collection and data series level for both content and end point capability through the Reference Architecture defined standards. Data and services can also be found at data.gov.au.
Access	Metadata	Advanced	NEII Catalogue Service for the Web endpoint is publicly available, and provides additional paths for data and service discovery that includes programmatic access. Metadata services can be obtained through GET requests only via Catalogue Service for the Web. Automated harvest is one way *from* the NEII instance *to* external catalogues (for example find.gov.au).
	Geographies	Advanced	GeoFabric feature-types are available as both Web Feature Services and Web Map Services. All mandatory Web Map Service interfaces are available for machine-to-machine operation. Partial Web Feature Service interfaces are available.
	Gridded data	NA	NA
Re-Use	Licensing	Mature	Data are available under a Creative Commons by Attribution licence
	Operational	Advanced	Geofabric NEII services are available on Bureau of Meteorology Category 3 infrastructure. This includes formal operating level-agreements. Use of the data services are logged. User support is in place to NEII data service users through system owner (ISS existing escalation processes) and Geofabric team. Formal testing and deployment have been completed before data services were released.
	Information	Advanced	Geofabric product is governed through formal mechanisms between the Bureau and Geoscience Australia, and Geofabric has an explicit information model. The governance of the content and formal management of the update process, including version redundancy, strengthens this aspect. Vocabularies have been established.
	Overall	**Advanced**	

6. DISCUSSION AND CONCLUSIONS

The National Environmental Information Infrastructure is a long-term programme with a key dependency on partners to build a standard-based federated environmental information system. Its implementation remains both a change-management and collective-impact challenge and the programme has been designed to address both of these. The NEII offers environmental information users, across government, industry and research, major benefits even in its early stages when the portfolio of conformant data is small. Learnings to-date have highlighted that a challenge confronting the programme is the differing partner maturity regarding their ability to implement NEII conformant standards for their data. This occurs because the cost of participation (i.e. delivering data through publically available web data services on operational infrastructure) can be high for some custodians in the absence of a strong internal business need to adopt NEII-based standards. To address this issue the NEII programme is designing engagement and support processes to at least partially decrease the cost of participation for data custodians. The development of the NEII conformance framework and its application of a capability-maturity model goes some way to addressing variation in maturity while making the possible progression explicit to users. Beyond the approach communicated in the conformance framework, technical solutions may also be needed to decrease the cost of participation for custodians. This could include offering an NEII data service hosting capability provided by a mature organisation, or developing cloud-based deployment and hosting solutions to overcome security and policy barriers that may exist in some organisations. And finally an emerging challenge confronting the programme, once adoption of NEII-based standards by data custodians is further progressed, is how to support non-experienced web data services users in their use of the infrastructure to support their decision-making needs. This later challenge recognises that the standards promulgated in the NEII Reference Architecture may be new to many users, and without a structured education and outreach component their uptake may be limited to the more technical users, such as those developing web applications that require data delivered using standards-based services.

ACKNOWLEDGMENTS

REFERENCES

Bureau of Meteorology, 2015a. National Environmental Information Infrastructure: information modelling discussion paper. Environmental Information Programme Publication Series Document no. 5. Bureau of Meteorology, Canberra.

Bureau of Meteorology, 2014b. National Environmental Information Infrastructure: Reference Architecture V1.1. Environmental Information Programme Publication Series Document no. 4. Bureau of Meteorology, Canberra.

Open Data Institute, 2015. Open Data Certificate. Open Data Institute: London. Last accessed 14 May 2015. Accessible via https://certificates.theodi.org/about.

Office of the Australian Information Commissioner 2015, Principles on Open Public Sector Information, Office of the Australian Information Commissioner. Last accessed July 1, 2015. Accessible via http://www.oaic.gov.au/information-policy/information-policy-resources/information-policy-agency-resources/principles-on-open-public-sector-information