

# EVOLUTION IN DATA AND PRODUCT MANAGEMENT FOR SERVING OPERATIONAL OCEANOGRAPHY, A GODAE FEEDBACK

F. Blanc<sup>(1)</sup>, V. Baralle<sup>(2)</sup>, J.D. Blower<sup>(3)</sup>, E. Bronner<sup>(4)</sup>, P. Cornillon<sup>(5)</sup>, J. deLaBeaujardiere<sup>(6)</sup>, C. Donlon<sup>(7)</sup>,  
A. Gemmel<sup>(3)</sup>, S.C. Hankin<sup>(8)</sup>, R. Keeley<sup>(9)</sup>, O. Lauret<sup>(1)</sup>, T. Loubrieu<sup>(10)</sup>, I. Petiteville<sup>(11)</sup>, S. Pouliquen<sup>(10)</sup>,  
M. Price<sup>(12)</sup>, T. Pugh<sup>(13)</sup>, A. Srinivasan<sup>(14)</sup>

- <sup>(1)</sup> Space Oceanography Division, CLS, Ramonville-St-Agne, France, Email: [fblanc@cls.fr](mailto:fblanc@cls.fr); [olauret@cls.fr](mailto:olauret@cls.fr)  
<sup>(2)</sup> JRC, Ispra, Italy, Email: [vittorio.barale@jrc.it](mailto:vittorio.barale@jrc.it)  
<sup>(3)</sup> Environmental Systems Science Centre, University of Reading, Reading, UK,  
Email: [j.d.blower@reading.ac.uk](mailto:j.d.blower@reading.ac.uk); [alg@mail.nerc-essc.ac.uk](mailto:alg@mail.nerc-essc.ac.uk)  
<sup>(4)</sup> Satellite Altimetry and Localisation Programme, CNES, Toulouse, France, Email: [emilie.bronner@cnes.fr](mailto:emilie.bronner@cnes.fr)  
<sup>(5)</sup> Graduate School of Oceanography, University of Rhode Island, Narragansett, RI, USA,  
Email: [pcornillon@gso.uri.edu](mailto:pcornillon@gso.uri.edu)  
<sup>(6)</sup> Program Office National Oceanic and Atmospheric Administration, Silver Spring, MD, USA,  
Email: [Jeff.deLaBeaujardiere@noaa.gov](mailto:Jeff.deLaBeaujardiere@noaa.gov)  
<sup>(7)</sup> ESA/ESTEC, Noordwijk, Netherlands, Email: [craig.donlon@esa.int](mailto:craig.donlon@esa.int)  
<sup>(8)</sup> Pacific Marine Environmental Laboratory, NOAA, Seattle, WA, USA, Email: [Steven.C.Hankin@noaa.gov](mailto:Steven.C.Hankin@noaa.gov)  
<sup>(9)</sup> DFO-MPO, Ottawa, Ontario, Canada, Email: [KeeleyR@DFO-MPO.GC.CA](mailto:KeeleyR@DFO-MPO.GC.CA)  
<sup>(10)</sup> IFREMER, Plouzané, France, Email: [Thomas.Loubrieu@ifremer.fr](mailto:Thomas.Loubrieu@ifremer.fr); [sylvie.pouliquen@ifremer.fr](mailto:sylvie.pouliquen@ifremer.fr)  
<sup>(11)</sup> ESA/ESRIN, Frascati, Italy, Email: [Ivan.Petiteville@esa.int](mailto:Ivan.Petiteville@esa.int)  
<sup>(12)</sup> Met Office, Exeter, United Kingdom, Email: [martin.price@metoffice.gov.uk](mailto:martin.price@metoffice.gov.uk)  
<sup>(13)</sup> Bureau of Meteorology Research Centre, Melbourne, Australia, Email: [T.Pugh@bom.gov.au](mailto:T.Pugh@bom.gov.au)  
<sup>(14)</sup> Rosenstiel School, Miami, FL, USA, Email: [asrinivasan@rsmas.miami.edu](mailto:asrinivasan@rsmas.miami.edu)

## ABSTRACT

The GODAE (Global Ocean Data Assimilation Experiment) project has operated for a decade to establish a common definition and description of what is a GODAE ocean data product, and tuned each production center to routinely serve quality-controlled ocean data. A legacy of GODAE has been in harmonizing the various products and the development of essential and generic functions to connect users to products tackling the problem of both the diversity of ocean data and volume of data flows.

The European directive 'Infrastructure for Spatial Information in Europe (INSPIRE) and the IOOS Data Management And Communication plan issues (DMAC) have defined the (distributed) architecture to serve data geospatially referenced. Since 2007, the two of them enter the implementation phase, addressing 3 technical issues to solve the problem of interoperability between software applications, across different organization: the metadata, the ontology, and the service bus. Pursuing their quest, the GODAE community is setting up large size structuring / implementation projects or programs in Europe. Those projects work on previously-addressed issues, with the goals of sharing expertise, reducing production costs, and producing high-quality, well-described, peer-reviewed datasets. Data sets are referenced in a central system which is developing an interoperable interface to discover, access and view them. The near future will also see a strong move

towards “operationalization” of the various data systems and the management of the timelines/accessibility criteria. The main challenge, however, for the future success of operational oceanography, is dependent upon the community to work together, to maintain a network of experts and to agree upon common approaches. The community therefore expresses recommendations or priority actions on how they envisioned the next steps in harmonization/standardization to help forge and tied up the links as well as to go on the convergence of solution and tools for efficient and sustainable marine core services.

## 1. INTRODUCING GODAE AND WHAT WAS ACHIEVED

The GODAE (Global Ocean Data Assimilation Experiment) project has operated for a decade to establish a common definition and description of what is a GODAE ocean data product, and tuned each production center to routinely serve quality-controlled ocean data.

**GODAE ocean data products** fall into one of two groups, earth observations and monitoring/forecasting information [Ref.5 and Fig.1].

Earth observations referred to science data acquired by remote sensing satellites (global and dense coverage, sea surface observations) or from in-situ instruments, (global and sparse coverage, 3D information). They are not raw engineering data or image products, but ocean parameters representative of a physical or biogeochemical phenomenon (e.g. sea level, currents, temperature, salinity, wind/wave, sea ice). They are characterized by their timeliness, which is the time interval between current time and product time, including time delay for accessibility of a delivery method (time span period).

Numerical simulations, called ocean modeling systems, routinely assimilate earth observations to derive monitoring and forecasting information of the ocean. They are targeted to open oceans, global ocean or regional sub-basins at higher resolution (gridded coverage, from ocean surface to depth). They are defined according to their fitness for purpose, spatial and temporal scales of exhibited ocean phenomena. For example, is a model suitable for simulating large ocean circulation or is it eddy-resolving? Is it suitable for daily or seasonal monitoring? Earth observations are also used to routinely validate the monitoring and forecasting information.

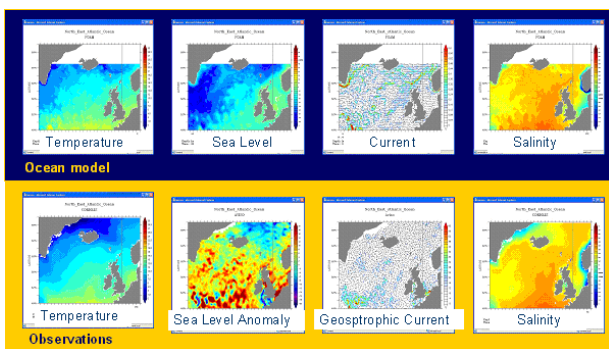


Figure 1. GODAE ocean data products<sup>1</sup>

A legacy of GODAE has been in harmonizing the various products and the development of **essential and generic functions to connect users to products** [Ref.1 and Ref.7]. The GODAE data service focused on 2 user needs:

- 1- Product discovery and understanding for the general public,

<sup>1</sup> GODAE ocean data products are geo and time-dependent (X,Y,Z,t), and ocean parameters in a data set collocated. The geographical coverage can refer to a point, a path along an observational swath, or a gridded area. The vertical coverage can refer to the surface level or address several ocean depths. The temporal coverage can address a precise time or a time window.

- 2- Data serving for regular or non regular users tackling the problem of both the diversity of ocean data and volume of data flows.

Almost each production centre implemented a „demonstration’ portal on the web for any user to discover and browse their products portfolios, with links to download functions. This has led to the harmonisation of the description of products (essential metadata and content as presented in previous section) and revealed the need to browse the data simultaneously to help the user better understand or evaluate/assess the products for his or her application before acquiring them [Fig. 2].

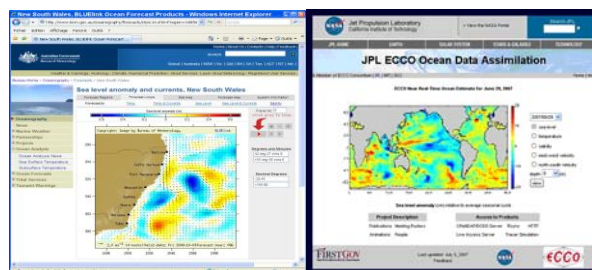


Figure 2. Product discovery functions implemented for BLUElink forecasts or ECCO reanalyses

Once the user knows which data products are required, he or she usually needs an efficient method for accessing them. The ground technology to publish the data depends on the product characteristics and user needs. For example, the meteorological community uses the European or Global Telecommunication System (Eumetcast, GEONETcast) to acquire in situ and satellite data which have been served using a secured FTP, whereas general users require tools and standards to find, discover, acquire and use the data (cf. the Google Earth fashion).

The Thredds/OPeNDAP technology (Unidata framework) has been widely and successfully used by the GODAE community to connect users to data products, allowing powerful and flexible access to gridded data. The Opendap has to be associated with:

- the NetCDF file format interface,
- the Climate and Forecast (CF) metadata conventions.

Use of this trio [Ref.2,15] allows to handle time series of gridded data to be handled. The hindcast/forecast window of model outputs can be managed to ensure that the forecasts are updated and past forecasts replaced by the best available data, such as a nowcast (current state) or hindcast (past state). Datasets are aggregated and presented as one product and, users to access/transport only the precise data they need (selection of parameters

and spatio-temporal window). The new product created is called "best-estimate" time series and characterised by its timeliness. A large number of analysis and display web-based systems or desktop toolboxes are compatible with this technology and widely used by the science community [Ref.5].

Recent advances in interactive data viewing functions have pushed the development of new web mapping service usable by the GIS community. This web mapping service is now mature for gridded products (ncWMS) and has been included in the latest version (4) of the Thredds/OPeNDAP Data Server. Research is ongoing to apply this technology to other data products such as in situ observations and satellite swaths, and to handle non-map products like time series, cross sections, vertical profiles, etc.

**Important applications** of GODAE technology exist already within the different national programmes of GODAE partners (e.g. marine safety in Canada to support the coast guard, fisheries in Australia to survey migration of the larvae of rock lobsters to maintain populations, maritime traffic in Japan to predict the Kuroshio large meander, marine pollution in France to monitor oil spill drift and to facilitate protective measures, water quality in United Kingdom to monitor the risk of problems from algae bloom, etc; [Fig.3 and Ref.5]. Strong contact and interaction is maintained by each GODAE partner for:

- Assessing data usefulness and user satisfaction, products to users,
- Enhancing the end-to-end data management to reach a high degree of service to user, i.e. critical needs for users (service commitment formalised in a Service Level Agreement, cf. ITIL process),
- Developing and testing interoperability of the systems either to acquire data, to enhance the visibility of products or use a typical distribution network.

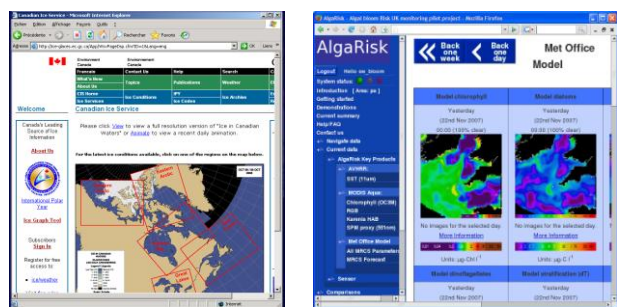


Figure 3. The Canadian Ice service and an application of the NorthWest Shelf Ecosystem (NCOF)

## 2. WHAT LESSONS WERE LEARNED DURING GODAE LIFETIME AND PLANNED EVOLUTIONS

- GODAE acted as a maritime cluster, **a learning place to increase in capacities**, sharing technologies and work practices and developing new capabilities for data management. The challenges of these activities have been addressed through the cooperative skills and energies of many individuals, the abilities to work with distributed systems and use of advanced technologies that were not tightly constrained by international standards. The challenges now are to solidify the procedures into standards as well as to operationally manage the data products (service engagement).
- Delivery of information via the Web has proved to be successful, but the large number of diverse web portals in current use can be confusing to users, particularly those outside the ocean community. End-to-end data management is almost less achievable by individual groups and therefore we need to build **integrated or semi-distributed infrastructure** to share effort, to build a common view of operational oceanography and to meet new goals for data serving.
- The European directive 'Infrastructure for Spatial Information in Europe (INSPIRE, see <http://inspire.jrc.ec.europa.eu/>) and the IOOS Data Management And Communication plan issues (DMAC, see <http://www.ocean.us/dmac subsystem>) have defined the (distributed) architecture to serve data geospatially referenced. Since 2007, the two of them enter the implementation phase, addressing 3 technical issues to solve the problem of interoperability between software applications across different organization, the metadata<sup>2</sup>, the ontologies<sup>3</sup>, and the service bus<sup>4</sup>.

<sup>2</sup> All the information required about the met-ocean product is called **meta-data** and ruled by the ISO/TC211 norms (International Standard Organisation), but mainly through use of the ISO 19115 and ISO 19119 metadata for discovery and service, and the XML representation by ISO 19139. The general mechanism is to create a schema from a general description common to all data products (GML/ISO19136 application schema) to specific information. Each level of specificity is an extension of the previous one. Final data model may also meet other needs like the knowledge of product dependencies or a way to organise the product among others and define back up or alternative solutions for users. Then XML metadata are instantiated for each node of the catalog.



Ontologies [Fig.4] improve the discovery process of resources by the common user viewpoint by describing, in machine-readable form, the key concepts used by the ocean community and the relationships between them. Definition and implementation of ontologies relevant to GODAE ocean data products is worked out through the following projects of communities: HMA (GSCDA), Climate/Forecast (CF), SEADATANET / Marine Metadata Interoperability (MMI), Global Change Master Directory (GMCD).



Figure 4. Ontologies items for Earth System Sciences (source: NASA, SWEET)

- Pursuing their quest, the GODAE community is setting up large size structuring projects or programs in Europe, Australia, U.S. & Asia [Tab.1 and Ref. 3,4,6,8,9,10].

Those projects work on previously-addressed issues, with the goals of sharing expertise, reducing production costs and producing high-quality, well-described, peer-reviewed datasets. Datasets are referenced in a central system which is developing

an interoperable interface to discover, access and view the datasets.

MERSEA / MYOCEAN	GMES applications – Marine Core Service	Europe
NOAA/IOOS DIF	Coast Watch – data access	US
IMOS	Marine Observing System – data access	Australia
US GODAE	GODAE data server	International
HMA	Satellite Earth Observations – data access	Europe
GHRSS-PP	sea surface temperature - data access	International
SEADATANET	In situ ocean observations – data access	Europe

Table 1. Present-day referenced large size structuring projects

The central system of this infrastructure deploys functions to:

- Manage the interfaces with the production centres with the aim of maintaining a full picture of products, (up to date and valid information on the data; delivery specification; traceability of events).
- Implement the standard interface for discovery of data products and web services of the service bus (all the connection between products and users, be they applications, web portals or services management).
- Specify also Rights Management Services (RM), which are needed to invoke e-commerce services.

3 **Ontologies** allow the exchange of data to take place not only at a syntactic level, but also at a semantic level, providing a shared understanding of common domains. Ontologies are handled under the naming of Semantic Web, an extension of the World Wide Web. The encoding is ruled with the following standards: W3C OWL, SKOS, RDF.

4 The **service bus** should deliver a number of web services; all ruled by OGC standards (Open Geospatial Consortium): CSW for discovery, WMS for viewing, WCS for downloading, and WPS for transformation.

- Many GODAE technologies have proven now their value, they are mature to serve gridded data, and ready for standardization: **Thredds/OpenDap/ncWMS is a candidate today for OGC** for download and visualization. Besides this first priority need, this suite of tools is now faced to economical issues or the need to maintain the tools, educate on their added value, configuration and use. These services are often overlain with other services that handle security issues and notification or alert needs.

### 3. GUIDANCE TO THE OCEAN COMMUNITY OR CHALLENGES FOR EFFICIENCY AND SUSTAINABILITY

The near future will thus see a strong move towards “operationalization” of the various data systems, with a continuing need to serve the ultimate goal of real user applications and commercial services. However, the foreseen evolutions are not without **threats**:

- The community is in mutation and the number of skills needed for end-to-end data management increased. The coordination was based on volunteer ship and conducted from small size project to medium size project. This collaborative mode is no more valid. How to pursue the GODAE momentum and further harmonize the data and interfaces? How to identify and tie up large size projects together?
- Future evolution in interface technology can be driven by pressures from individual projects, and not thought as a whole in system of systems. How to work on the new needs to define/standardize grades in data serving to allow each data management system to increase in capacities or benefit of functions of an external structure to increase in capacity?

The community therefore expresses 5 recommendations or priority actions on how they envisioned the next steps in harmonization/standardization to help forge and tie up the links as well as to go on the convergence of solution and tools for efficient and sustainable marine core services:

#### 1) **Edit a handbook for operational ocean forecasting data management**

The handbook will be a guide for data manager. It should document common nomenclature and symbology<sup>5</sup> as well as interfaces requirement<sup>6</sup> and best practices for the provision of marine core service.

#### 2) **Define monitoring needs to benchmark the performance across the system**

The monitoring is a tool for data manager to control and assess the system usefulness. It should give synthesis knowledge of the ocean data product:

---

<sup>5</sup> Cf. definition of products, critical needs for users, data management and service use cases – top level ocean scenarios -, data type and other criteria, discovery and service metadata, ontologies and conventions.

<sup>6</sup> Cf. upstream data provider and downstream service applications, standard and implementation tools.

type, format, access means and added value web services, data policy, storage, volume capacities, time window, period of validity, archival systems, network availability, timeliness, accessibility, uses and other key metrics for services commitments). It is also the tool to track the progress in implementing standard interfaces to acquire the data or to anticipate hardware, software and applications evolutions to operate, maintain and enhance the interfaces to data.

#### 3) **Define elements to retrieve feedback on end-uses from top to bottom, from users to producers.**

The full operational oceanography service can be met because of integration of many components. But with distributed systems, producers suffer nowadays of a lack of visibility on their utility and do not retrieve anymore user feedback. These feedback on system use and performance of products and services of each element is however a key issue for long term sustainability as well as give the knowledge of further user requirements, and priority to be given to future developments. The definitions should cover the elements (be it e-mail exchanges between services and users or transaction accounting messages) to be shared between systems.

#### 4) **Demonstrate interoperability of catalogue interfaces and how each system benefits the others**

This demonstration of interoperability is raised by the need to aggregate information, i.e. to provide users with a single point of access (one stop shopping portal) to search, query and discover any product. The demonstration of interoperability should address various thematic systems (e.g. operational oceanography data, meteorological data, climate data, coastal information, hosted anywhere in the world) but also the link with the GIS (Geographical Information Systems) community for the joint use of various types of products to enable spatial planning and development of maritime clusters.

#### 5) **Prioritize needs for further developments**

The final need is to provide guidance to improve interoperability; this is a high priority for research and development. For instance, a consistent data format is required for in situ data to allow the sharing of these data via OPeNDAP/THREDDS technology. Another example is the need for technology for managing the relationships between products (Dorandeu et al. 2009).

## 4. CONCLUSION & FUTURE OUTLOOK

Changes in technology and changes in society in the last ten years are both forcing production centers to rethink their role and modus operandi of data management, having well defined products and services to be useful to wider, non-scientific or multidisciplinary applications, as well as to realise the potential social and economic benefits they offer.

End-to-end data management must now be considered as an integral part of the design of a production/distribution system, and outlines the increasingly critical role played by ocean data stewardship in ensuring that ocean products are visible, manageable, accessible and utilized.

The main challenge, however, for the future success of operational oceanography, is dependent upon the community to work together, to maintain a network of experts and to agree upon common approaches.

As seen in this community paper, concrete implementation is done through large integrated projects or programmes. The first need is thus to tie up projects together, creating an official communication channel, and pursue the work on harmonizing interfaces and developing interoperability of systems. This need is partially met today:

1. Some projects identified external expert or adviser from other projects (cf. MyOcean & Seadatanet European projects requesting for IOOS expertise to revise the infrastructure to be deployed).
2. Some projects included tasks to implement interoperability (at a certain level) with other systems (cf. MyOcean with the Weather Information System – WIS – or GEO Portal – HMA –).
3. Participation and exchanges thanks to today major conferences which now clearly identify such themes in their programs (cf. OceanObs2009, PV2009).

Actions identified in previous section are done today at project level but not coordinated above. For further communication and wide adhesion, the community can use the JCOMM web site to publish online documentation as draft document and made available for public comment.

## 5. REFERENCES

- [1] Blower et al. (2009), "Serving GODAE data and products to the ocean community", *JOO*.
- [2] Hankin et al. (2009), "NetCDF-CF-OPeNDAP: Standards for Ocean Data Interoperability and Object Lessons for Community Data Standards Processes", *ESA publication*, OceanObs CWP.
- [3] de La Beaujardière et al. (2009), "Ocean and Coastal Data Management", *ESA publication*, OceanObs CWP.
- [4] Dorandeu et al. (2009), "MyOcean Concept Document", – synthesis for public dissemination –, *MyOcean*, project public deliverable.
- [5] Blanc et al. (2008), "Data and product serving, an overview of capabilities developed in 10 years", *GODAE*, conference proceedings.
- [6] Blanc F. (2008), "The relationship between SALP/AVISO and MyOcean, information systems view", *Eumetsat*, conference proceedings.
- [7] Blower et al. (2008), "Underpinning technologies for oceanographic data sharing, visualization and analysis: Review and future outlook", *GODAE*, conference proceedings.
- [8] Blanc F. (2008), "MyOcean Information System", *Eurogoos*, conference proceedings.
- [9] de La Beaujardiere (2008), "The NOAA IOOS Data Integration Framework: Initial Implementation Report," *Proc. MTS/IEEE Oceans '08*.
- [10] Loubrieu et al. (2008), "A distributed system for operational oceanography: the MERSEA System", *JOO*.
- [11] Blanc et al. (2008), "A Remote Sensing Ocean Portal for GMES Ocean", *Mersea*, *JOO*.
- [12] Blanc et al. (2004), "Mersea GODAE Model Intercomparison Framework", *GODAE*, conference proceedings.
- [13] Blanc et al. (2004), "Mersea Live Access Server", *Ocean Sciences*, conference proceedings.
- [14] Hankin et al. (2004), "Data and Communications Infrastructure for the U.S. Integrated Ocean Observing System", *ESDIM*.
- [15] Cornillon et al. (2003), "OPeNDAP: Accessing Data in a Distributed, Heterogeneous Environment," *Data Science Journal*.