

Flexible and modular visualisation and data discovery tools for environmental information

Yu, J. ^a, Hodge, J. ^d, Leighton, B. ^a, Seaton S. ^c, Vleeshouwer, J. ^b, Tickell, S. ^d and Car, N.J. ^b

^a Land and Water, CSIRO, Clayton, Victoria

^b Land and Water, CSIRO, Dutton Park, Queensland

^c Land and Water, CSIRO, Black Mountain, ACT

^d Oceans and Atmosphere, CSIRO, Dutton Park, Queensland

Email: jonathan.hodge@csiro.au

Abstract: Recent advances in environmental sensing, high-resolution satellite technology, advanced environmental modelling, information platforms, and virtual laboratories present significant opportunities for providing scientific communities and the general public with the ability to discover and gain insights into our ever-changing environment. However, barriers to realising the value of available environmental datasets exist, specifically, in the (lack of) streamlined access to datasets, and tooling to flexibly support the mash-up creation of various visualisations. In previous work, a Data Brokering Layer has been developed to support the streamlining of potentially heterogeneous access mechanisms and data formats.

The Data Brokering Layer provides functionality to handle data queries and access to registered datasets and data services, where in the past, visualisation portals would have required ad-hoc code to handle access to each dataset individually, thus limiting the extensibility of such data portals further. Thus, the Data Brokering Layer provides a basis on which a library of visualisation and data discovery modules can be developed. Such visualisation and data discovery modules provide flexible, reusable and mashable components for exploring datasets multiple contexts across multiple domains. Examples of visualisation modules include data portals, embedded visualisation panels, and real-time reports.

In this paper, we focus on the visualisation and data discovery capabilities developed in the context of the eReefs project which aim to provide an integrated information platform for discovery and visualisation of observational and modelled data of the Great Barrier Reef.

Keywords: *Great Barrier Reef, visualisation, data interoperability, data discovery, environmental information systems*

1. INTRODUCTION

Recent advances in computing, sensing and information technologies have given scientific communities the ability to capture and analyse environmental information at greater spatial and temporal resolution than ever before. These communities are increasingly publishing data online through the web and via cloud services, providing access to high resolution and broad environmental data e.g. soils and geology (Laxton *et al.* 2010) and oceans (Levitus *et al.* 2013).

However, a challenge is streamlining discovery, access and use of the range of data available across disciplines and across information platforms. Barriers exist for users in being able to query data in a precise way, such as, show me all data for ‘chlorophyll concentration’ within a given time period and spatial region, where the data itself is from different providers and uses different notation for the observation parameter, for example, ‘CHL’, ‘Chlorophyll a’, or ‘Chlorophyll-a’. Another barrier is the multiple access mechanisms and interfaces to the data, e.g. OGC Web Feature Services, OGC Map Services, OPeNDAP interfaces, bespoke REST Web APIs, and other proprietary services. Lastly, the integration of the data into client applications often requires knowing which data services to bind to a-priori and encoding the handling of the semantics of each data service in the client application (e.g. a private contract) resulting in static bindings to the data.

An example of this challenge is found in the Great Barrier Reef domain, where multiple data sources and formats exist around the movement and changes in water and water quality moving through catchments and into the Great Barrier Reef lagoon. The eReefs project aims to deliver a web-based information platform to link the various sources of such environmental data. Examples of data sources include in-situ water quality monitoring data, satellite remote sensed water quality data, and modelled hydrodynamic and biogeochemical data. Each of these data sources may involve accessing several services provided by multiple organisations.

In order for this challenge to be addressed, new methods and tools are required to enable systematic data description and discovery across a diverse range of web services to support the delegation and abstraction of the data and services layers to the underlying information architectures. This will subsequently allow discoverable and dynamic bindings to these components in client applications. In previous work, an ontology for describing an Observable Properties¹ (Cox *et al.* 2014b) and a Water Quality vocabulary² (Simons *et al.* 2013a, Simons *et al.* 2013b, Simons *et al.* 2014) was developed to enable precise semantic annotation of domain concepts in the data for eReefs and as best practice for encoding water quality data more generally via WaterML-WQ (Cox and Simons 2014). Also proposed in prior work, a methodology for enhancing data services with semantic annotations (Yu *et al.* 2014), the netCDF-LD format for enhancing netCDF data with Linked Data semantics (Yu *et al.* 2015a), the Data Provide Node ontology for describing data providers, data and data services³, and a mediator component, called the Data Brokering Layer (DBL) (Yu *et al.* 2015b), have been proposed and developed. These provide ontologies, controlled vocabularies, methods and tools to support the streamlining of potentially heterogeneous access mechanisms and data formats and in the case of eReefs, they apply specifically for water quality data.

In this paper, we present visualisation and data discovery tools developed for the eReefs project which leverage the above prior work on enabling semantic discovery of environmental data. The aim of this work is to address the above challenge of data discovery, aggregation and visualisation of environmental data in a way that allows modularisation and reuse of the components. We present the eReefs Visualisation Portal, which provides a lightweight client application for visualising eReefs data via map and chart views which leverage the underlying DBL for eReefs. We also present the eReefs Data Explorer, which showcases the eReefs DBL content including the various semantic facets. The eReefs Visualisation Portal and the eReefs Data Explorer provide examples of reusing the DBL functionality for exploring data across multiple contexts in multiple domains and disciplines.

2. EREEFS INFORMATION PLATFORM

The eReefs information architecture, proposed by Car (2013), provides a blueprint for the intended information platform for eReefs. The aim of the information architecture is to accurately capture the various groups, data services, and datasets for eReefs as part of a federated and distributed network of researchers, organisations and resources. As part of the architecture, the concept of a Data Provider Node was introduced

¹ Online here: <http://environment.data.gov.au/def/op>

² Online here: <http://environment.data.gov.au/def/>

³ Online here: <http://purl.org/dpn> and <http://purl.org/dpn/services>

to capture the governance of a particular group over a set of data services and datasets which constitute a Data Provider Node (DPN). The configuration of each DPN may vary from one to another, but is required to be appropriately described to allow a given DPN to be understood by the eReefs information platform, preferably in an automated fashion.

For a DPN to be appropriately described and understood, machine-readable formalisms for describing it are required. This includes any links between a given DPN and any scientific domain concepts. We use ‘Linked Data’ approaches to describe DPNs and link them to domain vocabularies in order to capture the semantics. DPNs are described using the DPN ontology (<http://purl.org/dpn>), which provides the formalisms and language for capturing a given DPN, associated contact persons and organisation (via the Org ontology⁴), and available services and datasets. The DPN descriptions are then registered to eReefs and a Data Brokering Layer (DBL) is able to query and harvest the additional metadata from the data service endpoints specified in the DPN descriptions. Currently, DPN descriptions captured for eReefs consist mainly of THREDDS Data Service (TDS) endpoints for the Remote Sensing and Hydrodynamic services, although any data service endpoint could be represented using the DPN ontology. The eReefs DBL API supports data discovery functions for eReefs client applications and is available online at <http://ereefs.org.au/dbl>. In addition to a lightweight API to query available DPNs, services and datasets, the DBL provides a mediating layer between the DPN implementations and the client applications.

In order to enhance the metadata captured for each DPN, Yu *et al.* (2014) developed a methodology for enriching the metadata published by each DPN with semantic annotations using HTTP URIs, which uniquely identify domain concepts relevant for each parameter being delivered for a given data service. This allows the DBL to index the services and datasets with these identifiers and thus provide enhanced discovery and search mechanisms to end-user client applications. References to the domain concepts resolve to the Water Quality controlled vocabularies which have been defined for eReefs and are deployed via a SISSVoc vocabulary service (Cox *et al.* 2014a). These are available online at <http://sisvdoc.ereefs.info/>.

The semantic annotations refer to any substance-or-taxon, quantity, units of measure and object being referenced for a given parameter. They can then be harvested by the eReefs DBL which in turn exposes them via the DBL APIs for the client applications upstream as shown in the figure below. The DBL provides the ability to index the semantics for the registered DPNs and facilitates semantic searching for relevant datasets. This allows precise querying of datasets from client applications such as the eReefs Data Explorer and the eReefs Visualisation Portal, which are discussed in the next section.

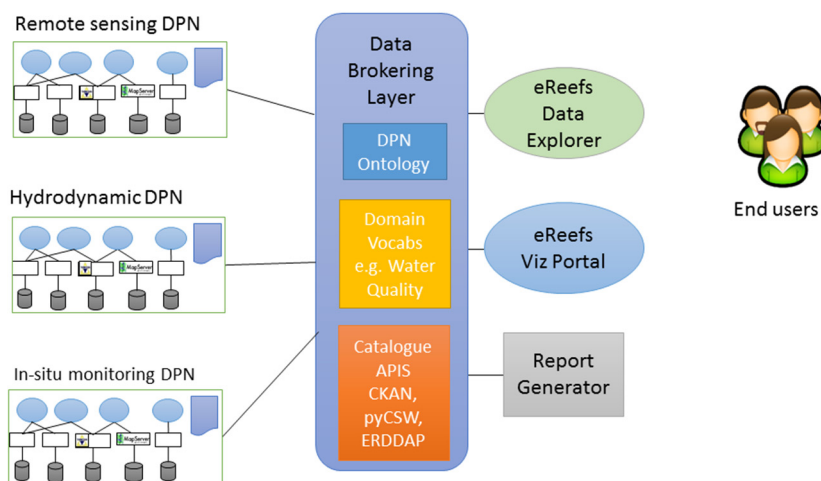


Figure 1. eReefs information platform.

3. VISUALISATION AND DATA DISCOVERY TOOLS

In this section, we introduce the eReefs Visualisation Portal and the eReefs Data Explorer which provide end-user client applications for eReefs. The eReefs Visualisation portal allows primarily spatial views of the data and a lightweight search function to assist users to add data layers over a map visualisation. The eReefs Data Explorer provides a data and domain concept perspective to assist users to explore and discover data and understand the semantic conceptual space for eReefs.

⁴ Available here: <http://www.w3.org/TR/vocab-org/> (Accessed 30 July 2015)

3.1. eReefs Visualisation Portal

The eReefs Visualisation Portal is designed to help users explore or search for eReefs data and have the selected data layers visualised via a map-based portal (see Fig. 2a). The Visualisation Portal primarily allows users to plot the data from distributed sources in the spatial context of the Great Barrier Reef. The eReefs DBL is used as the backend API for handling querying data using the underlying DPN and domain semantics, listing available data provider nodes, associated metadata about data services and datasets. These aspects are handled dynamically within the portal and demonstrates how a client application can leverage the underlying eReefs information infrastructure.

The eReefs DBL and the eReefs SSISSVoc Vocabulary service are used to power the search functions in the Visualisation Portal. An example is given in Fig. 2b, where a user has submitted a search query string of “chlorophyll”, which lists relevant controlled vocabulary concepts for the search term (and the URI which uniquely identifies the concept). The Visualisation Portal then queries the DBL in the background for data matching the vocabulary concept URI, e.g. chlorophyll-a concentration. The user is then presented with a list of relevant data layers which match the concept identifier or URI (shown in Fig. 2c) and is able to select a data layer to visualise it (shown in Fig. 2d). In this instance, the user is presented with plain English results instead of the complex variable names which were in the original dataset. Future work will expand these search responses to provide more detailed contextual results to the end user, including information about the data provider, data types and other information which will help the user identify datasets of interest.

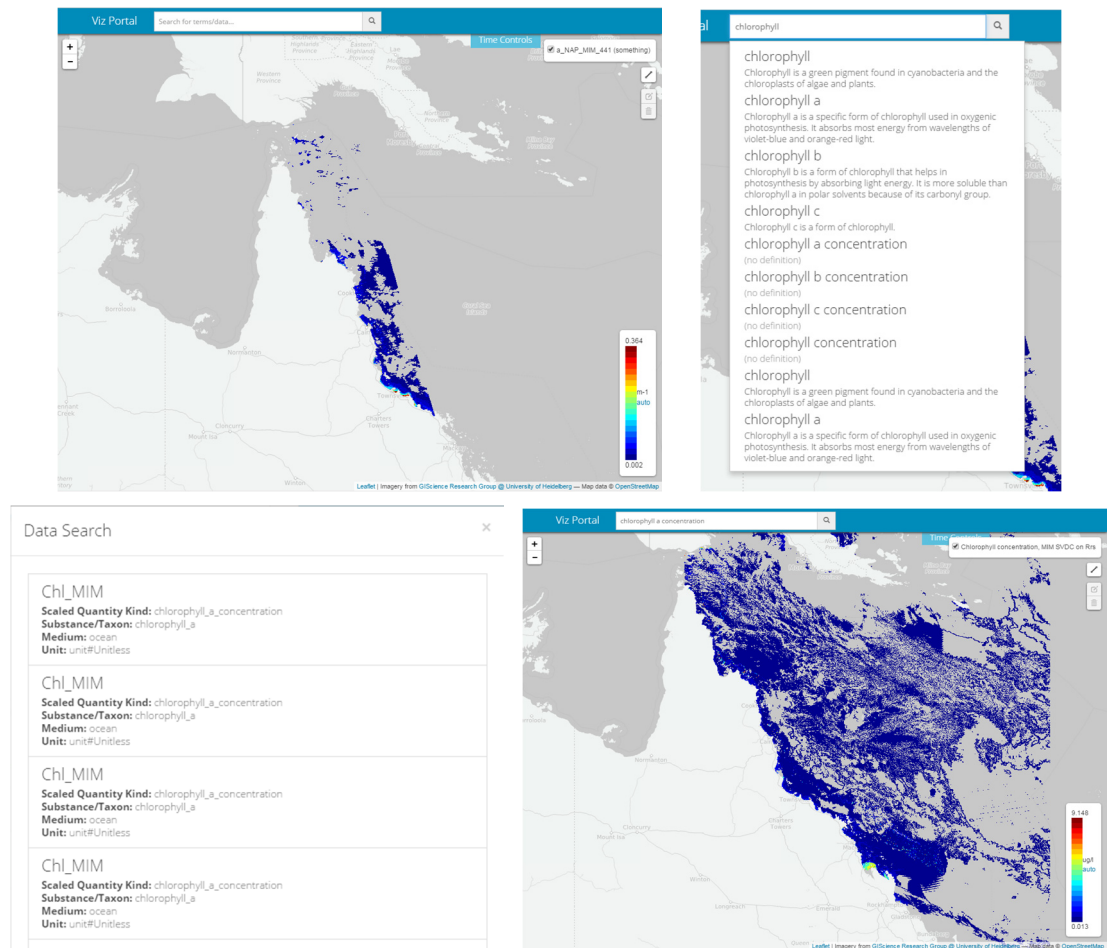


Figure 2. Clockwise from top to bottom - a) Visualisation Portal map-based user interface; and b) Search feature in the eReefs Visualisation Portal powered by eReefs SSISSVoc Vocabulary service; c) Data list showing matching datasets available for searched concept; d) Visualisation of Chlorophyll-a concentration from selected dataset.

The Visualisation Portal also allows end users to interact with datasets in multiple ways. A transect tool allows users to draw a line on the map interface to specify a transect across the spatial layer. This then displays the values for the data sampled along that transect. The figure below shows a user defining a multi-

point transect for Chlorophyll-a concentration and the data being displayed in a line chart. Again, this is achieved without the need to bind the chart component directly to a specific data service. The type of service is discovered via the Data Brokering Layer and the correct data service endpoint returned to the visualisation code. This allows the visualisation components to be built in highly reusable modules.

The modularity of the visualisation portal is further extended by utilising modern web technologies which combine HTML templating with data connected, and event-driven components. The eReefs Visualisation Portal is built using the Meteor (<http://www.meteor.com>) Javascript framework. The Meteor framework allows for rapid prototyping and deployment and helps to enable truly agile development.

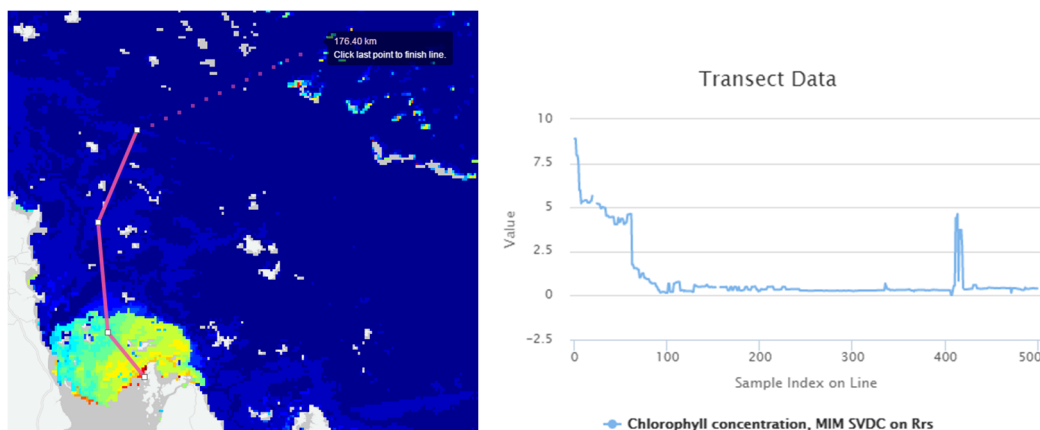


Figure 3. a) Defining a transect for Chlorophyll-a concentration; b) Data chart for defined transect.

3.2. eReefs Data explorer

While the Visualisation Portal provides a map-based view of the data and allows specific functions such as exploring data via transects, it doesn't allow users to explore data from a domain perspective. The eReefs Data explorer complements the Visualisation Portal and provides users with views from domain and data product perspective. Visualisations of eReefs domain concepts, including observable properties and water quality, have been developed to allow users to explore concepts and related data as well as explore the various related facets of domain concepts applicable to eReefs.

Figure 4 below shows a screenshot of the vocabulary explorer component which allows users to navigate the various controlled vocabulary concepts for eReefs. The main view shows the tree-based layout which lets users navigate between concepts and their narrower or child concepts in the concept hierarchy. The search bar on the top-right of the screenshot allows users to enter terms to search for particular concepts. The screenshot shows a view of the selected concept "Chlorophyll a concentration" rendered with concept highlighted in the tree structure. In the right panel, the concept details are shown to users as well as any matching related datasets known in eReefs. The information about the related datasets are queried dynamically from the eReefs DBL service allowing users to browse domain concepts and the datasets in a single view.

Figure 5 shows the drop-down panel which appears as the user enters a term in the query and suggests possible concepts and their related facets in the context of the eReefs Observable Properties conceptual model. The aim is to assist users to narrow their search to relevant terms with the facets. The details of the selected concept are provided to the user which is displayed in the bottom right panel and the relevant nodes are highlighted in the tree on the left panel as shown in Figure 4 for 'Chlorophyll a concentration'. It is important to note that the views shown below are being powered by the eReefs SISSVoc vocabulary service (<http://sisvoc.eereefs.info/>) which is maintained as a separate component to this client application.

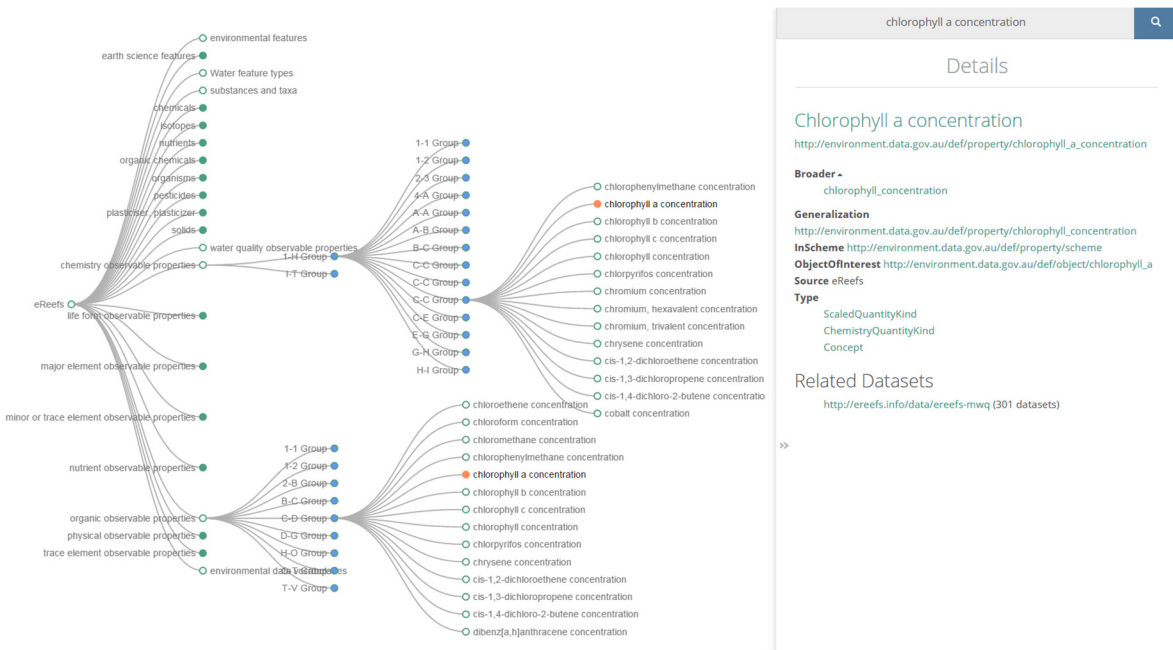


Figure 4. Visualisation of water quality domain concepts for eReefs powered by eReefs SISSVoc vocabulary services.

chlorophyll		
chlorophyll	substance or taxon	d
chlorophyll a	substance or taxon	
chlorophyll b	substance or taxon	
chlorophyll c	substance or taxon	
chlorophyll a concentration	quantity	
chlorophyll b concentration	quantity	
chlorophyll c concentration	quantity	
chlorophyll concentration	quantity	

Figure 5. Data Explorer domain search powered by eReefs SISSVoc vocabulary services. Dropdown list features groupings by facets from the Water Quality vocabularies.

4. DISCUSSION AND FUTURE WORK

In this paper, we presented the eReefs Visualisation Portal and the eReefs Data Explorer to support visualisation and data discovery respectively for the eReefs project. These client applications demonstrate how prior work on enabling semantic discovery of environmental data can be leveraged to flexibly reuse data assets and build client applications with different use cases.

The combination of the DBL and semantic descriptions, which identify datasets, data providers, data services and data service types and provide linkages between them, provides the advantages of being able to reuse the data infrastructure in a systematic way. We showed how the components in the architecture support systematic discovery, aggregation and visualisation of environmental data. The eReefs Visualisation Portal and the eReefs Data Explorer provide examples of leveraging the DBL to abstract complexity around access interfaces and syntax, and provide a semantically consistent view of the data across multiple DPNs.

Whereas in the past development of the client applications would bind to particular data services, the approach presented in this paper allows improved modularity and reuse of the components in the architecture. The DBL provides data discovery and access mechanisms for supporting the visualisation and data discovery modules without directly binding to the individual data services. This provides a flexible, reusable approach to both data provision and visualisation. The Data Provider Node abstraction provides a mechanism for defining both the data services, related semantic concepts and other metadata. This relatively minor addition results in significant improvement for client applications to discover, access and use datasets.

In this paper, we have introduced the eReefs Visualisation Portal, which provides a lightweight client application for visualising and exploring eReefs data via map views and novel visualisation techniques which leverage the underlying DBL for eReefs. Although, developed specifically for eReefs, the Visualisation Portal is able to be configured to retrieve data in broader contexts by connecting it to different DBL deployments. For example, if there was a DBL deployment for soils related data, then the Visualisation Portal would be able to render this data in a similar fashion. Compared with the Godiva web service, a commonly used visualisation tool for a single THREDDS data service, the Visualisation Portal is able to pull in data from a range of sources. This allows greater flexibility based on registered Data Provider Nodes in the DBL. A current limitation of the Visualisation Portal is the range of functions for users to filter search results. In future work, the authors intend to add features to filter the search results on criteria such as, time period selection, spatial extent selection, and data provider selections.

We also presented the eReefs Data Explorer which provides browse-based and search-based views of the data via graph visualisations, tree views and list views. This compliments the map-based visualisation of the Visualisation portal and allows users to explore the range of available data products and related components.

A limitation of the current implementations of the Visualisation Portal, Data Explorer and the DBL is that it has been more heavily tested on THREDDS data services. Future work is required to develop modules to handle other service interfaces such as Web Feature Services and Sensor Observation Services. This would extend the flexibility of the DBL and associated client applications to handle a wider range of data from other standardised web-based service interfaces and deployments.

ACKNOWLEDGMENTS

This research has been funded through the eReefs research project by CSIRO, the Science and Industry Endowment Fund (SIEF) and the Great Barrier Reef Foundation (GBRF). The authors acknowledge the contributions of Douglas Iacovelli, Yuan-Fang Li and Simon J.D. Cox for their input into the design and implementation of the eReefs vocabulary visualisation component.

REFERENCES

- Car, N.J. (2013). The eReefs Information Architecture, Proc. 20th Intl. Congress Modelling and Simulation, Adelaide, Australia, pp. 831–837.
- Car, N.J., Yu, J., et al. (2014). A services framework and support services for environmental information communities, Paper presented at 11th Intl. Conf. Hydroinformatics (HIC), Aug 17-21, New York, NY, USA, 8pp.
- Cox, S., Yu, J., & Rankine, T. (2014a). SISSVoc: A Linked Data API for access to SKOS vocabularies, Semant. Web., 01/2014, Pre-print:1-16, doi: 10.3233/SW-140166.
- Cox, S., Simons, B., and Yu, J. (2014b), A harmonized vocabulary for water quality, Paper presented at 11th Intl. Conf. Hydroinformatics (HIC), Aug 17-21, New York, NY, USA, 8pp.
- Cox, S. & Simons, B. (2014), WaterML-WQ – an O&M and WaterML 2.0 profile for water quality data, OGC Best Practice r14-003, Available <http://www.opengis.net/doc/BP/watermlwq/1.0> (Accessed 23 July 2015).
- Laxton, J., Serrano, J. J., & Tellez-Arenas, A. (2010). Geological applications using geospatial standards—an example from OneGeology-Europe and GeoSciML. International Journal of Digital Earth, 3(S1), 31-49.
- Levitus, S., Antonov, et al. (2013). The World Ocean Database. Data Science Journal, 12(0), WDS229-WDS234.
- Simons, B.A., Yu, J., Cox, S.J.D. (2013a). Defining a water quality vocabulary using QUDT and ChEBI, Paper presented at 20th Intl. Congress on Modelling and Simulation. pp. 2548–2554.
- Simons, B.A., Yu, J., Cox, S.J.D. (2013b). Water quality vocabulary development and deployment, Paper presented at American Geophysical Union Fall Meeting, Abstract IN53D-1586. San Francisco, USA.
- Yu, J., Simons, B.A., et al. (2014). Enhancing water quality data service discovery and access using standard vocabularies, Paper presented at 11th Intl. Conf. Hydroinformatics (HIC), Aug 17-21, New York, NY, USA, 8pp.
- Yu, J., Car, N.J., et al. (2015a). Towards Linked Data Conventions for Delivery of Environmental Data Using netCDF. Environ. Softw. Syst. Infrastructures, Serv. Appl. 448, 102–112. doi:10.1007/978-3-319-15994-2_9.
- Yu, J., Leighton, B., et al. (2015b). The eReefs Data Brokering Layer for hydrological and environmental data, Journal of Hydroinformatics, p. 1-23, in review.