National Environmental Information Infrastructure Information Modelling Discussion Paper

Contributing to the Australian Government National Plan for Environmental Information initiative

National Environmental Information Infrastructure - Information Modelling Discussion Paper

Environmental Information Programme Publication Series, document no. 5
ISBN to be included

Environmental Information Programme
Bureau of Meteorology
Email: environment@bom.gov.au
www.bom.gov.au/environment

Citing this publication
Bureau of Meteorology (2015), National Environmental Information Infrastructure Information Modelling Discussion Paper, Bureau of Meteorology, Canberra, Australia, pp. 13.

With the exception of logos or where otherwise noted, this report is licensed under the Creative Commons Australia Attribution 3.0 Licence. The terms and conditions of the licence are at:
www.creativecommons.org/licenses/by/3.0/au

Contents

1 Introduction

The National Plan for Environmental Information (NPEI) initiative was established in 2010 to improve the quality and accessibility of Australian environmental information. The initiative is being jointly implemented by the Bureau of Meteorology (the Bureau) and the Department of the Environment. A core activity under the NPEI is the development of a National Environmental Information Infrastructure (NEII) to improve discovery, access and re-use of national environmental data.

The NEII is an information platform designed to improve discovery, access and re-use of nationally significant environmental data. Its primary focus is on data that are already well-managed, but have limited application beyond its original business purpose. The NEII is envisioned as distributed environmental data services adopting standards for interoperability. 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. These design elements are more fully described in the NEII Reference Architecture (Bureau of Meteorology, 2014) and the NEII Roadmap (Bureau of Meteorology, 2014).

The development of *information models* is a core component of the NEII. Information models are a tool that can be used to formally capture the semantics of data in a system; to describe the structure and meaning of data in the system. They are independent of particular data formats, but by adopting Spatial Data Infrastructure (SDI) principles and international standards for geographic information we can benefit from exchange formats that can be automatically derived from information models. By defining shared information models, data can be more readily exchanged between different parties and services, something that is essential in a distributed information infrastructure like NEII.

Although critical to the long term delivery of NEII, widespread awareness of the role of information models does not exist. The purpose of this discussion paper is to provide a high level overview of the role of information modelling in the context of NEII, introduce key elements of their development and introduce some of the standard to be adopted for information modelling within NEII.

2 What is information modelling?

Information modelling is the process of reaching a shared understanding of the nature and meaning of the data we are interested in, and documenting this shared understanding in a formal way.

The NEII Reference Architecture defines information models and classes as:

- **Information Model:** a formalised description of the logical structure and semantics of one or more information classes and their relationships.
- **Information Class:** a set of information objects with sufficiently similar characteristics that they may be dealt with in a common way within a system.

An information model is an implementation-neutral tool for describing key data concepts and the meaning of those concepts. Implementation-neutral means that the information model isn't limited to data stored in a particular format or delivered using a particular technology; it is an independent definition of the data that can be applied across different systems and tools.

An information model that is agreed upon within a community of practice is a direct representation of the community's knowledge and understanding of data within that domain.

Different organisations often choose to capture and manage similar information in very different ways. Furthermore organisations frequently have implicit information models which are understood only by those who work closely with the data or data format, that is, their format defines their information model. Difficulties often arise when organisations wish to share data or interpret data from other organisations. Without a clear, well-understood information model it can be difficult or time-consuming to exchange data with others or to understand data from others with a high level of confidence.

An information model takes a step back from particular formats or encodings and seeks to specify the common information types (classes) that are of primary interest to a domain, independent of how they might be currently captured or encoded. The information model is formalised using an implementation-neutral format (typically Unified Modelling Language, UML) that can describe: the classes of information we are interested in, the properties these things have, and the relationships between other classes of information.

The information model provides value because it:

- formally documents a community's domain understanding in a controlled way.
- is implementation-neutral (it is not linked to a particular format or system).
- acts as an agreed reference point for the community when exchanging data or developing systems to work with data.
- provides a common structure that heterogeneous systems can map to for conversions, exchange and delivery of data.

- offers a way to automatically derive an exchange format.
- also provides a way to derive other products, such as documentation or database schemas.

The development of an agreed, shared exchange format is often a primary driver for the development of a shared information model. In distributed systems like NEII, organisations have a need to share their data with others without fundamentally changing their underlying systems. When multiple parties are involved, the most efficient approach for exchanging data is to establish a shared exchange format that can be used to share or expose data either directly or via web services. It is possible to automatically derive an exchange format from a well-defined information model; this is known as the Model Driven Approach.

There are many well-documented information models in existence already. Most of these are associated with an Extensible Markup Language (XML) exchange format. Table 1 lists several of these.

Information Model	Description	Source
AIXM	Aeronautical Information Exchange Model	European Organisation for the Safety of Air Navigation (EUROCONTROL)
GeoSciML	Geoscience information model and exchange format	Multiple Geoscience organisations (via OGC)
INSPIRE Data Models	A broad range of environmental information models	European Commission
ISO/OGC Observations and Measurements	Generic information model and encoding for Observations	ISO and OGC standard
OGC WaterML2	Water Information Model	Multiple Hydrology Organisations (via OGC)
WDTF	Water Data Transfer Format	Bureau of Meteorology and CSIRO
WXXM	Weather Information Exchange Models	Federal Aviation Administration (FAA) and EUROCONTROL

Table 1: Examples of established information models and exchange formats

3 Standards for information modelling

Building an information model is a collaborative effort and requires a framework of agreed rules and patterns to enable effective collaboration.

The National Environmental Information Infrastructure, through the Reference Architecture, has adopted the International Standards Organisation Technical Committee 211 for Geographic Information/Geomatics (ISO TC211) suite of standards for Geographic Information across both services and information. The ISO TC211 suite of standards is a structured set of standards for digital geographic information. Since almost all environmental data has a geographic (spatial) data component the TC211 standards are highly relevant to environmental data.

From an Information Viewpoint the ISO TC211 standards provide a set of fundamental information models and encodings in Geography Markup Language (GML) for spatial, temporal and geometric information, as well as providing a 'meta-model' for information modelling; defining the 'General Feature Model' which can be applied to develop information models for particular domains or communities.

ISO TC211 has been widely adopted as a core framework in other national and international Spatial Data Infrastructures (SDIs) and by adopting this framework NEII is well aligned with wider, global information sharing activities. The Open Geospatial Consortium has also co-adopted many of the ISO TC211 concepts in its architecture, which is documented in freely available specifications.

NEII follows the ISO TC211 standard *ISO 19110 Geographic Information Methodology for Feature Cataloguing* and uses Unified Modelling Language (UML) as the format in which information models are defined. UML provides a visual representation using diagrams which are useful for unambiguous communication of concepts. UML is also machine-readable so it can be used as a basis for automatically deriving products such as encoding schemas or documentation.

The methodology for developing an information model is outlined in ISO 19110 Methodology for Feature Cataloguing, along with ISO 19109 Rules for Application Schema. The key process is outlined in Figure 1. The process describes starting the information modelling by defining the 'Universe of Discourse', that is, defining the scope and limits of what is being modelled. Then the process describes how the feature types are formalised and an application schema (exchange format) may be generated. In practice these steps are an iterative, consultative process, starting at a high level and leading up to a formalisation of the feature types, their attributes, operations and relationships with other feature types.

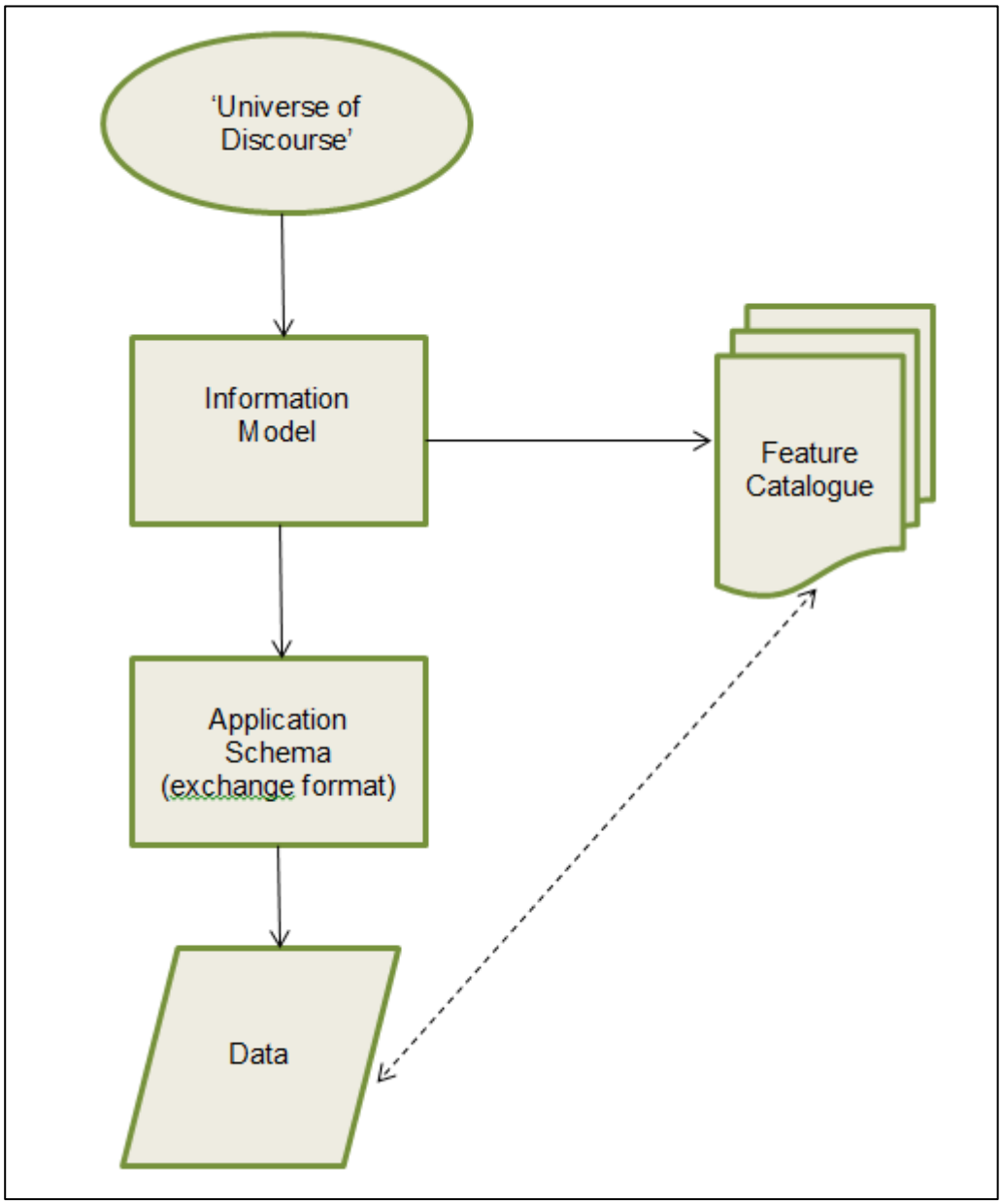

Figure 1: Method for Feature Cataloguing, based on ISO 19110

A significant benefit in adopting standard processes is that there are existing standards and information models that NEII can re-use or build upon. All of these information model 'building blocks' can be re-used as needed in domain specific information models. This avoids having to reinvent the wheel for individual information models and allows NEII to benefit freely from the significant effort that has gone into these standards.

4 Scenario—Air Quality Observations

This hypothetical scenario demonstrates how information modelling can work in a particular environmental domain.

Scenario: *Several NEII data providers wish to exchange observational Air Quality data via web services. There are no known common exchange formats in existence.*

Air Quality gap analysis

Having established the scope of the problem a 'gap analysis' is performed whereby existing information models are examined for suitability. No ideal Air Quality information model is found, although the ISO standard 19156 Observations and Measurements is deemed to be highly relevant and a strong candidate for re-use.

Air Quality information model

An information model for 'air quality observations' is collaboratively developed which specifies that there is an information type known as an 'Air Quality Observation' and each 'Air Quality Observation' has certain properties such as:

- the location at which the observation was made
- the time the observation was made
- the chemical or phenomenon that was observed (e.g. sulphur dioxide concentrations)
- information about the sensor used to take the observation
- metadata about who made the observation, and in what context
- … plus any other properties key to the understanding of the information
- … plus any relationships to other key information types.

These properties and relationships are captured formally as an information model in UML. Given that the Observations and Measurements model was deemed relevant the Air Quality model extends Observations and Measurements rather than building from scratch.

Information models frequently need to be supplemented with controlled vocabularies. In the Air Quality case a controlled vocabulary for the observed phenomena and chemical elements is adopted and agreed upon. These controlled vocabularies can be thought of as the 'values' that may be used for particular information properties in the model (such as chemical element names).

Air Quality exchange format

Once the information model is established there are many ways a set of air quality observations can be encoded according to the model. They could be stored in a spreadsheet or a database, or plotted on a chart or map.

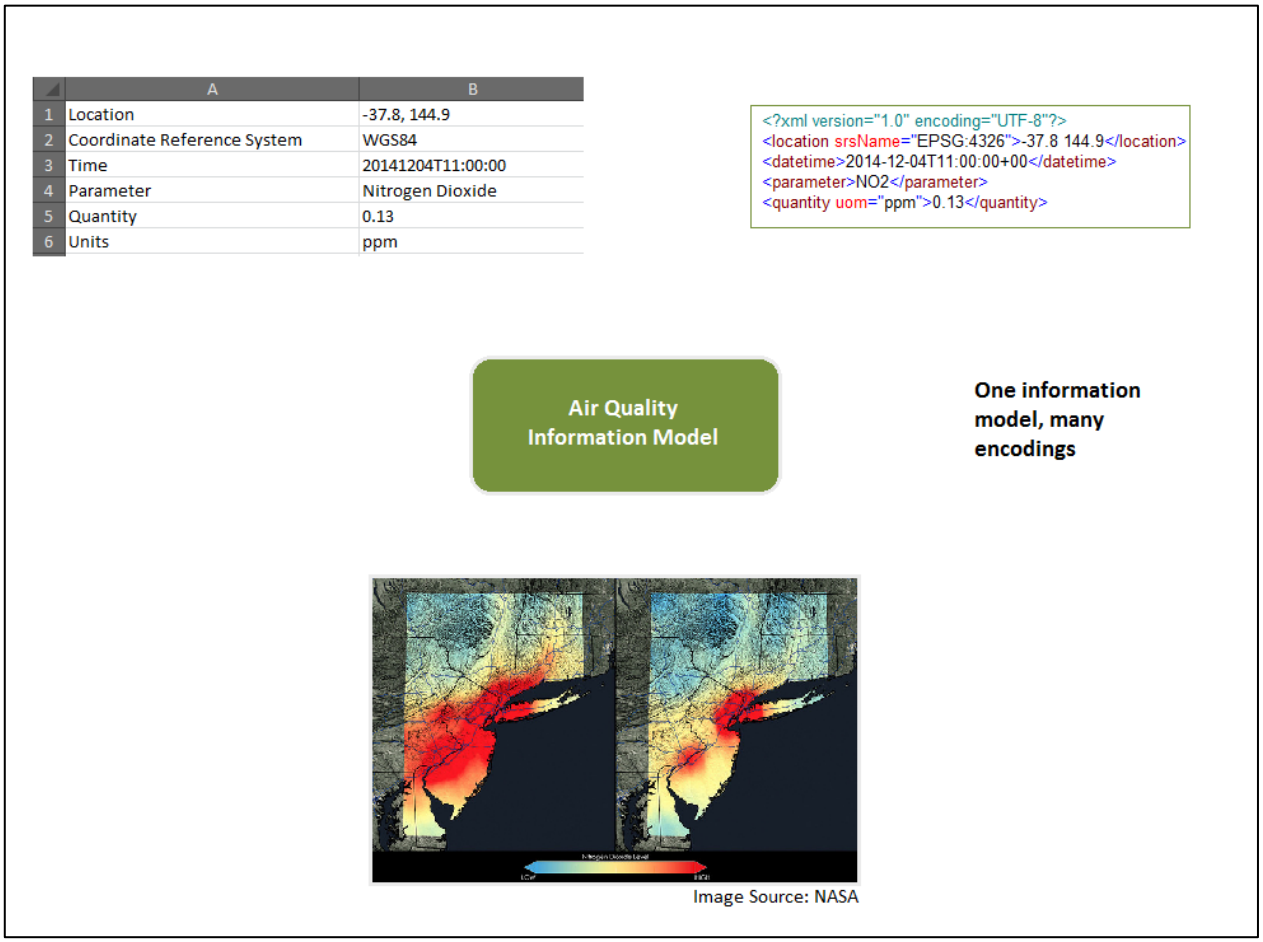

Figure 2: Figures showing different representations of the same information model

The primary driver for the scenario was to exchange Air Quality observations via web services. In order to meet this requirement an XML based interchange format is derived automatically from the UML model. This is done using the UML to GML rules identified in the ISO standards.

Air Quality data services

Having established an agreed encoding, based on ISO Observations and Measurements, the observations can be exchanged between data providers and users with OGC Sensor Observation Services. This is in line with the NEII Computation Viewpoint as defined in the NEII Reference Architecture.

5 Community Case Study: OGC WaterML 2.0

Information modelling is as much a social challenge as it is a technical one. A good information model requires broad engagement across a community.

Like many forms of environmental information, data about water (flow, level, quality) needs to extend across jurisdictional and organisational boundaries to provide adequate understanding and enable planning and mitigation around resources. Australia addressed this problem at a national level by developing WDTF (Water Data Transfer Format) to support the delivery of water data to the Bureau in a common format.

Meanwhile organisations in other countries developed similar water information models to deal with their problems locally.

In 2012 *OGC WaterML 2.0 Part 1 - Timeseries* was developed and standardised via the Open Geospatial Consortium (OGC) with input from several national organisations, providing an internationally agreed information model and encoding based on the best of the national standards as shown in Figure 3..

This model is now implemented in several water data management tools and provides a solid shared understanding of water timeseries observations for the water data community.

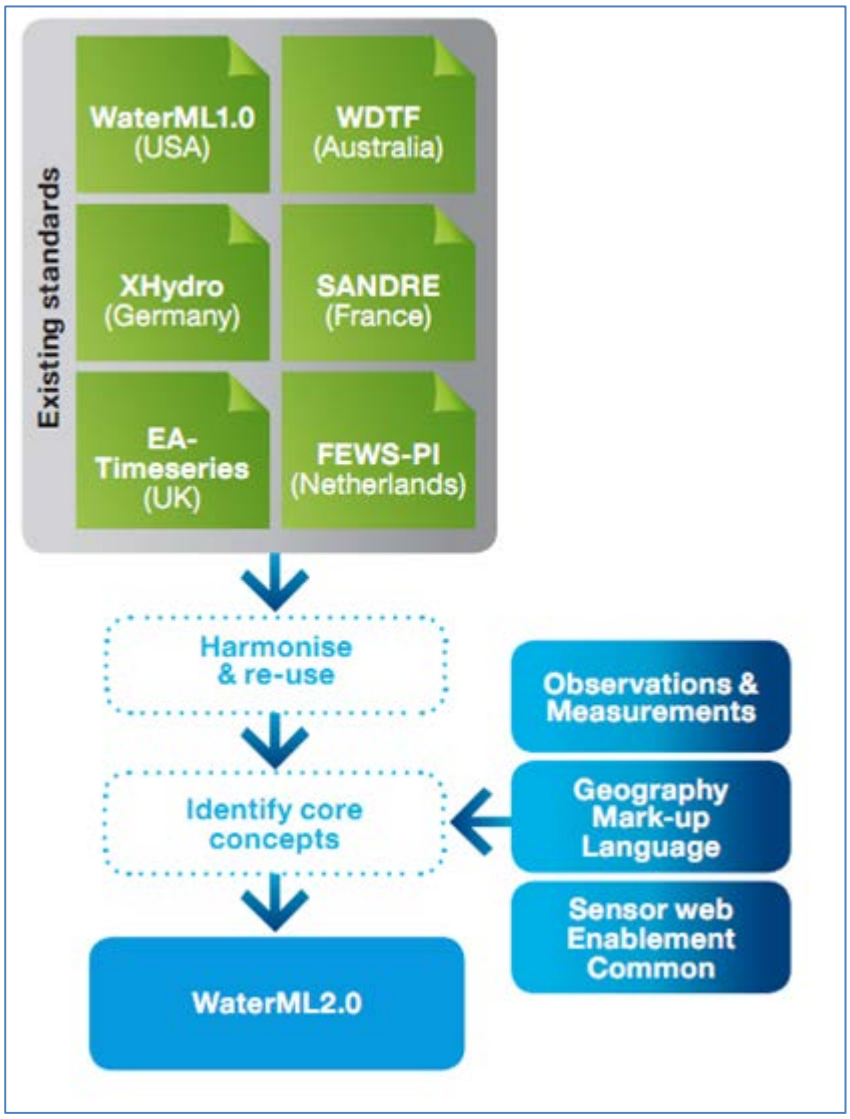

Figure 3: OGC WaterML2.0 Part 1: A model developed by taking into account existing standards. (Image: CSIRO)

OGC WaterML 2.0 is an example where a well-targeted information model can be established across many different organisations. OGC WaterML 2.0 Part 1 is narrowly focussed on the accurate description of timeseries data for hydrological observations. Other aspects of water data were ruled out of scope. Defining the clear scope of an information model from the outset is seen as a critical success factor. In ISO terminology this means defining the 'Universe of Discourse' before beginning to develop the information model.

6 Information modelling in NEII

NEII is broad in scope and covers a diversity of environmental data. Advanced data sharing will be achieved in particular domains when common information models are adopted to facilitate this exchange. For example, several organisations distributing data via OGC Web Feature Services can only do so effectively if they agree on a common information model for the 'Feature Types' that are served by the service.

Similarly for OGC Sensor Observation Service (SOS), it is helpful to tightly constrain the structure of the observations being delivered by the service. OGC WaterML 2.0 is a good example of this type of information model where the general Observations and Measurements model is constrained for a particular domain.

NEII will endorse particular information models (pre-existing or developed by NEII participants) as appropriate for use within the context of NEII, and will maintain a register of these information models.

NEII is not a standards organisation and therefore endorsement is not equivalent to standardisation by a formal group like ISO or OGC. Nevertheless endorsement by NEII is recognition that a particular standard has been subjected to a satisfactory level of rigour and formalism by domain experts from a particular community and is suitable for the exchange and sharing of environmental data within a distributed infrastructure.

The Bureau of Meteorology in its leadership of NEII will also help facilitate the development of information models as resources allow. However other patterns of participation are equally valid, for example the development of a community-driven information models..

NEII is based on established best practice and will make use of existing information models before developing new ones. In particular, the NEII Reference Architecture makes reference to the ISO 19156 Observations & Measurements standard. This is a key information model that can be applied, or profiled, across multiple domains. Where new information models are seen to be needed NEII will focus on enabling the development of narrow, high benefit, information models to assist in the delivery of key datasets.

7 Conclusion

Information modelling for specific environmental domains or applications is a key enabler for delivering a functional National Environmental Information Infrastructure. As the NEIIs understanding of the business drivers underpinning NEII matures the programme will establish a register for information models and associated mechanisms for the endorsement of specific information models. This will occur primarily through the activities of the NEII Reference Group.

8 References

Bureau of Meteorology (2014), *National Environmental Information Reference Architecture Version 1.2*, Environmental Information Programme Publication Series, Canberra, No. 4

Bureau of Meteorology (2014), *National Environmental Information Infrastructure Roadmap 2014-2019*, Version 1.1, Canberra, Australia, pp. 20.

National Plan for Environmental Information (NPEI): http://www.environment.gov.au/topics/science-and-research/national-plan-environmental-information

ISO TC 211: http://www.isotc211.org/

Open Geospatial Consortium, *Abstract Specifications*: http://www.opengeospatial.org/standards/as

Open Geospatial Consortium 2010, *Observations and Measurements - XML implementation*, OpenGIS standard 10-025r1, http://www.opengeospatial.org/standards/om

Open Geospatial Consortium 2010, *Topic 20 Observations and Measurements, OpenGIS standard 10-004r3*, http://www.opengeospatial.org/standards/om

Open Geospatial Consortium 2012, *Sensor Observation Service Interface Standard*: OpenGIS standard 12-006 http://www.opengeospatial.org/standards/sos

Open Geospatial Consortium 2012, *WaterML 2.0 Part 1 – Timeseries*, OpenGIS standard 10-126r4, http://www.opengeospatial.org/standards/waterml

Open Geospatial Consortium 2014, *Web Feature Service 2.0 with Corrigendum*, OpenGIS standard 09-025r2 , http://www.opengeospatial.org/standards/wfs

WaterML2.0: development of an open standard for hydrological time series data exchange, Journal of Hydroinformatics, Vol 16 No2 pp 425-446 http://doi.org/10.2166/hydro.2013.174

Appendix 1 Glossary of Terms

Bureau	Bureau of Meteorology
CSIRO	Commonwealth Scientific and Industrial Research Organisation
GML	Geography Markup Language
ISO	International Standards Organisation
ISO TC211	International Standards Organisation Technical Committee 211 for Geographic Information/Geomatics
NEII	National Environmental Information Infrastructure
NPEI	National Plan for Environmental Information initiative
OGC	Open Geospatial Consortium
OGC SOS	OGC Sensor Observation Service
OGC WFS	OGC Web Feature Service
O&M	Observations and Measurements (OGC and ISO standard)
RM-ODP	Reference Model of Open Distributed Processing
SDI	Spatial Data Infrastructures
UML	Unified Modelling Language
XML	Extensible Modelling Language