Aviso and Mercator serving MyOcean, or how a European marine user accesses ocean products derived from satellite altimetry and numerical simulations (past and present missions, and forecasts)

Frédérique Blanc (1), Dominique Heulet (2)

*(1) **Collecte Localisation Satellite (CLS)***
8-10, rue Hermès, Parc Technologique du Canal, 31520 Ramonville Saint-Agne, France
EMail: fblanc@cls.fr

*(2) **Centre National d'Etudes Spatiales (CNES)***
18, Avenue Edouard Belin, 31401 Toulouse Cedex 9, France
EMail: dominique.heulet@cnes.fr

1- Abstract

MyOcean (http://www.myocean.eu.org) is a 3-year project granted by the European commission within the GMES Program (7th Framework Program) whose objective is to provide a common denominator data for all users requesting an information on the ocean: the ocean information for downstream services. The paper is organised in four sections. The first section further describes MyOcean and its commitments. The second section describes the CNES/SIPAD-NG to access historical archival data. The third section describes the CLS/ATOLL to access timely ocean data time-series. The fourth section clarifies the concept of constellation of catalogues, showing how each catalogue remains in its own scope and enlarges its content by showing/delivering products of other catalogues (long term preservation versus operational oceanography). The clarification addresses also the gaps to fulfill implementation, be it either technical or organizational.

2- Presentation of MyOcean

MyOcean (http://www.myocean.eu.org) is a 3-year project granted by the European commission within the GMES Program (7th Framework Program) whose objective is to provide a common denominator data for all users requesting an information on the ocean: **the ocean information for downstream services.** This common denominator data describes the ocean currents, temperature, salinity, sea level, primary ecosystems, ice coverage (*(see figure 1)*. It is delivered thanks to added value functions which are for instance, data combination and assimilative models, continuity and consistency, expertise on the ocean, **information system, service desk**.

By 2012, the ocean service should be operational, based on **a « service oriented » organization[1]** to bring up a new pan-european value for ocean monitoring and forecasting for the benefit of the marine service providers on duty at national or European levels. This ocean service should also get aboard the link for **continuous innovation and improvements and the feedback for sustainabilit**y.

[1] Ruled by an ITIL management process.

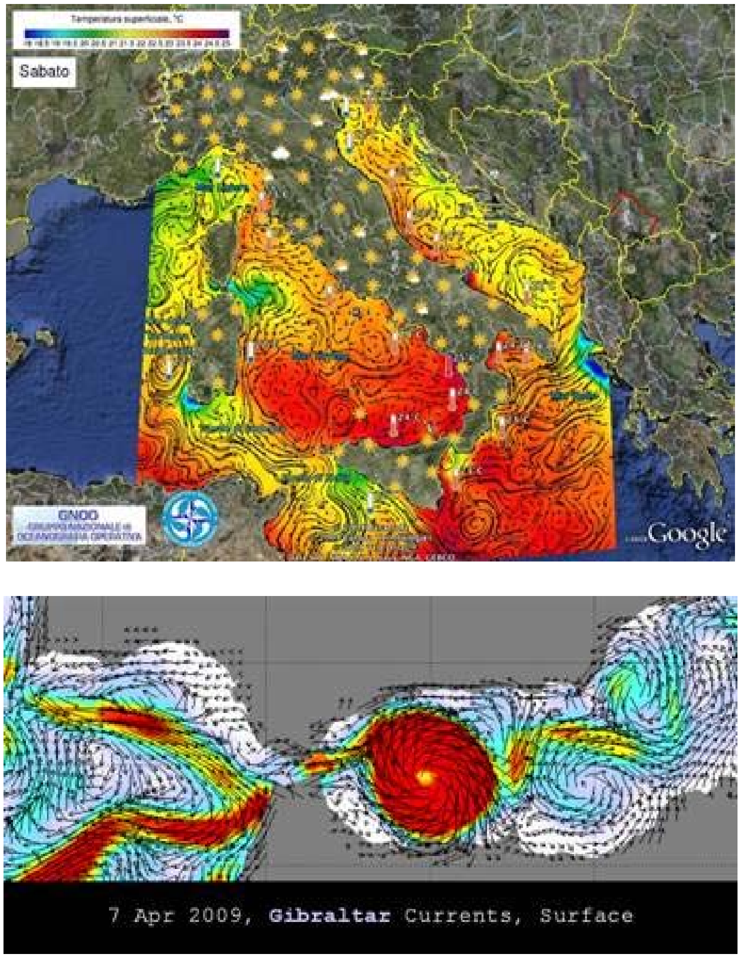


Figure 1: sea surface and currents in the Mediterranean Sea (© MyOcean)

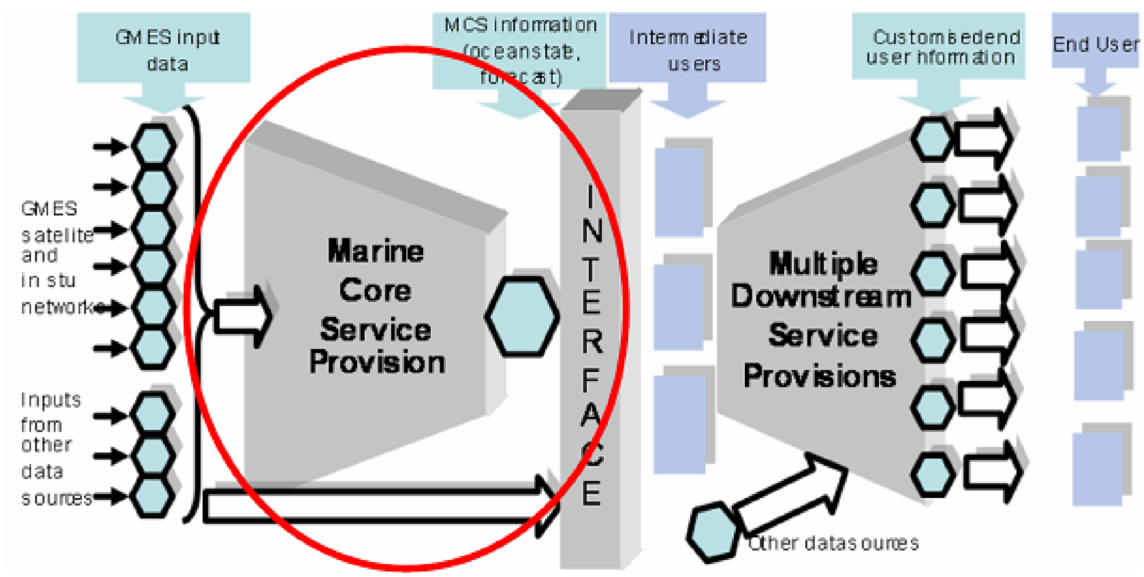


Figure 2: MyOcean clear (and limited) mission and roles in the value chain linking observations to end-users.

MyOcean has thus **a clear (and limited) mission and roles in the value chain** *(see figure 2)*, establishing the link between upstream activities (satellite and in situ observation networks – measurements –) and downstream activities through targeted users (specialized service providers). Downstream activities have been identified according to 4 market segments with key users in the project to establish **permanent link with the end-user community**, including requirements capture and system assessment: 1 – Marine Safety (marine operations, oil spill combat, ship routing, defense, search & rescue), 2 – Marine Resources (fish, stock management, ICES, FAO), 3 – Marine and Coastal Environment (water quality, pollution, coastal activities), 4 – Climate & Seasonal Forecasting (climate monitoring, ice, seasonal forecasting). **Access to products is open and free**, and delivery commitments managed through Service Level agreement (SLA) between users and myOcean.

MyOcean organization is made of production centers and shared facilities. The production centers are categorized as follows:

- 5 Thematic Assembly Centers (TACs) to provide past and current ocean information derived from satellite observations – sea level, ocean color, sea surface temperature, sea ice and wind – and in situ observation network.
- 7 Monitoring and Forecasting Centers (MFCs) to provide ocean analyses and forecast information fitted to the global ocean and European basins, Arctic, Baltic, Northwest shelves, Iberian / Biscay / Ireland (Ibiroos), Mediterranean sea, Black sea.

TACs and MFCs *(see figure 3)* **are engaged in the system through Operating Level Agreement (OLA)**, to regularly and systematically deliver qualified products identified in the products and services portfolio. These engagements will define level of qualification, access and monitoring tools.

Figure 3: MyOcean production centers in Europe.

Mercator Océan, **CLS** and **MetOffice** are among the major partners of myOcean:

- **Mercator Océan** lead the overall project, a system of system, drives the strategy, governance and data policy, is responsible of user engagements and finally responsible of the global and IBIROOS MFCs.

- **CLS** is in charge of system engineering and configuration management for development, responsible of external and internal interfaces (OLA implementation), ensures quality management and more specifically product quality for TACs, responsible of myOcean Information System (MIS) [2], fitness for purpose and operational performance and the sea level TAC, and contributes to the global MFC.

- **Met Office** is responsible for overall product quality relying on a transverse expert group to meet the needs, responsible for operations, technical issues management and evolutions, responsible of the central service desk and finally responsible of the NorthWest Shelves MFC.

From now on, the paper will focus on a user cases requesting sea level products, and how upstream and core systems are linked together. The link with the upstream systems is discussed here via the availability/accessibility of **historical and real-time along track Corrected Sea Surface Heights**. Satellite altimeter measurements are key inputs of the Sea Level TAC which transforms them into inter-calibrated high resolution[3] sea level anomalies directly usable by **MFCs** like the global system of Mercator Océan (either along track data for assimilation or internal gridded products for routine validation) and **end-users** (timely and accessible high resolution along track products).

The management of the altimetry data products is met by **SALP/AVISO**, which relays on the interoperability of two technology solutions which complete quite well each others, the **CNES/SIPAD-NG** archival systems and the **CLS/ATOLL** timely delivery systems *(see figure 4)*. The first one is describe in section 3, the second one in section4. Finally the paper will conclude on the advantages of the two systems and what is still missing to go further away in meeting user requirements.

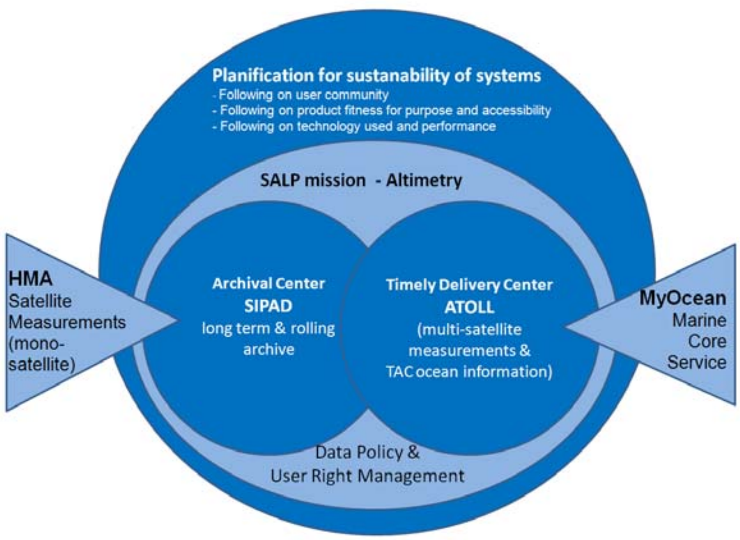


Figure 4: Constellation of information systems – Altimetry user case

3 - Presentation of SIPAD-NG to access historical archival data

SIPAD-NG (Information System for Data Preservation and Access – New Generation) is a generic software system allowing web consultation of scientific data catalogs and access to data. *Sipad-NG* is

[2] Implementation is shared with Ifremer, who is also responsible for the In Situ TAC.

[3] Targeted to each MyOcean regions, the global ocean and European sea regions.

adaptable to any scientific domain and, therefore, can be used by **Data Centres** to implement their data access function without having to develop it from scratch.

The aim of *Sipad-NG* is to avoid recurrent development of data access systems in Mission Centres or Data Centres financed by CNES. In particular, *Sipad-NG* is intended to be used by thematic Data Centres whose long-term mission is to regularly enrich the list of data and added value services made available to a user community. *Sipad-NG* is thus an efficient, open-ended operational system with sophisticated administration functions.

Functions provided by *Sipad-NG* are: acquiring and cataloguing data, searching for and selecting data, consulting information on data, managing users and user profiles, extracting archived data and distributing data through networks or particular media. To be adapted to a thematic Data Centre, *Sipad-NG* uses a Data Dictionary technology that allows adapting the information model to match the characteristics of a particular scientific theme and to define and control the interface for supplying information to the system.

The functional architecture of *Sipad-NG* matches the **OAIS functional model** (Open Archival Information System), *as shown on figure 5*. *Sipad-NG* is an implementation of the **ingest function**, **data management function** and **access function** of the OAIS model, adapted to the needs of space Data Centres. At CNES, the OAIS **storage function** is implemented by a dedicated service: the **STAF** (File Transfert and Archiving Service). Thus, *Sipad-NG* and STAF are an implementation of the OAIS model.

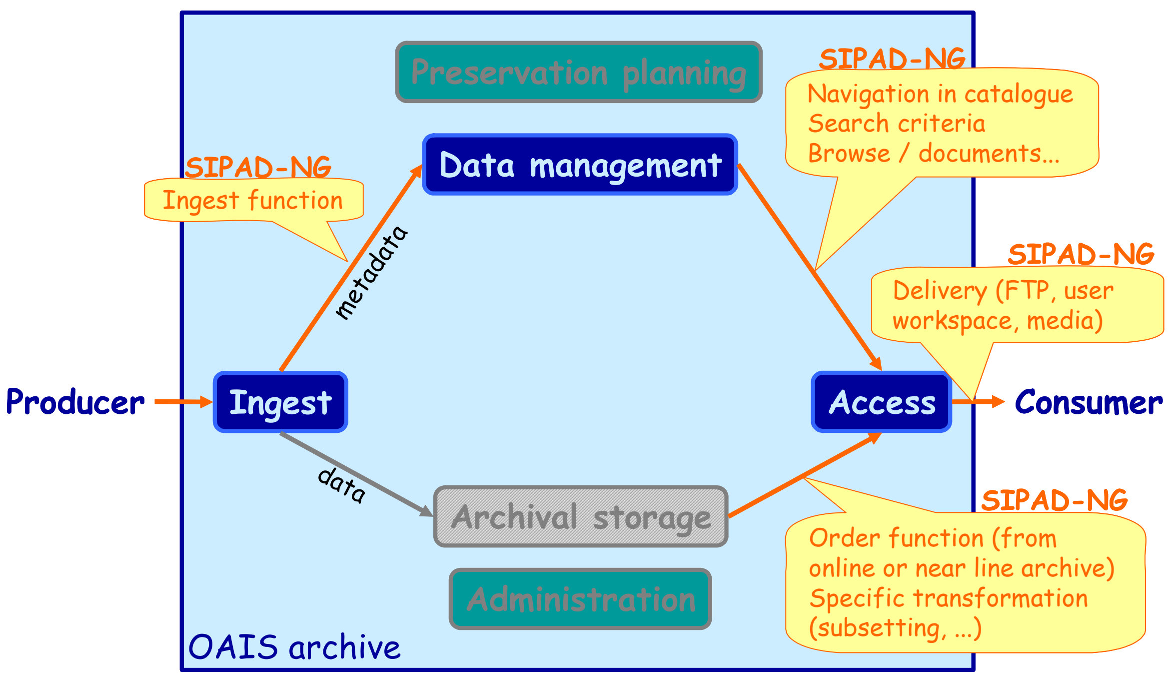


Figure 5: OAIS functions implemented by Sipad-NG to meet Data Centres requirements

An important element allowing large capabilities for adapting *Sipad-NG* to a Data Centre is the **architectural design of the system**. *Sipad-NG* is very modular *(see figure 6)*. A **relational database** and a set of applications called **"basic services"** constitute the kernel of the system. The "basic services" can be interfaced by **client applications** that can be web servers or any kind of local or distant application having to consult the catalog or to retrieve data from the archive. Figure 5 shows this very modular architecture.

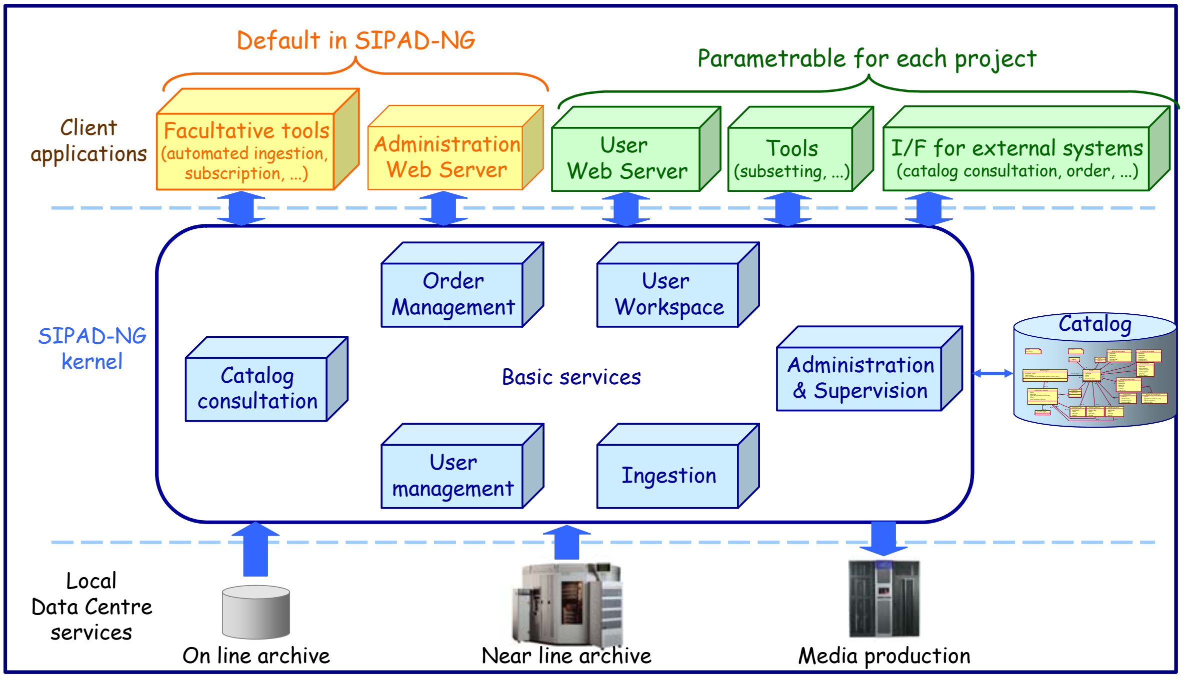


Figure 6: Sipad-NG architecture – client applications interface "basic services" that share information trough a relational database

Thus, *Sipad-NG* offers two ways for accessing Data Centres: **direct access** through the user Web Interface or **access through web services**, allowing the constitution of **"catalog of catalogs"** accessing several archives based on *Sipad-NG*, *as shown by figure 7*.

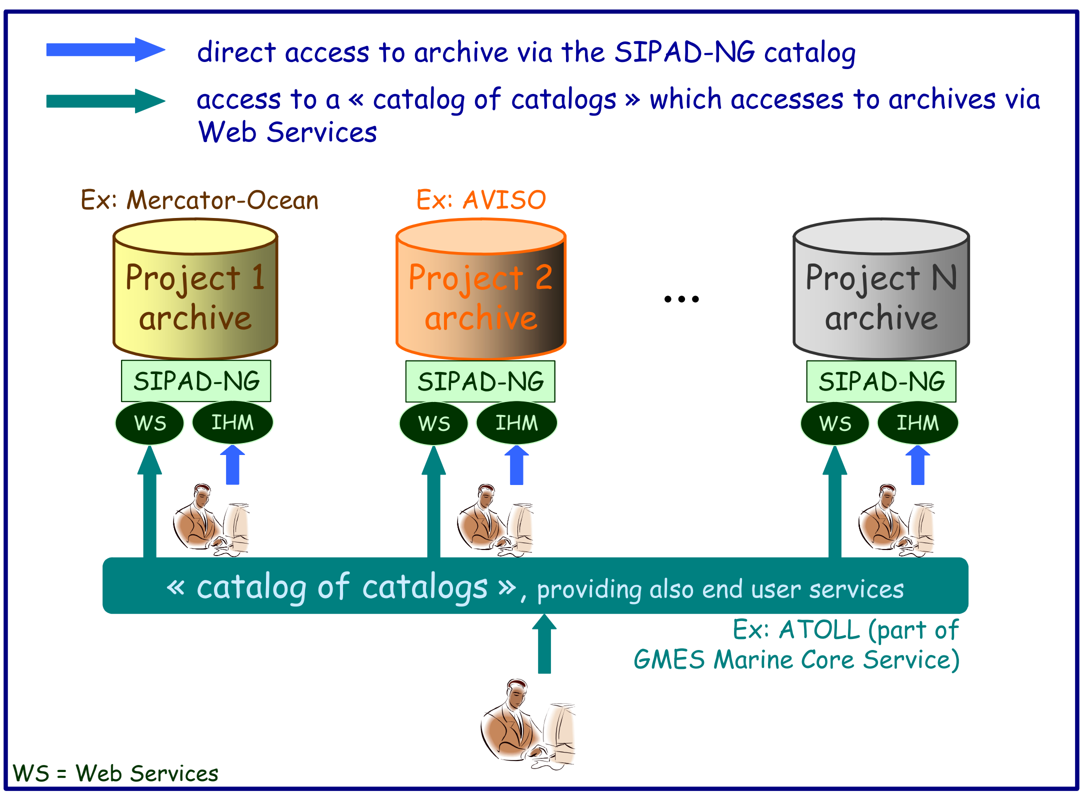


Figure 7: The two different ways to access Sipad-NG

The web services interface provided by *Sipad-NG* is based on **SOAP protocol**, which is used to consult the catalog and to order data. For performance reasons, FTP protocol is used to deliver data.

In 2010, *Sipad-NG* will provide an interface compliant with the **ESA HMA (Heterogeneous Mission Accessibility)** protocol.

4 – Presentation of ATOLL to access timely time-serie ocean data

Capitalizing on its participation to European and international projects (Mersea / MyOcean -GMES-, Seadatanet -FP7-, Humboldt -Inspire-, Godae -IOOS-), CLS developed ***Atoll,*** an end to end Information System for managing, referencing and sharing of ocean data and services. *Atoll* is an Inspire Spatial Data Infrastructure which meets also ITIL service Management. It is a generic software suite which can be activated through 4 user views:

1) systematic use of metadata and network Web Services (scheduled or advanced delivery, GIS), in compliance with the normative frames promulgated by ISO, OGC and W3C.
2) monitoring and reporting portal giving a view of Key Performance Indicators like availability and timeliness of products, product dependency traffic light, or transaction statistics.
3) discovery or application web portal (catalogue for full discovery, one stop thematic portal, front-end demonstration window).
4) interoperability of catalogues (data mapping, or harvesting of network services).

Atoll acts as an end-to end data management system, semi distributed system, which ensured that ocean products are visible, manageable, catalogued, accessible and utilized. It tackles the problem of both the diversity of data, the variety and distribution of interfaces to access the data, and volume of data flows to be served in a timely manner.

Atoll allows sharing expertise to describe and present well-described, peer-reviewed datasets, to apply a complex data policy and deploy the Inspire functions for data and service sharing. It ensures clarity on data origin and transformation processes as well as the various actors who intervene in the process to produce the requested ocean information.

Atoll resolves the inventory need and manage the timeliness and accessibility of products (and any other need reported in the Operating Level Agreement – OLA – of data centers).

Atoll avoids cost effective implementation of an Inspire infrastructure at each data center and enables to present datasets with various point of view.

The logical infrastructure which is implemented relies on three main components *as shown on figure 8*:

1) The first component of *Atoll* is dedicated to manage the interfaces with the production centres in the scope of maintaining a full and up to date picture of products to users, and implement a service bus for data access. Among others, it reduces the complexity of filling in the metadata by sharing its edition with multiple actors together with an automation of the inventory/accessibility function per pair "product – delivery mean" and harmonise/standardise/control the interfaces for users (standardised interfaces for along-track products, for gridded products and forecasts).

2) The second component of *Atoll* implements the Inspire referencing and network services architecture as mandated by the directive to discover data products, visualize, access and transform them, and how those services connect to portals and applications through the Inspire Service Bus. Among others, it has to tackle the issue of user authentication / rights management which are needed to invoke e-commerce services or data and service sharing in a distributed context. It has also to report on transactions (performance and use feedback) from top – use level – down to the data centre place.

3) The third component of *Atoll* is to monitor service commitment to react and anticipate incident handling, and report on uses and system usefulness. Services to user are rendered through Service Level Agreement, link with distributed sub-systems by Operating level Agreement.

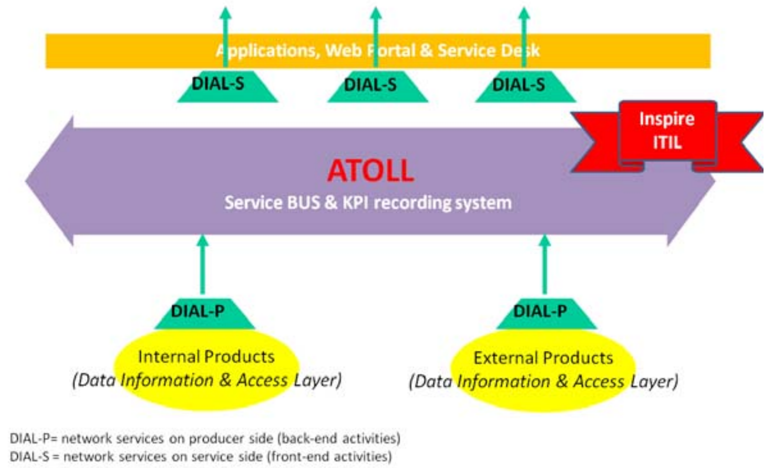


Figure 8: Conceptual top level view of Atoll end-to-end data management system

Atoll has performance and quality skills that make it a fully operational server. From an architectural point of view, *Atoll* is based on the modern principle of cutting into levels of responsibility for which functional boundary is perfectly demarcated *(see figure 9).* Those levels are then involved in a collaborative and, eventually distributed, environment. The inputs to the bus are handled within conceptual connectors called **DIAL** - Data Information and Access Layer -, a DIAL-P for interfaces with Productions Units, and a DIAL-S for client connections from Applications, Web Portal, Service Desk, etc. DIAL-S is directly constrained by the Inspire network services requirements.

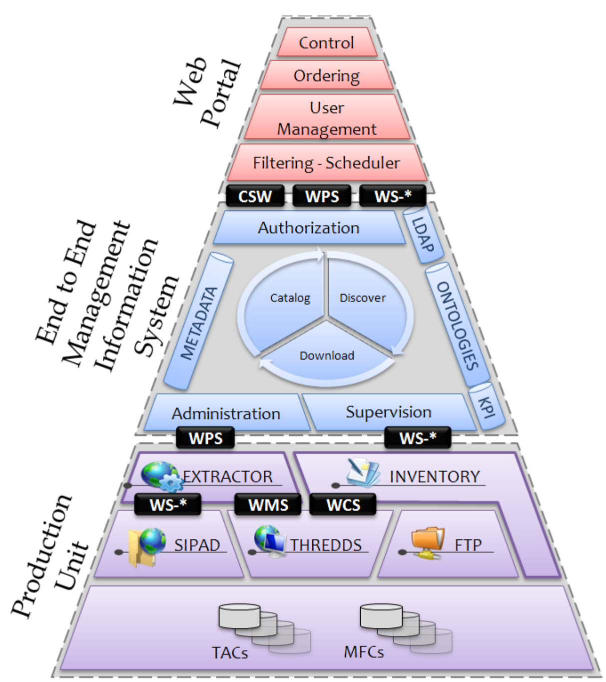


Figure 9: Atoll architectural design

The base line technology relies on an ocean data model, a ThREDDS data connector, an inventory module coupled to an extractor engine (MOTU), a service bus and a supervision system assessing the service performance (central *Atoll*):

1. The ocean data model

The ontology is the baseline to identify and model our referential; the knowledge of data relationship constrained with taxonomy being the key to offer many user point of views. CSML ocean data model is tested in the ocean science community and has recently been submitted to ISO TC211 as a candidate standard developing a suite of ontology data models for points, grids, profiles and swaths to be implemented principally as GML Application Schema (Geography Markup Language - ISO 19136). It describes the production environment (for instance the observing system, datasets used to produce it, ocean parameters and reference systems, application regions and schemas, qualification status) as well as the delivery means together with its accessibility and timeliness commitments or other thematic extensions, specificity of the information system.

For short term commitments, an intermediate ocean schema has been implemented, and relies on reference vocabularies as developed within Seadatanet project. 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. The encoding is ruled with the following standards: W3C OWL, SKOS, RDF.

Then XML metadata are instantiated for each node of the catalog, generating metadata. The metadata information covers ISO 19115 for description of data and ISO 19119 for description of services). The XML profile is ISO 19139 and Inspire compliant (tested within the Humboldt Inspire implementation project). This profile participates also to the international process (concrete exchanges with similar community than European myOcean, IOOS US ocean community,I MOS Australian ocean community, and the wider CSML community).

2. The ThREDDS data connector

The ThREDDS[4] technology software, provided by Unidata (cf. http://www.unidata.ucar.edu/projects/THREDDS/), is a middleware that bridges the gap between data providers and data users. It has been widely and successfully applied by the ocean community to connect users to data products, and thus allowing powerful and flexible access to gridded data. The software is mainly based on three pillars:

- The NetCDF file format as units of data,
- The Climate and Forecast (CF) metadata conventions.
- The OPeNDAP protocol as a means of network access to the data.

Use of this trio (Hankin et al. 2009) allows handling 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 (Blanc et al., 2008). Recent advances in interactive data viewing functions has pushed the development of new web mapping service usable by the GIS community. This web mapping service is now mature for gridded products (like ncWMS) and has been included in the Version 4.0 of the ThREDDS Data Server. Research is ongoing to apply

[4] ThREDDS stands for “Thematic Realtime Environmental Distributed Data Services”.

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.

3. The inventory module coupled to the extractor engine (MOTU)

The inventory module is responsible to brings to *Atoll* the knowledge of the concrete treats of the managed products and the way to access them. An inventory is lively and progressive. For that, the inventory module explores producer repositories and maintains a list of hosted products. It then records it into an inventory file.

A monitoring function in XML format which is implemented for each pair (product line – delivery mean) and report when needed on product delivery stream and status to be recorded in the information system (cf. production window and update frequency, real time window, last production date, volume size ….). This monitoring function is directly linked to the supervision system.

The extractor engine called MOTU is a web processing service for bulk or advanced delivery to users. Requested data, for one shot or regular delivery, may undergo several processing that include the extraction of parameters, the spatiotemporal (4D) extraction, the processing (to other formats, other projections), the packaging (compression, file organization, delivery note inclusion), and then deliver (deposit into an FTP directory, mail sending).

From this point of view, it is considered as the operational arm of *Atoll*. The synopsis presented on *figure 10* describes the internal logical view and in particular the processing chain of data.

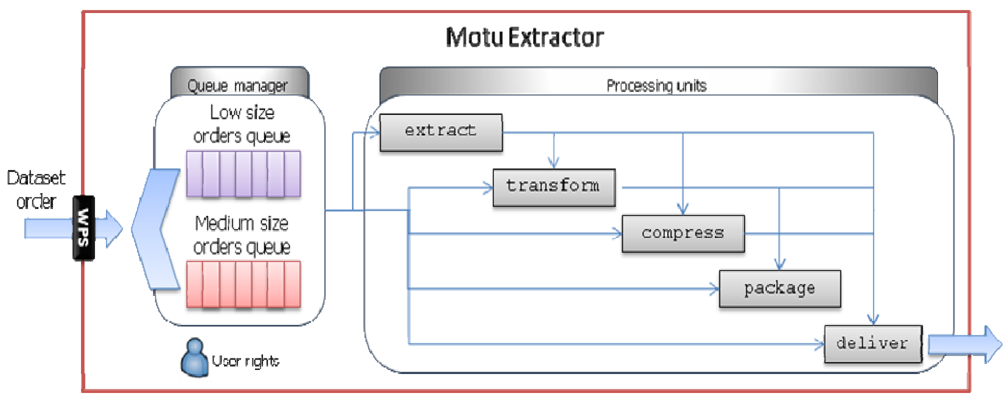


Figure 10: Internal logical view of Motu Extractor

Motu Extractor handles incoming requests through the WPS protocol and puts each of them into an appropriate queue. It manages several queues and has the ability to distribute an order according to its characteristics, like the expected size of the delivery. Each queue has a policy to stress a specific processing aspect and thus reconfigure the way requests are carried out. So, by controlling the bandwidth and curbing the deny of service, Motu reaches a status of operationnality.

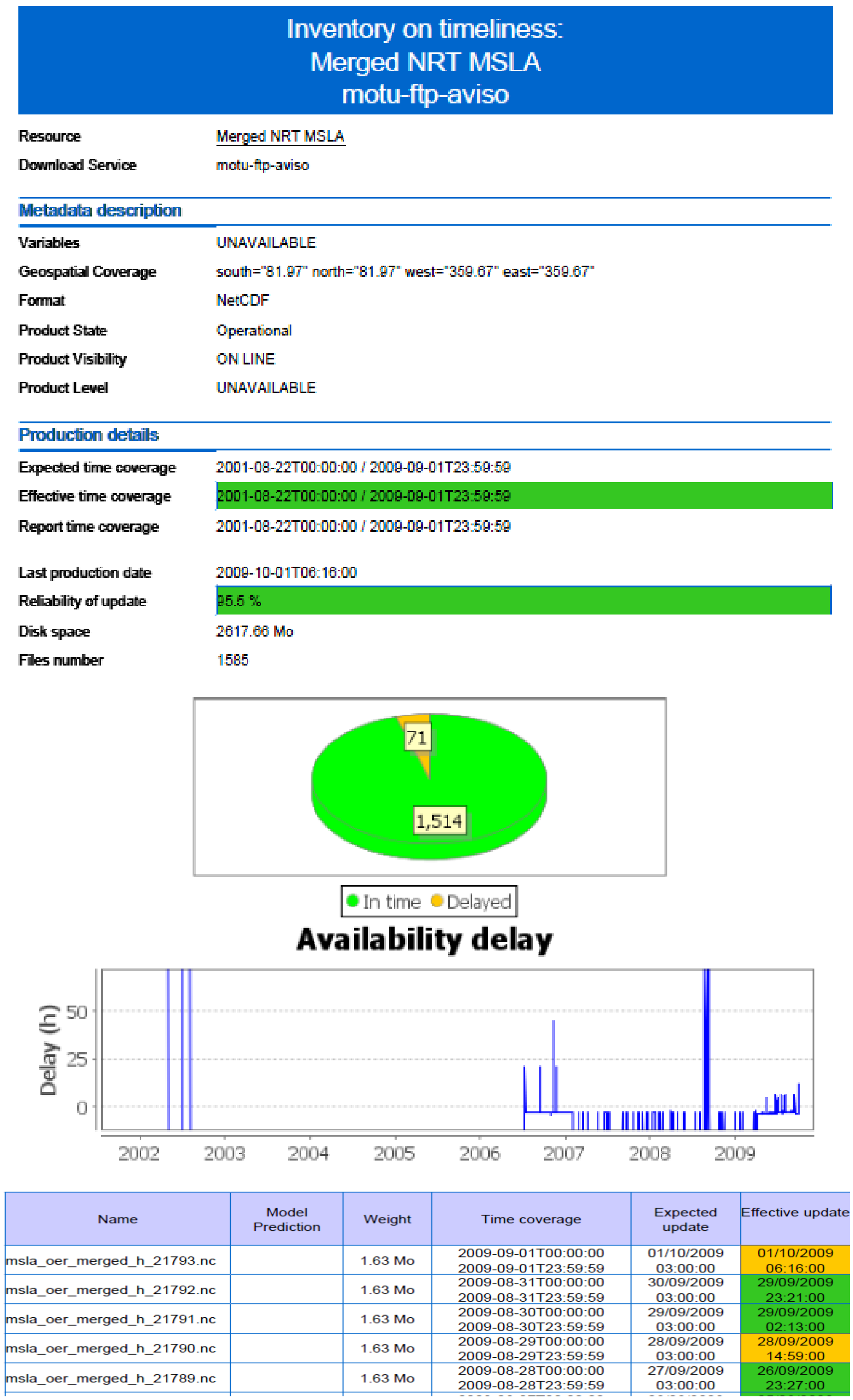
The processing units are atomic elements that perform a dedicated task in the processing flow of the request. Those elements can be chained to get a higher functional level, from the physical data manipulation to an end user delivery.

4. The service bus and the supervision system assessing the service performance

The service bus publishes 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. At this stage, those services are directed at authentication rules and monitoring of activity.

The supervision system acts as a recorder to monitor product availability, accessibility and timeliness on a short term system reactivity and on a longer term system usefulness. This system relies essentially on 2 supervision functions:

- A well-known Cacti graphing solution and Nagios monitoring system.
- Visualisation of results of the monitoring function addressed above as shown on *figure 11*.



Inventory on timeliness:
Merged NRT MSLA
motu-ftp-aviso

Resource	Merged NRT MSLA
Download Service	motu-ftp-aviso

Metadata description

Variables	UNAVAILABLE
Geospatial Coverage	south="81.97" north="81.97" west="359.67" east="359.67"
Format	NetCDF
Product State	Operational
Product Visibility	ON LINE
Product Level	UNAVAILABLE

Production details

Expected time coverage	2001-08-22T00:00:00 / 2009-09-01T23:59:59
Effective time coverage	2001-08-22T00:00:00 / 2009-09-01T23:59:59
Report time coverage	2001-08-22T00:00:00 / 2009-09-01T23:59:59
Last production date	2009-10-01T06:16:00
Reliability of update	95.5 %
Disk space	2617.66 Mo
Files number	1585

Name	Model Prediction	Weight	Time coverage	Expected update	Effective update
msla_oer_merged_h_21793.nc		1.63 Mo	2009-09-01T00:00:00 2009-09-01T23:59:59	01/10/2009 03:00:00	01/10/2009 06:16:00
msla_oer_merged_h_21792.nc		1.63 Mo	2009-08-31T00:00:00 2009-08-31T23:59:59	30/09/2009 03:00:00	29/09/2009 23:21:00
msla_oer_merged_h_21791.nc		1.63 Mo	2009-08-30T00:00:00 2009-08-30T23:59:59	29/09/2009 03:00:00	29/09/2009 02:13:00
msla_oer_merged_h_21790.nc		1.63 Mo	2009-08-29T00:00:00 2009-08-29T23:59:59	28/09/2009 03:00:00	28/09/2009 14:59:00
msla_oer_merged_h_21789.nc		1.63 Mo	2009-08-28T00:00:00 2009-08-28T23:59:59	27/09/2009 03:00:00	26/09/2009 23:27:00

Figure 11: Monitoring product timeliness

A web portal is however required to overlay the needed thematic layer to guide the user to the "product" he needs, taking into account the various definition of products organize products between them, and to support operations, administration and supervision. *Atoll* monitors consequently the production for operational delivery.

5 – Final words

As shown on figure 2 and described in section 3, the CLS *Atoll* portal accesses CNES *Sipad-NG* by using the web services interface. Two independent *Sipad-NG* systems are interfaced with *Atoll*: the SALP/AVISO archive and the Mercator Océan archive. *Atoll* sends consultation requests (to get a list of datasets available in the archives) and order requests (to retrieve data). Data are transferred from *Sipad-NG* systems to *Atoll* by FTP protocol. Then *Atoll* transfers these data to end users (*Atoll* is in charge of the user management function).

Atoll will be enhanced and shared with Ifremer Information system to meet MyOcean Information System objectives (MIS), that is data management and data accessibility in a strong operational and sustainability context.

Within MyOcean, we will develop interoperable link between MIS and

1. HMA (space earth science measurements) and Seadatanet (In situ ocean measurements) to get access to measurements data,

2. IOOS (US system) and IMOS (Australian system) to enlarge the spectra of Myocean (products and users ina similar context),

3. WIS (Weather Information System) and Geonetcast/Eumetcast to benefit on a new distribution media and reach the metorological user community.

However the work will remain at the discovery level unless the use feedback (downward traceability of the transaction process) has been well defined and agreed by the various parties. This is the needed condition today to move forward on the concept of constellation of catalogues for an increased use, and still giving visibility at the lower level of the data usefulness and use. This is the key issue for long term sustainability of all the actors intervening in the distribution process, as well as to give the knowledge of further user requirements, and priority to be given to future developments.

We would like to end this paper by a focus on the organizational aspects. 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, and how to connect users and products. Major roles identified today to organize data management activities are:

- System of systems architecture designer and maintenance of interfaces,
  including order of connection. integration between systems and well filled and up to date metadata,
- System and service performance definition and measures,
  operations and system capacity monitoring
- Knowledge and execution of complex data policy and product governance,
  and user right management

6 – References

2009, Blanc et al., "Evolution in data and product management for serving operational oceanography, a GODAE feedback", OceanObs,

2009, Hankin et al., "NetCDF-CF-OPeNDAP: Standards for Ocean Data Interoperability and Object Lessons for Community Data Standards Processes", OceanObs, 2009, de La Beaujardière et al., “Ocean and Coastal Data Management”, OceanObs,

2009, Blower et al., “Serving GODAE data and products to the ocean community”, JOO,

2008, Baudel et al., “Ocean view via an integrated downloading and visualization system benefits from ThREDDS data servers”, Mersea

2008, Blanc et al., “Data and product serving, an overview of capabilities developed in 10 years”, GODAE,

2008, Blanc F., “The relationship between SALP/AVISO and MyOcean, information systems view”, Eumetsat,

2008, Blower et al., “Underpinning technologies for oceanographic data sharing, visualization and analysis: Review and future outlook”, GODAE,

2008, Blanc F., “MyOcean Information System”, Eurogoos,

2008, de La Beaujardiere, “The NOAA IOOS Data Integration Framework: Initial Implementation Report,” *Proc. MTS/IEEE Oceans’08,*

2008, Loubrieu et al., “A distributed system for operational oceanography: the MERSEA System”, JOO,

2007, Heulet et al., "SIPAD-NG & SITools: Two systems for accessing scientific data and Added Value Services - Comparison between the two systems - Feedback on experimentation", PV 2007

2004, Heulet, "SIPAD-NG: a multi-field system for accessing scientific data and added-value services", PV 2004

2004, Hankin et al., “Data and Communications Infrastructure for the U.S. Integrated Ocean Observing System”, ESDIM.

2003, Cornillon et al., "OPeNDAP: Accessing Data in a Distributed, Heterogeneous Environment," Data Science Journal,