

The eReefs Information Architecture

Nicholas J. Car^a

^a *CSIRO Land & Water, Environmental Information Systems*
Email: nicholas.car@csiro.au

Abstract: eReefs is a large, multi-agency project that aims to deliver an information platform (eReefs-IP) which integrates existing and new data products. The eReefs IP will be used by both researchers and operational decision makers to better manage the Great Barrier Reef. Core to eReefs is the principle of interoperable systems whereby data generated and managed by a range of agencies such as the Queensland Government, the Bureau of Meteorology, AIMS and CSIRO, can be used together seamlessly.

The project runs for 5 years (2012 – 2017) and is implementing an information architecture for the eReefs-IP that has been developed from previous Australian and international natural resource distributed computing projects such as the Water Resource Observation Network, CUAHSI and INSPIRE. It is an NEII (National Environmental Information Infrastructure) demonstrator project and the lessons learned from eReefs will feed back in to the NEII and other NEII projects. In addition to delivering the eReefs-IP and new components for information architectures, eReefs is also extending the features and power of several water quality remote sensing and modelling systems.

eReefs is implementing a new distributed computing system approach to perform data service and support service tie-ins at the data provider level (data owner). This allows data providers to expose data set's data services in ways that enables them to be associated with the metadata and other services needed to ensure that it can be understood correctly. eReef contributions to information architectures include a new "orchestration service" that will help both the eReefs operators and data providers keep the eReefs community of systems running and growing over time. Additional architectural contribution is the first integrated use of a provenance service, PROMS, and extended use of the SISSVoc vocabulary service for data term definitions within data and metadata services.

This paper overviews the eReefs architecture at the highest level and provides a background outlining progression from earlier work to the current architecture. The new architectural concepts and components are discussed and some examples of implementation are given.

Keywords: *Great Barrier Reef, distributed computing, eReefs, WRON, National Plan for Environmental Information, data services, information architecture, Linked Data*

1. INTRODUCTION

Over the last decade in Australia, government agencies, research institutions and other environmental organisations have made significant progress in publishing, sharing and exposing environmental information. The use of standardised services and open licensing has enabled discovery and re-use however this evolution in data availability has occurred in an uncoordinated manner meaning: specific data and metadata service standards (such as Open Geospatial Consortium standards) are commonly used but what is missing is a widely adopted pattern of multi-service use and their interactions. This is important as data services can rarely be used without support services.

Data services must not only be self-describing but also discoverable linked to other support services that allow their integration into domain-specific communities. The challenge in Australia therefore remains to not only continue to implement standardised data and metadata services for data currently in silos but also to establish multi-service use patterns and whole-of-community interoperability.

Rather than developing a fully customised information suite which needs to be deployed in participating organisations, projects such as eReefs are developing frameworks which will allow existing systems and services to be better described, discovered and accessed in a consistent manner. eReefs acts as a pathfinder for the National Environmental Information Infrastructure (NEII) project so that architectural forms adopted by eReefs will influence a wide range of environmental data projects in Australia.

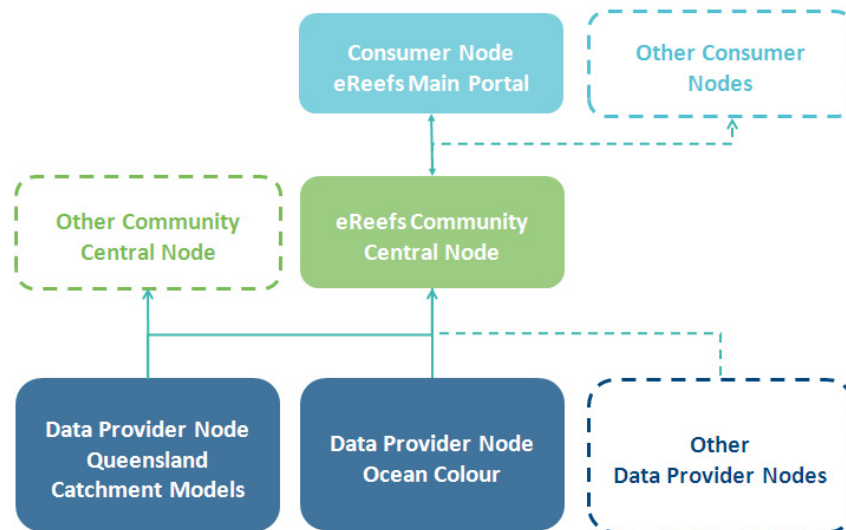


Figure 1: A high-level view of the eReefs architecture showing implementations of the three main architectural conceptual entities: Data Provider Nodes (DPNs), a Central Node and Consumer Nodes.

eReefs is developing the concept of distributed *Data Provider Nodes* (DPNs) which describe to contributing organisations how to supplement standardised data and metadata services with additional services such as vocabulary, persistent identifier and provenance services, and an overarching access methodology with a view to enabling their integrated use. The new supporting services can be implemented in-house, alongside a contributor's existing data systems, or in third-party locations linked to via the Linked Data paradigm. DPNs then expose data, metadata and other supporting information which any community – through a *Community Central Node* (CCN) – may integrate, add value to and then use for particular purposes. The eReefs project will support the development of an eReefs CCN, however the DPNs used by eReefs may also be used by any other community through their own CCN implementations. Data from DPNs accessed either directly or through a CCN are able to be utilised by any number of *Consumer Nodes* (ConNodes) which may be web portals, desktop programs or other systems that consume data and metadata. Figure 1 shows a high-level view of the eReefs information architecture including DPNs, CCNs and ConNodes.

This approach will allow for the next generation of data interoperability by augmenting established, standardised, services and allowing for the integration of multi-service use rather than adding a new standard.

2. BACKGROUND

eReefs builds on a number of interoperability initiatives, particularly the Water Resources Observation Network (Lemon *et. al.*, 2007), and has attempted to reuse its architectural principles and development techniques where possible. In addition to considering previous distributed computing architectures, an eReefs pilot was developed and tested in 2010 (Chen *et. al.*, 2011).

2.1. The eReefs Pilot

The eReefs pilot demonstrated the integration of a number of catchment, lagoon and ocean modelling systems and produced visualizations of their outputs in Google Earth (*ibid.*). In early published eReefs work including the pilot study, eReefs was referred to as an “integrated operational system of data sources” (*ibid.*) and undertook “model orchestration” running models in sequence through a workflow. The Pilot showed a single processing chain (series of models) and an end-user outcome (the visualization of final model results) but it did not demonstrate an architecture that employed principles of interoperability to facilitate multiple processing chains and multiple products, as is required of the eReefs project today. The Pilot developed two distinct tools: a Water Quality Visualisation Platform and an Orchestration Tool, both of which can be fitted into the eReefs architecture discussed in this paper as distinct ConNodes (see discussion in Section 3).

2.2. Developments since the eReefs Pilot

In accordance with objectives in the eReefs Collaboration Agreement, signed after the eReefs Pilot, eReefs will develop “Integrated spatial and temporal data *from a wide variety of sources;*” [emphasis added] (eReefs CJV, 2011). This requirement and the provisioning of a Work Package 2 team (WP2) within the project structure dedicated to “Interoperable data and information systems” means eReefs now has both the requirement and means to build an information platform to both deliver current core eReefs data and provide mechanisms for the inclusion of other data sources into eReefs now and in the future.

2.3. 2012 Scoping Study – using the WRON techniques

In September 2012, the WP2 team delivered a Scoping Study parameterizing the scope of future WP2 work (Car *et. al.*, 2012). This study surveyed strategic goals, stakeholder interests, known datasets and architectural requirements for eReefs’ information architecture and proposed the use of the Reference Model for Open Distributed Processing (RM-ODP) (ISO, 1998) as a general methodology for design and the Water Resources Observation Network Reference Model (WRON-RM) (Lemon *et. al.*, 2007) as an architectural model.

By adopting RM-ODP methodology, eReefs uses 5 ‘viewpoints’ for system description which are:

1. **Enterprise** - describes the business needs/drivers for the information system;
2. **Information** - describes what information flows around the system;
3. **Computational** - describes how the service interfaces are defined and interact;
4. **Engineering** - specific implementations of system components in their particular environments;
5. **Technology** - describes specific components required to perform the required system functionality.

The main implications for adopting WRON-RM architecture are:

1. **Adherence to interoperability ‘best practice’** - specific architectural principles ;
2. **Adoption of generic Use Case categories** - which prompt the consideration of Use Cases outside normal end-user Use Cases (the Use Case most commonly thought of in software design).

Finally the Scoping Study provided recommendations to guide WP2’s work. In dealing with stakeholder requirements, Use Cases etc., the Scoping Study is the RM-ODP’s *Enterprise Viewpoint* for eReefs.

2.4. Architecture document

An architectural document (a wiki website) was initially prepared in December 2012 and updated with continuing WP2 work (eReefs WP2 team, 2013). The wiki proceeds through the *Information, Computational, Engineering* and *Technology* RM-ODP *viewpoints* picking up from where the Scoping Study left off. It specifies that existing data services should be used to deliver datasets via the Internet and also what particular services should be used for particular data types. Due to the nature of eReefs data, these mostly consist of Open Geospatial Consortium standardized data services, such as the Web Feature Service (OGC, 2013). Particular service variants are chosen to match the expectations of the NEII team at the BoM.

3. ARCHITECTURE

In addition to providing an overall architecture for eReefs using an accumulation of existing standards and architecture principles, the eReefs architecture defines a series of components that, while similar to existing components seen elsewhere, are new in terms of their constitution (both the collection of existing services and standards used and also new services and standards) and their system integration. The overall eReefs architecture is given in Figure 1 **Error! Reference source not found.** and a detailing of DPNs, the CCN and ConNodes follows.

3.1. Data Provider Node concept

Data Provider Nodes (DPNs) are a conceptual entity that provides a collection of data and support services for one or more data sets according to a set of requirements. They are self-sufficient entities that need not have any knowledge of the eReefs or any other community. The set of requirements for an eReefs DPN, a subset of which can be thought of as the requirements for a generic DPN, are:

1. **Web Services:** datasets provided are to be available via standardized web services for data;
2. **OGC standards:** if they are geospatial datasets, they must implement relevant OGC standards;
3. **ANZLIC metadata:** datasets provided must have ANZLIC-compliant records¹;
4. **Metadata service:** metadata must be available via OGC’s Catalog Service for the Web standard;
5. **Vocabularies:** controlled vocabularies of specialized terms need to be provided for data within datasets. These terms must be delivered via the SISSVoc service²;
6. **Provenance:** data provenance for datasets designated by eReefs as requiring it must be delivered as a Provenance Management System (PROMS) service (see Section 4.3);
7. **Features:** a list of the features for which data is collected must be delivered via a Feature Type Catalog service³
8. **Brokering Layer:** the DPN constituent services must be described using eReefs’s DPN Service sBrokering Layer protocol (see Section 4.2);
9. **Registration:** all DPNs that wish to be considered part of the eReefs community must be registered in the Barry Orchestration Service (see Section 4.1) and pass Barry’s conformance tests;

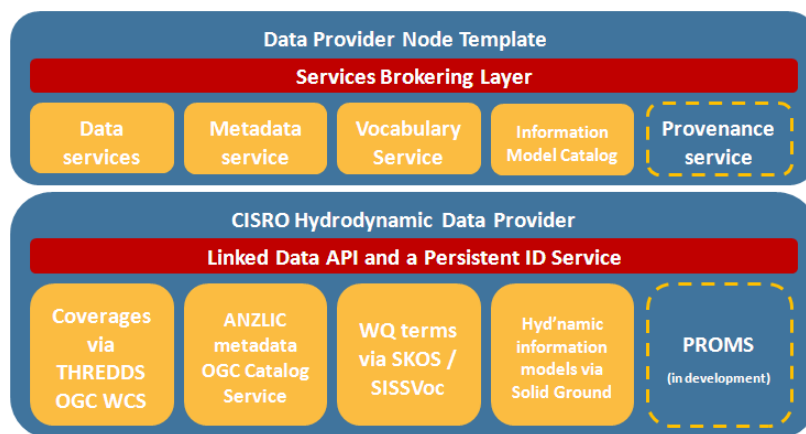


Figure 2 shows both the generic DPN structural template and an eReefs instantiation of that template: the CSIRO Hydrodynamic DPN which is currently under construction. In addition to the DPN structural requirements listed above, there are functional requirements which require that parts of the architecture operate together in certain ways. The navigability of the Services Brokering Layer (SBL) to service endpoints is one example, another is the linking of specialised terms used in data

Figure 2: Generic template of an Data Provider Node (top) as well as a specific implementation (bottom) for the eReefs CSIRO Hydrodynamic DPN. Provenance service implementation is not yet required for this DPN as it is in development.

service headers (e.g. netCDF files delivered via THREDDS) and metadata records to vocabulary service terms. This ensures DPN service actually interoperate, rather than merely coexist.

An agency may host DPN constituent services in any manner and in any location they choose; utilisation of a DPN only requires that its constituent services are discoverable and accessible by a community’s orchestration service and adhere to relevant standards. Using an SBL means that, at a minimum, a DPN owner need only host the SBL and use it to indicate where the other services are.

Any DPNs established for eReefs may also be utilised other communities that also implement a distributed information architecture, indeed even though eReefs DPNs will be registered with eReefs, they will deliver their constituent services in the same manner to any other consuming entity.

3.2. eReefs Data Provider Nodes

eReefs expects to see the implementation of at least nine DPNs which are:

- CSIRO’s Hydrodynamic data ([test link](#));
- Queensland Government Great Barrier Reef spatial layers⁴;

¹ ANZLIC Metadata Profile: An Australian/New Zealand Profile of AS/NZS ISO 19115:2005: <http://spatial.gov.au/sites/default/files/legacy/osdm.gov.au/Metadata/ANZLIC%2BMetadata%2BProfile/default.html>

² SISSVoc V3 Overview: <https://www.seegrid.csiro.au/wiki/Siss/SISSvoc3Overview>

³ See <https://www.seegrid.csiro.au/wiki/AppSchemas/FeatureModel>

- Queensland Government River Gauge data⁵;
- Queensland Government's Catchment Modelling outputs (perhaps same DPN as other QG data)⁵;
- Bureau of Meteorology's (BoM) Ocean Colour. This DPN's entry point is ([test link](#));
- Aust. Inst. Marine Science (AIMS) weather data ([link](#));
- Integrated Marine Observation System's (IMOS) timeseries data ([link](#));
- BoM's Geofabric Gazetteer ([link](#));
- Great Barrier Reef Foundation's (GBRF) Citizen Science data⁵;
- Other small datasets in a single 'eReefs' DPN;

It is reasonable to expect that the DPNs will be implemented by the agencies as listed above, not on their behalf by others, and that each DPN will contain more than one dataset. Data from many of the candidate DPNs listed above are already available online as services (see the given 'links') but none are yet full DPNs. Once established, it is expected that many of these DPNs could be utilised by communities other than eReefs. For example, the Bioregional Assessments project⁵ could utilise the Bureau of Meteorology's Geofabric Gazetteer and Queensland River Gauge data to assess water quality of water in particular catchments.

3.3. Central Node (concept and eReefs implementation)

The eReefs CCN is the only piece of the information infrastructure that an on-going, operationalised, eReefs team will need to maintain indefinitely. Its job is to coalesce data and other services from the DPNs determined to be in, or of relevance to, the eReefs project. Its role is *not* to present or visualise data from DPNs or to supply new data of its own unless that data is required to bridge two or more datasets. Its role is to *broker* access to DPNs and provide community support services. In one mode it acts as an SBL across multiple DPNs. Core services or architectural components provided by the CCN are:

1. An eReefs **Identifier Service** operating similarly to a DPN SBL
2. An **Orchestration Service** to monitor the eReefs collection of DPNs and their constituent services;
3. A **Services Master Catalog** listing all eReefs DPN's services;
4. A **Master Metadata Catalog** listing all eReefs DPN's metadata entries;
5. A **Master Feature Type Catalog** listing all features addressed by eReefs DPNs' data;
6. A **Vocabulary Master Service** containing vocabulary terms specific to eReefs (mostly water quality-related) and caching vocabulary terms served by DPNs;
7. One or more **Spatial Gazetteers** delivered as services. They may take the form of a DPN of spatial layers hosted by an eReefs partners, such as Queensland Government.

3.4. Consumer Nodes (concepts and examples)

Traditionally, distributed systems of similar data have delivered access to their data holdings via a portal or other similar interface. An example close to eReef's domain is the Australian Ocean Data Network Portal⁶ which provides access to IMOS' datasets as well as others. eReefs aims to facilitate a number of data consumption methods, thus the concept of a ConNode which is then any program or service that consumes data and services from the eReefs community either via the CCN or directly from a single, or several, DPNs. The eReefs portal⁷, modelled on the AODN Portal, is indeed a ConNode but other ConNodes are also possible. While other ConNodes have not yet been completed for eReefs, the following are anticipated to be in development by December, 2013:

- A **mobile 'app'** indicating water quality based on a fusion of the Hydrodynamic and Ocean Colour datasets. The target users are SCUBA dive operators and scientific field staff;
- A **workflow engine** that is both a ConNode, as it will collect data from more than one DPN and generate a derivative datasets, and a DPN as it will deliver its own results for use in eReefs;
- An **eReefs Portal** that acts as a general, visual, access point for eReefs spatial data layers;
- A **Marine Water Quality Dashboard** provided by the BoM that delivers water quality data from eReefs and possibly other, non-reef-based, sources;
- An **automated report card generator** used to generate future versions of the Reef Report Card⁸;
- An **alert service** that sends notifications (email, SMS and other) triggered by users' rules.

⁴ Data services not publicly available yet

⁵ The "Bioregional Assessments project" mentioned here is a CSIRO-led project implementing the project detailed at <http://www.environment.gov.au/coal-seam-gas-mining/bioregional-assessments.html>

⁶ <http://portal.aodn.org.au/aodn/>

⁷ <http://ereefs.org.au/Portal2/>

⁸ <http://www.reefplan.qld.gov.au/measuring-success/report-cards.aspx>

4. NEW ARCHITECTURAL SERVICES & COMPONENTS

The majority of architecture componentry and services use in eReefs are not new in eReefs but are mature, often internationally standardised, systems. There are, however, some new components and services that will move eReefs beyond current systems. They are detailed in the following subsections.

4.1. Barry – the Orchestration Service

Central to eReefs continuing functionality is a model-based *orchestration service*, named “Barry”⁹ that can be expected to be implemented by future communities with eReefs-style architecture. It will perform several roles, firstly acting as a web crawler service, similar to Google’s Webmaster Tools¹⁰, iterating through all of eReefs’ DPNs’ services and some datasets running a series of tests on them. Figure shows the central and peripheral tests Barry will run. It will check to see if DPNs and their services are online, if they are configured correctly and have all of the attributes that the eReefs requires them to have. Secondly, it will also report on DPN and other architecture component’s status both to future eReefs staff and also to DPN owners. Since a DPN may belong to several communities which may run variants of Barry, DPN owners may receive multiple reports. Thirdly, it will allow the owners of candidate DPNs and new datasets presented in existing DPNs to build towards eReefs Architecture conformance using a test-driven approach. They will be able to submit their DPN or new dataset to Barry and see its performance against the test suits.

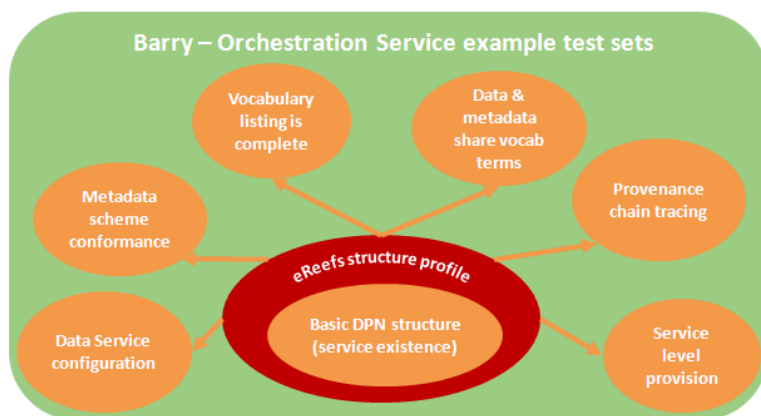


Figure 3: Barry Orchestration Service central and peripheral test modules.

The “Basic DPN structure” test set is a subset of the “eReefs structure profile” that can be expected to be implemented by systems wishing for eReefs compliance on top of generic DPN compliance.

Since tests are modular and modules are executed in a specified order, particular modules are only run if appropriate for a particular DPN based on that DPN’s given service description in OWL.

Tests are run on a DPN simply by registering its entry point URI with Barry.

4.2. Services Brokering Layer

An SBL (see Section 3.1 for context) is a brokering layer that, through an established a well-known URI pattern set, links together the various constituent services of a DPN. It uses an OWL¹¹ ontology to describe the DPN’s constituent services, instances of which are used to generate a series of Linked Data “views” of the services, such as persistent URIs with redirection facilitated by the Persistent ID Service¹². Part of a persistent URI pattern set linking to services, the testing of which would amount to a portion of the tests within the “Basic DPN structure” indicated in Figure 2, look as shown in **Error! Reference source not found.** The links in Table 1 de resolve to a test implementation of CSIRO’s Ocean Colour DPN.

4.3. Provenance Service – PROMS

The Provenance Management System (PROMS¹³) is a new web service designed to capture PROV ontology¹⁴ provenance data from automated workflows. Systems processing eReefs data then delivering it back to eReefs will be required to record process provenance and make it available via a PROMS installation. Since PROMS is currently under active development, exact implementation details are still unknown.

5. ENGAGEMENT AND ROLLOUT

As eReefs serves as a pathfinder for the National Environmental Information Infrastructure which is part of the National Plan for Environmental Information (DSEWPac, 2012), architectural work done in eReefs will

⁹ Named after the Australian composed Barry Singh, see <https://wiki.csiro.au/display/NE21/Barry+-+Orchestration+Framework>

¹⁰ <https://www.google.com/webmasters/tools/home?hl=en>

¹¹ Web Ontology Language: http://en.wikipedia.org/wiki/Web_Ontology_Language

¹² <https://www.seegrid.csiro.au/wiki/Siss/PIDService>

¹³ The PROMs wiki at <https://wiki.csiro.au/display/proms> is the most authoritative source of information on this developing service.

¹⁴ <http://www.w3.org/TR/prov-o/>

Table 1: Basic DPN structure URI pattern set

both be watched by and need to watch developments in other NEII projects.

We expect design choices here to be modified as the needs of other projects are taken in to account since achieving cross-project interoperability is an NEII imperative.

As of September, 2013, the Bioregional Assessment project is the only other NEII project, however smaller projects are looking to use the DPN concept and other parts of the eReefs architecture. Currently there is no cross-community index of NEII-compliant DPNs however this is expected to be a very useful register to establish.

| URI (relative to base URI) | DPN Element |
|--|--|
| http://dpn-oc-vc.nexus.csiro.au/ | DPN base URI |
| / | A human-readable web page |
| /config | Services Brokering Layer entry point |
| /config/datasets/ | Index of the DPN's datasets |
| /config/datasets/{dataset_ID} | Entry point for a dataset, redirects to a the alternates view (list of other views) |
| config/datasets/{dataset_ID}/?_view=meta&_format=html /config/datasets/{dataset_ID}/?_view=meta&_format=csv | The metadata entry for a particular dataset in HTML and then XML via the Catalog Service for the Web |
| /config/datasets/{dataset_ID}?_view=vocab | A listing the vocabularies in which the terms used by this dataset are defined. |
| /config/datasets/vocab/term/{term} | Persistent URI for a vocab term. Likely to redirect to a more authoritative URI. |
| /config/datasets/{dataset_ID}/?_view=prov | Entry point to a PROMS register corresponding to a particular dataset |

6. CONCLUSION

The nodal architecture presented here lends itself to distributed deployment for whole communities, as does the service-brokered approach for data providers. It should also allow DPN, CCN and ConNode reuse. The data provider services meta-framework, given by way of an SBL, allows maximal existing service reuse and give a methodology for the inclusion of new services and, by them, new community requirements. This should prevent obsolescence and allow communities to evolve while retaining interoperability.

REFERENCES

- ANZLIC, the Spatial Information Council (2007). ANZLIC Metadata Profile: An Australian/New Zealand Profile of AS/NZS ISO 19115:2005, Geographic information – Metadata (implemented using ISO/TS 19139:2007, Geographic information Metadata XML schema Implementation. ISBN: 978-0-646-46940-9.
- Car, N.J., Fitch, P.G. and Lemon, D. (2012) Scoping Study: eReefs Work Package 2 – Interoperable data and information systems. CSIRO, Aust. ISBN 978-1-922173-07-2 (print), 978-1-922173-08-9 (electronic).
- Chen, Y., Minchin, S.A., Seaton, S., Joehnk, K.D., Robson, B.J. and Bai, Q. (2011). eReefs – a new perspective on the Great Barrier Reef. Proceedings of MODSIM2011. pp. 1652-1658. ISBN: 978-0-9872143-1-7. <http://www.mssanz.org.au/modsim2011/C4/chen.pdf>. MSSANZ.
- eReefs Collaborative Joint Venture (CJV) (2011). Collaboration Agreement in relation to the development of the eReefs system. Final copy signed, 24th November, 2011. Unpublished.
- eReefs Work Package 2 (WP2) team (2013). eReefs Architecture. Wiki web page, online at <https://wiki.csiro.au/display/NE2I/eReefs+Architecture>. Accessed 7 July 2013. CSIRO Land & Water.
- International Organization for Standardization, (1998). ISO/IEC 10746-1:1998 (E). Information technology – Open Distributed Processing – Reference model: Overview. Geneva: ISO.
- Lemon, D., Cox, S., Walker, G. Atkinson, R., Fitch, P., & O'Hagan, R. (2007). *The Water Resources Observation Network – Reference Model Version 0.1*. Canberra, Australia: CSIRO: Water for a Healthy Country National Research Flagship. <http://www.csiro.au/files/files/pku7.pdf>.
- Department of Sustainability, Environment, Water, Population and Communities (2012). The National Plan for Environmental Information. Web page, online at <http://www.environment.gov.au/npei/>. Accessed 7 July 2013. Environmental Information Policy Section, DSEWPaC.
- Open Geospatial Consortium (OGC) (2013). OGC Standards. Web page, online at <http://www.opengeospatial.org/standards/is>. Accessed 7 July 2013. Open Geospatial Consortium.