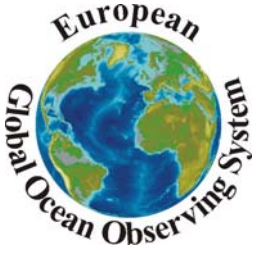


EuroGOOS- DATAMEQ-WG

EuroGOOS DATA Management , Exchange and Quality Working Group

Recommendation for a PAN-European data management system for operational oceanography within EuroGOOS

Ref: DataMEQ-01-2008
Date: December 2010
Version 2.1

AUTHORS : EuroGOOS DataMEQ working group members

Table of Content

Document Change Record

Author	Modification	Issue	Date
S Pouliquen	First Draft	0.1	July 2008
S Pouliquen	Included remarks from DataMEQ Members	0.2	01 Sep 08
S Pouliquen	Included remarks from EuroGOOS board	1.0	22 Sep 08
S Pouliquen	Accepted with no change by Eurogoos members at annual meeting in Galway	1.0	20 Oct 08
S Pouliquen	Update after the 2010 DATAMEQ meeting	2.0 2.1	Dec 10 Feb 11

DATAMEQ- WG members

Chair Sylvie Pouliquen
EuroGOOS Patrick Gorringe

BOOS and NOOS representatives
- BSH — Kai Soetje
- DMI — Vibeke Huess
- SMHI — Thomas Hammarklint
- UKMET — Simon Good

IBI-Roos representatives
- Ifremer — Thomas Loubrieu
- Puertos Del Estado — Marta de Alfonso

MOON representatives
- ENEA — Giuseppe Manzella
- HCMR — Leonidas Perivoliotis, Dimitris Kassis

Artic representative
- IMR — Helge Sagen or Sjur Ringheim Lid

Other program representatives
- Argo/Gosud/OceanSITES — Thierry Carval
- SeaDataNet — Gilbert Maudire
- Satellite — Frédérique Blanc
- ECOOP — Declan Dunne
- INSPIRE — Keiran Millard
- Ferrybox — Wilhelm Petersen
- ICES — Neil Holdsworth or Hjalte Parner

1. INTRODUCTION

EuroGOOS has set up the Data Management, Exchange and Quality working group with the following terms of references

1. *Develop an overall concept for the management of EuroGOOS observation data taking into consideration data management systems which are developing within GMES and JCOMM*
2. *Identify, in consultation with the EuroGOOS Task Teams and OOS/N, as appropriate, the type of observations which can be made available either in real-time or in delayed mode*
3. *Propose the most effective ways to make observation data readily available for operational purposes in a sustained matter*
4. *Propose mechanisms to ease access to delayed mode observation data in cooperation with NODCs, keeping aware of the progress in SeaDataNet*
5. *Draft a minimum set of standards for data quality control which is related to observation data collection, processing and exchange procedures*
6. *Each TT or OOS/N should appoint 2 persons to represent them in the working group and promote internal coordination.*

The DATA-MEQ chair will liaise with the JCOMM Data Management Programme Area coordinator.

The purpose of this document is to address the points 1-3-4-5 of the term of references based on the present development made at national, EuroGOOS/ROOSes and EU level.

2. WHY DO WE NEED A PAN-EU DATA MANAGEMENT SYSTEM FOR EUROGOOS ACTIVITIES?

The scope of the EuroGOOS ROOSes is wide and their needs will be partially addressed by the on-going development carried on within GMES, SeaDataNet or other EU initiatives. It is essential to improve the quantity, quality and accessibility of marine information for decision making and to open up new economic opportunities in the marine and maritime sectors of Europe, for the benefit of the European citizen and the global community.

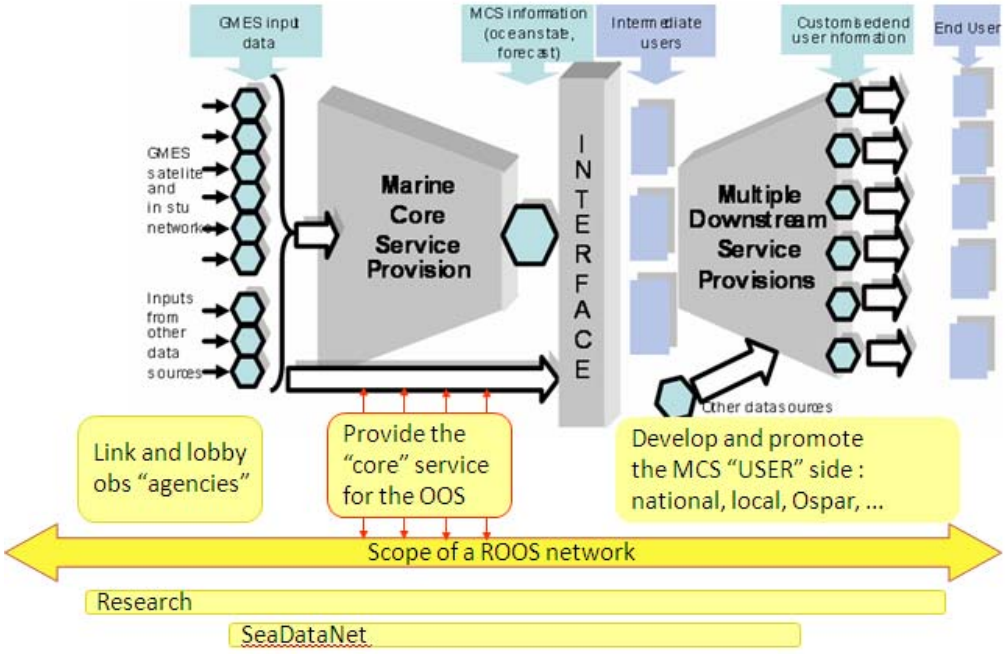


Figure 1 EuroGOOS ROOSes and GMES

To address this wide scope the ROOSes need

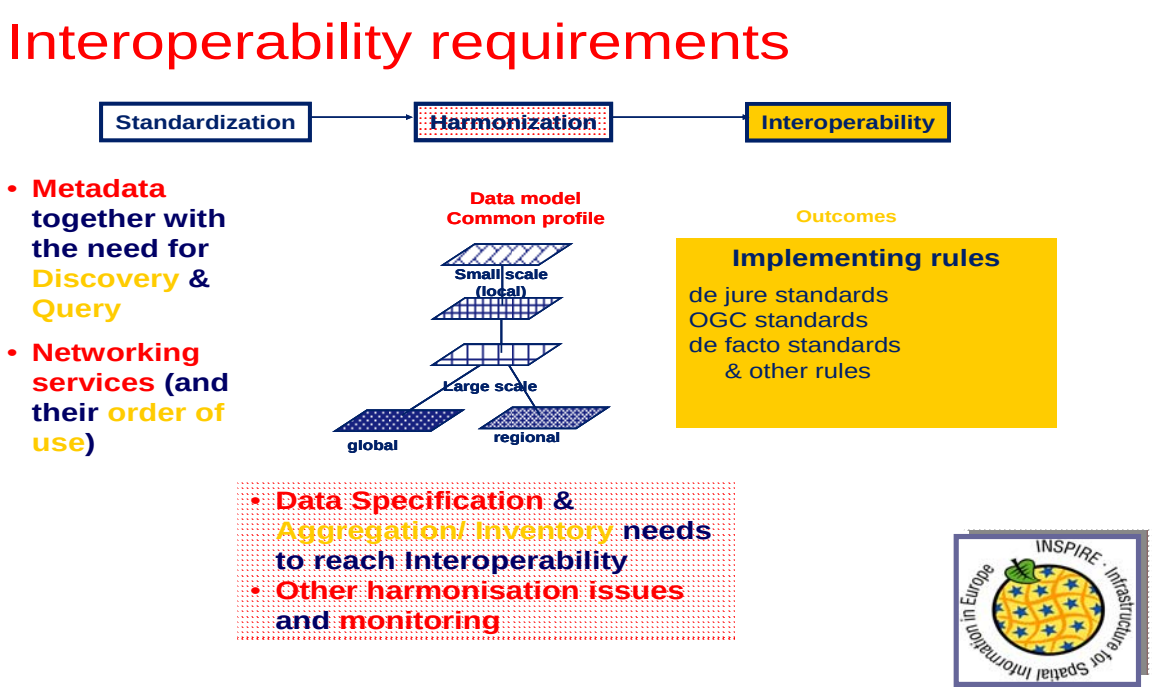
- To understand their requirement for data and know what data are available → this will help to improve the observation strategy and coordination
- To access data easily through standard generic tools: easily means using the data without having to be concerned about who processes them providing this is done in a coherent way and adequate metadata are available to know how they have been processed, .
- To combine in situ-observation data with other information (satellite or model outputs,..) in order to derive new products, build new services or to take decisions.

Building this Pan-European system will require input from various actors

- Data provider catalogues issued by operational agencies, research laboratories and national oceanographic Data Centres & Satellite Data centres (a catalogue provides the product portfolio, documentation on products, production plan and commitments to users) and index or inventories as tools to access these data.

- Users requirements: from operational forecasting centres, research community, EU bodies and the private sector setting up services

EuroGOOS will benefit of such a Pan –European data management system at various levels. Organising the data management and exchange process within EuroGOOS will reduce duplication of effort among agencies, improve quality and reduce costs related to geographic information, make geographic data more accessible to the public and increase the benefits of using available data. It will also help to establish key partnerships with states, counties, cities, nations, academia and the private sector to increase data availability. An operational EuroGOOS data management system will help to prevent imposition of a system from outside that will not fit our needs in terms of the use of standards and the tasks of organisations



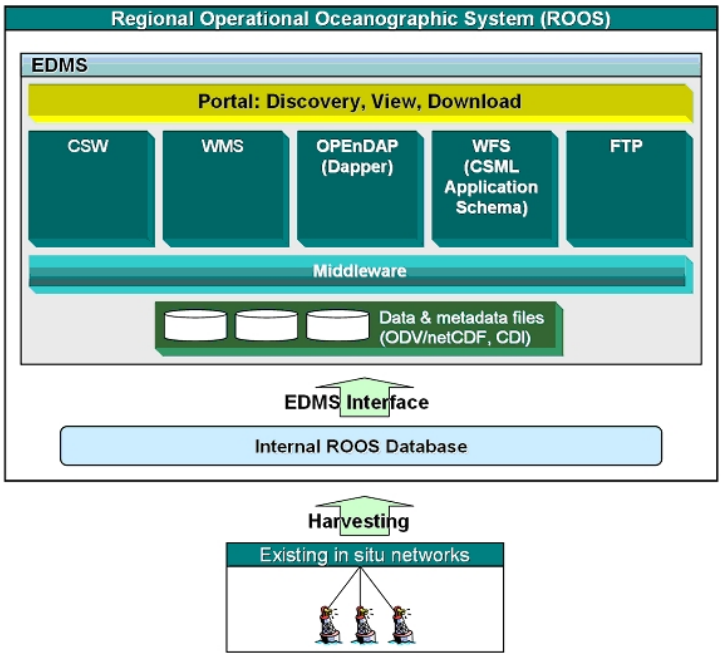
.

If we are not able to set up such a system the cost to meet the needs of the various EU directives and other national and EU requirements will continue to increase and we will rapidly loose advantage that we have gained by combining our efforts compared to other institutional bodies.

3. DESIGN OF A PAN-EUROPEAN DATA MANAGEMENT SYSTEM FOR EUROGOOS

The Pan–European data management system for EuroGOOS needs to provide a reliable, sustained access to integrated products on a limited number of variables critical for EuroGOOS ROOSes activities. The tasks have to be distributed in order to share the work load, avoid unnecessary duplication of data and benefit from the regional expertise of the data providers. It will provide a single system access point to European marine observation data including real-time and delayed mode data as well as facilitating the setting up of specialised portals to be built above this architecture.

This system will be one element of the GMES and international systems. It benefits from ongoing progress being made within the ROOSes, and within EU funded projects such as MERSEA, ECOOP, SeaDataNet and MyOcean. It will also benefit from progress made on standards and protocols at international levels through JCOMM, INSPIRE and EMODNET



European data management system (EDMS) installed in regional centres

4. WHAT IS AVAILABLE TO BUILD THIS SYSTEM ?

4.1. The ROOSes

Each ROOS has started data exchange activities to satisfy the needs they identified. Some have chosen to exchange data in real-time through ftp-boxes, others to set-up data centres that are aggregating data either according to the type of platforms that acquired the data or according to the variables that was measured, and others have decided to set up portal(s) for their ROOS. Therefore for most of the ROOSes the potential data providers for their activities have been identified and the willingness to share the data at ROOSes level at least has been established..

4.2. SEPRISE (Sustained, Efficient Production of Required information Services)

SEPRISE is an FP5 EU project coordinated by EuroGOOS office that demonstrated that data exchange in real-time was possible at European level. The focus was mainly on fixed point stations. No harmonisation on data format, metadata or quality control was handled by the project. This project showed that it was possible to unlock some access to data at EU level for operational oceanography purposes.

4.3. ECOOP (European Coastal Sea Observing and fOrcasting Programme)

ECOOP is an FP7 project led by DMI that aims to consolidate, integrate and further develop existing European coastal and regional observing and forecasting systems into a Pan-European system. Considering the data management system, it is designing a distributed system of ROOSes data portals. It will benchmark technology such as Dapper and WFS to see how it can handle the in-situ data to be used within information and decision systems by end users. It is also prototyping the regional portals using FTP at least or on voluntary basis OPeNDAP- DAPPER version, recommending data formats, product description and setting up an EU product catalogue and viewing service extending developments made within the FP6 MERSEA project.

4.4. SeaDataNet

SeaDataNet is an FP6 project setting up a Pan-European Infrastructure for Ocean and Marine Data Management based on a distributed infrastructure of National Oceanographic Data Centres. It will provide on-line access to integrated databases of standardised quality by using adapted communication & information technology. SeaDataNet is an important source of data for the EuroGOOS activities. Moreover SeaDataNet will provide standardised discovery, delivery and viewing services for water column parameters (physical, chemical, biology). It will maintain catalogues of data centres, observatories and datasets as well as common vocabularies, metadata and catalogues standards, quality control procedures as well as interoperability guidelines that should be used within EuroGOOS.

4.5. Motiive (Marine Overlays on Topography for Annex II Valuation and Exploitation) and MarineXML (Marine eXstensible Markup Language)

These two projects have worked closely with ISO TC211 and OGC (Open Geospatial Consortium) on exchange standards for measurement and observation data. This is important as the forthcoming INSPIRE (Infrastructure for Spatial Information in Europe) Directive will mandate data exchange based on these approaches. The two key building blocks in this process has been the OGC 'Observation and Measurements' recommendation that sets a generic data model for observation and the Climate Science Modelling Language that describes GML Application Schema for the result of an observation. Although Motiive and MarineXML had a marine focus, there are working closely with the international water and meteorological communities. This recognises the fact that the marine community will need to exchange data with these communities so common standards are important. The standards developed in Motiive are being testing as part of ECOOP. A WFS (Web Feature Service) has already been established to provide SmartBuoy data as CSML PointSeries through a Web Feature Server

4.6. MyOcean

MyOcean is an FP7 EU project , following up MERSEA FP6 project that demonstrated the feasibility of integrated operational oceanography system in Europe. It started early 2009 and delivers the GMES Marine Core Service (Left wing of the butterfly in Figure 1). It has set up Thematic Assembly Centres (TACs) that are essential elements of the operational oceanography infrastructure. These TACs provide data and products needed by the modelling and data assimilation systems (MFCs)) with monitoring of the production chains from external interfaces to product provision to users. In particular MyOcean has developed the In-Situ TAC that is a distributed system, coordinated at regional level for the 6 ROOSes involved within MyOcean that commit through OLA (Operational Level agreement) to deliver in-situ products the MyOcean MFCs. These OLA define the level of service and monitoring tools. MyOcean also implement an information system that provide discovery, viewing and downloading services for the MyOcean products. These services are developed taking into account interoperability with SeaDataNet, the methods demonstrated by ECOOP and other international recommendations. The data management system will serve the EuroGOOS community and also provide an infrastructure that can be extended to fulfil additional ROOSes' needs .

4.7. EmodNet

EmodNet aims at assembling in situ data into a coherent framework to allow new economic activities on the sea and in monitoring ecosystem health. As a result many valuable data are now inaccessible because of how, why and by whom they were collected and how and where they are now held. These datasets need to be unlocked and made easily accessible so that such records of change over time could begin to answer questions as yet undreamed-of. The "proof of concept" of EMODNET is being tested through preparatory actions. Portals for a number of maritime basins are being set up for hydrographic, geological, biological ,chemical and physical data as well as functional habitat maps. These portals will provide access to marine data of a standard format and known quality and identify gaps in coverage. The projects will identify the main challenges in moving from an ur-EMODNET to an operational EMODNET.

4.1. Standardisation bodies

There is emerging consensus that data exchange related to measurements and observations should be related to the sampling regimes that form scientific common practice. These regimes need to relate to interoperable data models that in term can be realised in a particular encoding. A common technology platform should be capable of delivering all these encodings.

The sampling regimes that form common practice include conceptualising data as Points, Profiles, Grids, Swaths, and etcetera. These are defined as "Coverages" according to the standard ISO 19123. Four main communities are engaged with providing encodings of these sampling regimes and associated tools:

- UNIDATA. Unidata are developing a common data model for points, grids, profiles and swaths to be implemented principally as NetCDF.
- CSML (Climate Science Modelling Language). CSML are developing a suite of data models for points, grids, profiles and swaths to be implemented principally as GML Application Schema (ISO 19136)
- EOS (Earth Observing System). EOS has developed a suite of data models for points, grids, profiles and swaths to be implemented principally as HDF.
- SensorML (Sensor Markup Language, Open Geospatial Consortium) . SensorML is describing a wide range of sensors, including both dynamic and stationary platforms and both in-situ and remote sensors.

Of course there are slight differences in each of these implementations, but the commonality of the approach means that interoperability between these three approaches should be straightforward. For example NetCDF4 is based on HDF (Hierarchical Data Format) and CSML is designed to 'wrap' NetCDF files. The biggest obstacle to EuroGOOS is that all three approaches are still evolving.

This problem of evolution extends to the systems that can be used to deliver and view data in these standards. As the NetCDF CDM (Common Data Model) is not fixed, no technology to deliver it exists, hence technology for well-established conventions such as NetCDF-CF with THREDDS (Thematic Realtime Environmental Distributed Data Services)are used. CSML can be delivered via a Web Feature Server, but there are few examples of doing this, especially for large datasets.

Related to the above is the work of the OGC with its Observation and Measurement framework. This is essentially a meta-model based on ISO standards that relates the 'what', 'how', 'when', 'who' of a measurement with its associated result (dataset) and dataset metadata. This work has recently been submitted to ISO TC211 as a candidate for standardisation.

At a European level the biggest driver for exchange standards will be the INSPIRE directive. This will specify standards for generic exchange of observation data. Whilst these will not be mandated for exchange within the oceanographic community, they will be for exchange from the oceanographic community to other communities, especially for national reporting objectives. This situation will be the same for other communities such as water and air. These exchange standards will be based on ISO TC211 standards and so will be similar to the approaches used in CSML.

5. HOW TO IMPLEMENT A PAN EUROPEAN DATA MANAGEMENT SYSTEM FOR EUROGOOS ROOSES

The key challenges are to organise a common data management approach to be adopted by all actors. It must engage and build upon existing initiatives to harmonise data management practices within EuroGOOS. The key elements of the harmonisation are:

- A common data policy enabling free and open access to data
- Common standards and protocols for quality control procedures, metadata formats and descriptions, and data exchange formats (ie clear definitions of interfaces)
- Common (OGC compliant) tools to share between partners to implement the infrastructure
- Development as a system of systems where interoperability between systems will be achieved by adopting the EU (INSPIRE) principles under development.

To be able to implement this data management system the elements, the interfaces, and the common functions to implement have to be defined. Standards and protocols have to be recommended to describe the quality of exchange data and products. Moreover a technological survey has to be maintained in order to update regularly this referential to benefit from enhanced standards that will emerge in the future.

5.1. Building a distributed data management system

Considering that the Pan–European data management system for EuroGOOS will provide a reliable, sustained access to integrated products on a limited number of variables critical for EuroGOOS ROOSes activities, the ROOSes have to set up regional portals extending what is benchmarked in ECOOP and consolidated/certified in MyOcean as basic infrastructure for EuroGOOS Data Exchange.

5.2. Recommendations on standards and their evolution

As the SeaDataNet scope is wider than EuroGOOS's, SeaDataNet is defining very complete vocabularies . A subset should be extracted from SeaDataNet to fit the variables relevant to EuroGOOS activities.

OceanSITES NetCDF, SeaDataNet ODV, should be used for data exchange between ROOSes' portals within EuroGOOS. These formats will be revised periodically by the DataMEQ working group to benefit from technical progress.

The DataMEQ working group will advertise the standards chosen by EuroGOOS and the way to implement them at JCOMM level and work on adopted standards from JCOMM.

The DataMEQ working group will continue the collaborative work started within ECOOP on standards to be able to influence their elaboration (implementation of OGC services, CSML, CSW, OPeNDAP, DAPPER,OceanTron THREDDS, format, …) . DataMEQ will help their deployment within the ROOSes and maintain a forum/wiki with guidelines to the ROOSes as well as the data providers .

5.3. Maintaining catalogues for EuroGOOS

SeaDataNet has developed and is maintaining a set of catalogues that are useful for the ROOSes but difficult to keep up to date at SeaDataNet level. The ROOSes meet once a year and can easily update the Fixed Observatory and data set catalogues.

Therefore, DataMEQ recommends the ROOSes to maintain a catalogue of their observing platforms (simplified EDIOS) that would be updated on regular basis and could be provided to SeaDataNet to maintain their central catalogues. A subset of metadata mandatory to discover observatories within EuroGOOS should be defined in collaboration with SeaDataNet and EmodNet-Physic projects.

DataMEQ recommends the ROOSes maintain a catalogue of their products and federate them into an EuroGOOS catalogue. A link with SeaDataNet EDMED catalogue is desirable. Product specification should include parameters, quality control (QC) procedures, update frequency, delivery services. The Product Portfolio should be provided by the EuroGOOS Product WG.

ROOSes catalogues should be ISO19139 catalogues to be compatible with SeaDataNet and JCOMM standards.

5.4. An interoperable Information system

To provide visibility to the products generated within EuroGOOS' ROOSes , INSPIRE technical services should be implemented and more specifically the following services:

1. the discovery service has to be able to locate rapidly products by browsing the catalogues described in previous paragraphs, and
2. the viewing services at regional level have to be homogeneous and OGC compliant.
3. Downloading service should be provided whenever possible based on common data distribution standards

These services should be developed enhancing the tools developed within MERSEA, ECOOP and MyOcean.