Joint European Research Infrastructure network for Coastal Observatories

First Data Management Report
D5.3

Grant Agreement n° 262584
Project Acronym: JERICO

Project Title: Towards a Joint European Research Infrastructure network for Coastal Observatories

Coordination: P. Farcy, IFREMER,

jerico@ifremer.fr, www.jerico-fp7.eu:

Authors: C. Fanara, A. Crise, L. Petit de la Villeon, S. Pouliquen
Involved Institutions: OGS, IFREMER
Version and Date: v. 5.0 – November 2013

Table of Contents

1. Document description

REFERENCES

Annex 1 to the Contract: Description of Work (DoW) version 2011-02-22

Document information	
Document Name	First Data Management Report
Document ID	
Revision	
Revision Date	
Authors	C Fanara, A Crise, L Petit de la Villeon, S Pouliquen
Security	

History			
Revision	**Date**	**Modification**	**Author**
1.0	2013.03.15	Creation of the document	C Fanara, A Crise
2.0	2013.04.24	Integrations	L Petit De La Villeon, S Pouliquen
3.0	2013.04.29	Integrations	C Fanara, A Crise
4.0	2013.09.25	Integrations	C Fanara, L Petit De La Villeon
5.0	November 2013	Final draft	C Fanara

Diffusion list				
Consortium beneficiaries	X			
Third parties				
Associated Partners				

1. Summary

The JERICO approach is strongly based on the “use of what exists” through the creation of suitable partnerships with ongoing European data management initiatives for the minimization of duplication of efforts. Thus, there has been no development of a specific data management structure for JERICO. Instead, the use of, and integration with, already available data management infrastructures is being promoted. This strategy is consistent with the policy behind SeaDataNet and MyOcean, the major ongoing European initiatives for the establishment and coordination of infrastructures for the management and distribution of marine data and products.

By treasuring the lesson learned from these projects, the JERICO data management framework for delayed-mode data uses the SeaDataNet (SDN) infrastructure, while real-time data are being handled through MyOcean (MyO).

The JERICO approach is driven by the great importance that MyO and SDN initiatives have had in the last few years, since both systems proved to be robust and successful in the archiving and distribution of marine data, and correspond to a perspective of long-term sustainability for the European marine infrastructures.

Continuous interaction with MyO/EuroGOOS and SDN takes place in order to facilitate seamless integration with both these established infrastructures for managing the JERICO data stream.

This Data Management Report wants to illustrate the JERICO approach and its links, together with the efforts made during the first 24 months of Project with reference to the Data Management and Distribution (WP5) and the results obtained so far as the contribution to the activity performed in Services and Data Access (WP7).

This first version of this Document will be updated (month 48) to a final version in order to account for the achievements made during the whole Project.

2. Introduction: the JERICO approach

The JERICO project comes from the growing necessity to bring together the representative European coastal observatory systems, enhancing their coordination and promoting the cost-effective use of their facilities, with the final aim of supporting the continuous and efficient provision of high quality environmental data and information products related to the marine environment in European coastal seas.

The JERICO system is “innovative” in the in the sense that, for the first time, there is a common European organization that consolidates and harmonizes the currently dispersed coastal observing activities in a sustainable and coherent framework to the benefit of data quality, availability and costs.

JERICO proposes a pan-European approach for a European coastal marine observatory network that integrates distributed infrastructures and different technologies (such as moorings, drifters, ferryboxes and gliders). Through its networking activities JERICO intends to facilitate the optimal use of existing equipment and promote interactions with other European infrastructures (such as MyO, SDN, EMODNET). This necessarily implies the definitions of common best practices for maintenance and quality control, as well as quality standards for sensors and data exchange.

In this context, JERICO aims to provide a platform for the identification and dissemination of best practice, the definition of quality standards and the promotion of interoperability. This comprehensive approach is geared to the delivery of a reliable, cost-effective observational infrastructure based on an end-to-end concept of coastal monitoring that covers all steps leading from data acquisition to data dissemination.

With reference to data management, the success of the JERICO Project mainly depends on the goal of assuring that the flow of real-time and delayed mode data, coming from the participating observing networks, will be reliable, accessible and easy to distribute. To get this purpose it was necessary to promote the use of common procedures for both real time and delayed mode data management within the JERICO community.

In this frame, the first version of two Data Management Handbooks, one for delayed mode data and the other for real time data, have been of practical advice in harmonizing the data management procedures for the JERICO community.

The procedures here recommended were those that treasured the lesson learned from SDN and MyO/EuroGOOS, as the major European initiatives for the establishment and coordination of infrastructures for the management and distribution of data and products.

Though most of the JERICO Partners already contribute to (or will contribute to) SDN and MyO, there are other “external” users that are not acquainted with the specific data management procedures used within both systems. This is why JERICO intends to support those users not only through practical manuals dealing with data management but also with dedicated activities performed by specific sub-contractors, as detailed in paragraph 3.4.

3. Main Report

3.1. SeaDataNet: overview and links

SeaDataNet (SDN) is an efficient distributed European Marine Data Management Infrastructure for managing, indexing and providing access to large and different, marine and ocean sets of data, deriving from in situ and remote observations.

The SDN system was implemented, in the frame of EU FP6, during the first phase of the SeaDataNet Project that started in 2006 and had 5 years duration.

The second phase of SDN (SeaDataNet 2 Project) started on October 2011 and will last 4 years. It aims to upgrade the present SDN infrastructure into an operationally robust and state-of-the-art Pan-European infrastructure for providing up-to-date and high quality access to ocean and marine metadata, data and products.

SDN is based on the interconnection between currently existing infrastructures active in data collection (National Oceanographic Data Centres and Marine Data Centres of 35 countries). The networking of these professional data centres, in a unique virtual data management system, provide integrated data sets with common standards for metadata, vocabularies, data transport formats, quality control methods and flags, access.

SDN is a multi-disciplinary system: its architecture is designed to support a wide variety of data types. The Partners manage large and diverse sets of marine and ocean data, originating from their own institutes and from other organisations in their country. The majority of those data concerns physical oceanography, marine chemistry, hydrography, and an important volume of marine biology, geology and geophysics.

The key objective of SDN is to provide an integrated and harmonised overview and access to these data resources, using a distributed network approach, where interoperability between platforms is ensured through the development and adoption of common communication standards and improved technology.

A fundamental aspect within SDN is the development and the implementation of the Common Data Index (CDI) metadata service that provides the link between discovery and delivery of data sets (such as samples, timeseries, profiles, trajectories, etc.). The CDI user interface enables users to have a detailed insight of the availability and geographical spreading of marine data, archived at the different interconnected data centres. Besides, this service provides the means for downloading data sets in common formats via a unique user interface.

The CDI, initiated in the Sea-Search project, was further developed and upgraded within the SDN project. Extensions and updates are being made in SDN2 in order to apply the CDI for an extending range of ocean and marine data types and meet requirements of on-going associated projects. Recently the ISO 19115 content model, which is the basis for the XML formats and exchange schema's (XSD), has been upgraded to ISO 19139 format in order to be compliant with the INSPIRE directive and make the exchange between the national providers and the central portal operators more effective.

SDN also points to share its technologies and expertise, to develop its approach and to build bridges to other qualified infrastructures active in the marine sector.

With the aims of achieving a wider coverage of data sources and an overall interoperability between data infrastructures, SDN is currently pursuing active cooperation on a national scale with other data holding organisations via its NODC networks and on an international level with other data management networks/projects. These are Geo-Seas, UpgradeBlackSeaScene, CaspInfo, EMODNET and EMODNET 2. All projects are adopting the SeaDataNet standards and extending its services.

Active cooperation takes also place between SDN and EuroGOOS/MyOcean2 in the field of real-time and delayed mode meteo-ocean monitoring data.

3.2. The MyOcean Project and the future ECOMF

MyOcean (MyO) is the EU FP7 project responsible for the development and pre-operational validation of the Ocean Monitoring and Forecasting component of the GMES (Global Monitoring for Environment and Security) Marine Core Service. It is represented by a consortium of 61 partners (both public institutions and private enterprises), coming from 29 different European countries.

The Project started in January 2009 and ended in March 2012. The follow-up project MyOcean2 (MyO2) is now ensuring the continuity of services, in terms of development and improvement. MyO2 wants to work for the sustainability of the infrastructure and aims at pooling the resources coming from the various European Ocean Forecasting Centres in order to create a fully operational European Ocean Monitoring Service by 2014.

MyOcean is centred on the activity performed by different production units that generate and operate observations, model-based data, and added-value products. These components are divided into:

- 5 Thematic Assembly Centres (TAC), all dealing with a specific set of observation data: Sea Level, Ocean colour, Sea Ice & Wind, and In-Situ data;
- 7 Monitoring and Forecasting Centres (MFC) to serve the global ocean, the Arctic area, the Baltic Sea, the Atlantic North-West shelves area, the Atlantic Iberian-Biscay-Ireland area, the Mediterranean Sea and the Black Sea.

MyOcean does not operate in situ observing systems but collects observations from data providers outside MyOcean, mainly from JCOMM networks (Argo and Euro-Argo are among the most important in-situ observing systems for MyO) and EuroGOOS regional alliances.

EuroGOOS has developed a European regional structure for the integrated production of observations and services, so called "Regional Operational Oceanographic Systems" (ROOSs) that are related to six different regional sea areas. They are:

- ARTIC-ROOS: the Arctic Regional Ocean Observing System
- BOOS: the Baltic Operational Oceanographic System
- NOOS: the North West European Shelves Operational Oceanographic System
- IBI ROOS: the Ireland-Biscay-Iberia Regional Operational Oceanographic System
- MONGOOS: the Mediterranean Operational Network for the Global Ocean System
- BLACK SEA GOOS: the Black Sea Global Ocean Observing System.

Since MyO is very dependent on the availability of global and regional in-situ observations, EuroGOOS members and partners play a very important role in the success of MyO2 and the construction of an operational GMES Marine Service.

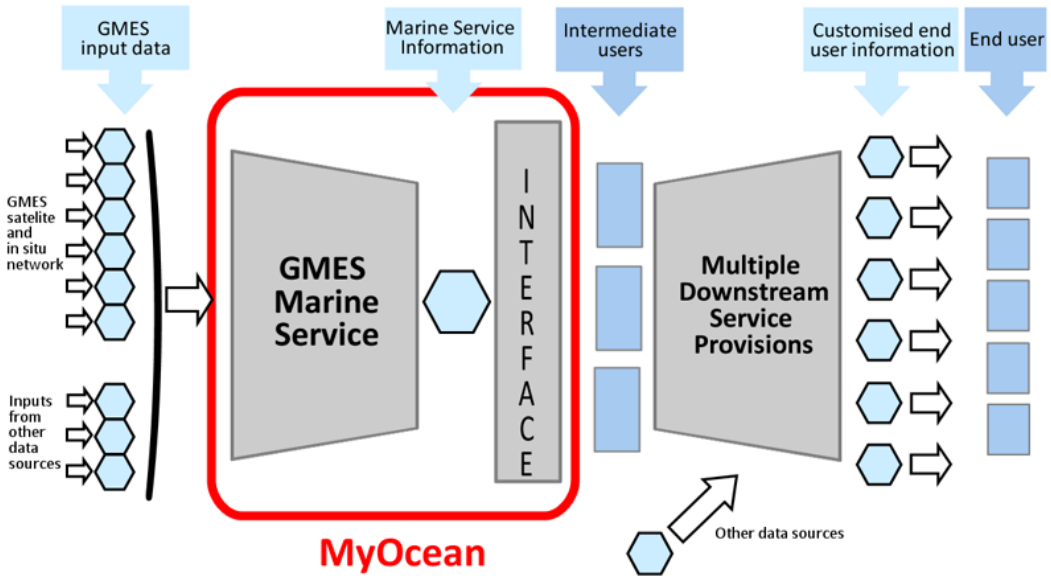
The MyO approach is “user driven”: services and products offered are open and free to all kinds of customers. The service is intended for any actors operating in the following marine application areas: maritime safety; marine resources; marine and coastal environment; climate seasonal and weather forecasting. Quality, simplicity and sustainability are the most important features of the service.

Users can benefit from homogeneous quality and validation information for most of the products. The quality assessment activity is performed by all the Production Centres.

Through its Portal (http://www.myocean.eu/), MyOcean offers a reliable and easy access to a single Catalogue of products, regularly updated. From the online Catalogue, users can download products according to their needs in a unique format (netCDF).

The available information (about 100 products) includes observations, analysis, reanalysis and forecasts describing the physical state of the ocean (through the main parameters: temperature, currents, salinity, sea ice, sea level, wind) and its primary biogeochemical parameters. MyOcean also contributes to research on climate by providing long time-series of reanalysed parameters.

The Catalogue is interactive and this allows users to select products according to the 7 geographical areas; the parameters/variables; the product. User can also select a product by using a keyword search or the pdf catalogue; or by accessing the full online catalogue.



The GMES Marine Fast-Track Service chain and its Ocean Monitoring & Forecasting (OMF) component. Source: www.myocean.eu

MyOcean2 represents the ultimate step before the transition to ECOMF (European Centre for Ocean Monitoring and Forecasting) during 2014.

ECOMF will be the organization delivering the ocean monitoring and forecasting component of the GMES Marine Service. This “European Centre” will take care of the core service part and will aim to pool high-level investments, to simplify interfaces and to deliver commonly needed data.

Based on the MyO and MyO2 experience, ECOMF will be the key component of the future of Operational Oceanography in Europe.

MyOcean2 in context. Source: www.myocean.eu

3.3. Actions taken for the harmonization of Delayed Mode and Real Time data management procedures. Implementation phase.

Since the start of the project, discussion have been carried out with both the SeaDataNet, MyOcean and the EuroGOOS ROOSs to define what was the best way to integrate both Delayed Mode (task 5.2) and Near Real Time (task 5.3) observations from the JERICO coastal observing systems.

The proposed solutions were discussed with the EuroGOOS ROOSs in some workshops, and agreement was reached to support a common approach. The discussions with the EMODnet-PP project that was facing similar issues were also taken in account so that the solutions proposed within JERICO and EMODnet-PP were consistent.

3.3.1. Harmonization of the management procedures for Delayed Mode (DM) data

The Delayed Mode Data Management Handbook (v.1) provided a practical basic guide to data contributors: the document described the procedures to follow in order to facilitate data and meta-data flow to the SeaDataNet data repository infrastructure.

Other concrete actions to harmonize DM data management procedures were subcontracted to MARIS and MARUM. These activities are detailed in paragraph 3.4.

3.3.2. Harmonization of the management procedures for (near) Real Time (nRT) data

The Real Time Data Management Handbook (v.1) provided a practical basic guide to data originators: the document described the procedures to follow in order to make data available and circulate through the MyOcean real-time data pipeline.

This latter is based on the Regional Operational Oceanographic Systems (ROOSs) that provide data to the MyOcean in situ TAC (Thematical Assembling Centre) hosted by the Coriolis data centre.

At the beginning, the solutions suggested in the Handbook proved to be applicable for most of the data providers that began taking part in the Services and Data Access activity on January 1th, 2013. However, the implementation of the procedures for the preparation and integration of datasets proved more difficult than expected.

The following table shows the platforms involved and the status of the datasets integration (until September 2013) through the nRT data stream. Further details can be found in WP7 (Services and Data Access).

1) MOLIT & Mesurho buoys
2) RECOPESCA (158 vessels)
3) Alg@line 3 Ferrys : - Finnmaid (call sign = OJMI): data reaching the Coriolis/MyOcean data flow. - Silja Serenade (call sign = OJCS) and Kristina Brahe (call sign = OIEC) : No data. Contact taken. Data will flow through NIVA
4) CRS - Coastal Research Station 1 coastal station and 1 mooring - Contact taken - Data integration process started
5) NorFerry - Norwegian Ferrybox network 3 Ferrys : - Norbjorn (call sign = LAKM4) - Trollfjord (call sign = LLVT) - Bergensfjord (call sign = OUZI2)
6) NorFerry – ColorFantasy Color Fantasy (call sign = LMSD)
7) IMR - Coast observatories No data. Would like to identify contributor in the NetCDF files. Discussion pending.

8) OGS-NACObs - FVG-MMS Development for data integration is started but not completed
9) OGS-NACObs – MAMBO Data reaching Coriolis/MyOcean data flow since June 2013
10) CNR - NAMS Data reaching Coriolis/MyOcean data flow since July 2013
11) CNR – FOS Contact taken Data will flow to Coriolis/MyOcean through HCMR
12) POSEIDON Buoy Network 8 stations
13) POSEIDON Buoy Network 3 stations 1 Ferry : Olympic Champion (call sign = SYWD)
14) POL - COBS No answer to a mail sent by coordinator
15) COSYNA 3 Ferrys : - Hafnia Seaways (call sign = 2AMH9) : No data - FunnyGirl (call sign DFPZ) : Data reaching Coriolis/MyOcean Database - LysBris (call sign = LJLN3) : Data reaching Coriolis/MyOcean Database - Wadden Sea Piles : Data integration process started
16) SMHI - MOS Contact to be confirmed
17) SMHI - Laesoe
18) SmartBay Galway
19)Puertos del Estado Deep Water Network

Status of the integration of the JERICO datasets through the NRT data stream
(from WP7 report - last update September 2013)

Data set circulating in NRT
Data set not yet integrated in a NRT data stream
Contact taken with the data provider. Development for data integration is started.

3.4. Sub-contracted activities

With the aim of giving specific support to the JERICO community dealing with data management activity, some specific tasks were sub-contracted (by IFREMER) to the following companies/institutions: MARIS bv., MARUM – University of Bremen and CAPGEMINI Consulting.

The planned activities of each subcontractors are summarized below. Wherever possible, details on the progress made are given, too.

3.4.1. MARUM – University of Bremen

As set out in the DoW, there is a plan to move up one level and to design a SensorML profile for observatories within JERICO, taking into account the work that has been done so far in the ESONET NoE project.

The activities subcontracted to MARUM are mainly dedicated to the application of standards from OGC's (Open Geospatial Consortium) Sensor Web Enablement (SWE) suite to:

a) set up procedures to describe accurately marine observatories and their on-board sensors (including calibrations, processing, accuracy limits, …) using SensorML standards;
b) access observatory descriptions and data using a Sensor Observation Service (SOS) and the corresponding XML O&M (Observations and Measurements) encoding;
c) publish and access digital catalogues of metadata for geospatial data, services and related resource information by using OGC's Catalogue Service for the Web (CSW).

Since the SensorML is a very comprehensive standard, which has to be downsized to be easy to use in a specific domain like marine observations, the MARUM activities include:

- the definition of a SensorML standard subset (“profile”) which records only the descriptors (i.e. SensorML fields) which are useful for JERICO;
- the establishment of a guide for best practices to enter JERICO descriptions (metadata) of observing systems;
- the gathering of these metadata within a sensor registry accessible through the JERICO portal.

3.4.2. MARIS

The activities of MARIS within the JERICO project are mainly dedicated to give support to the observation providers of historical data which are not able to access NODCs in order to utilize the SDN infrastructure.

Principally, MARIS is in charge of providing assistance to partners that cannot use the SDN procedures and software in:

- populating the EDIOS catalogue (European Directory of the Ocean-observing Systems) with JERICO data descriptions;
- generating the CDIs (Common Data Indexes) that will allow access to the JERICO datasets through the SDN portal;
- setting up a prototype of a portal (based on SDN technologies) which will allow direct access to JERICO data;
- installing and configuring SDN tools for data management/data access and services (installation of components; usage of formatting tools, …).

Specific attention is being given to partners who have decided not to transmit their data to National Oceanographic Data Centres (NODCs) but to directly interface their data repositories with the SDN system.

3.4.3. CAPGEMINI Consulting

A sub-contract has been signed with CAPGEMINI Consulting to develop a specific indexing and data distribution scheme and a link between observatories and the MyO TACs.

An appropriate tag (JERICO index) is important for identifying and selecting JERICO data when they are embedded in larger archives. This measure will also permit to clearly identify the JERICO contribution to the global observation system, and are consistent with the PERSEUS and MOON projects that follow the same approach.

3.5. Data providers and their data management activity

The first versions of the DM and (n)RT Data Management Handbooks were prepared and distributed to the JERICO community for comments and application.

Later on, at the start of the Services and Data Access (January 2013), some critical feedbacks came from those data providers that had experimented the procedures provided in the Handbooks.

The management of the integration with the SDN and MyO systems turned out to be more complex and demanding than expected.

The data providers found the Handbooks only partially useful, and asked for descriptions of step-by-step procedures that, in their opinion, were not sufficiently clear in the documents.

Problems were mainly encountered by those partners that were not acquainted with the MyO data format and specifications.

Some providers opted to send their data directly to the in-situ TAC in the original in-house format and ask for the conversion of the data format and the application of the correct data quality control procedures.

Other JERICO partners performed the generation of the data files into the MyO format (OceanSites NetCDF) on their own. In these cases, there were some technical problems in adapting databases to the new procedures and in mapping them onto the different NetDCF codes. The delay in providing data is also (but not only) due to this reason.

4. Conclusions

The JERICO approach, based on the “use of what exists”, promotes the application of, and integration with already available systems devoted to the management and distribution of marine data and products, such as SeaDataNet and MyOcean.

By treasuring the lesson learned from these initiatives, the JERICO data management framework has been implemented, using the SeaDataNet infrastructure for delayed-mode data and the MyOcean architecture for the real-time data.

Among the WP5 actions accomplished in order to facilitate the management, distribution and integration of data within both these infrastructures, there was the preparation of the first versions of two Data Management Handbooks, one for the Delayed Mode and the other for the (near) Real Time data.

The application of the procedures suggested in the Handbooks following the start of the Services and Data Access activity (January 2013), highlighted some criticalities. The management and preparation of datasets to be integrated in MyO and SDN systems turned out to be more complex and demanding than expected, especially by those data providers that were not acquainted with these processes.

Some specific activities have been subcontracted to third parties, with the aim of giving support to the JERICO community in dealing with data management, and are still in progress.

Future activities will be dedicated to strengthen the actions assisting data providers: more details on the description of the data management procedures will be included in the final version of both handbooks; a stronger and closer collaboration with WP7 (Services and Data Access) will be set up. Finally, a greater participation by HCMR, as the infrastructure responsible for the acquisition, processing and storing of JERICO data, will be sought to help with technical issues faced by data providers.

5. Annexes and References

http://www.myocean.eu/

http://www.seadatanet.org/

JERICO Deliverable 5.1 Delayed Mode Data Management Handbook (Version1) - http://www.jerico-fp7.eu/deliverables/d51-dm-data-handbook

JERICO Deliverable 5.2 Real Time Data Management Handbook (Version1) - http://www.jerico-fp7.eu/deliverables/d52-rt-data-management-handbook

6. Acknowledgements

We would like to thank the OGS colleagues, A. Giorgetti and R. Nair, for their valuable support and constructive suggestions on some technical aspects of this report.