ODYSSEA - Operating a network of integrated observatory systems in the Mediterranean Sea

C. Figueiredo (1), I. Pedro (1), A. Silva (2), P. Galvão (2). J. Rodrigues (2) and P. Leitão (2)

(1) Edisoft,Sa. Rua Calvet Magalhães, 245, 2770-153 Paço de Arcos. www.edisoft.pt, carlos.figueiredo@edisoft.pt, ines.pedro@edisoft.pt

(2) HIDROMOD, Lda. R. Rui Teles Palhinha, 4, 1º, 2740-278 Porto Salvo.www.hidromod.com, adelio@hidromod.com, pedro.galvao@hidromod.com, joao.rodrigues@hidromod.com paulo.chambel@hidromod.com

Abstract: This paper presents an overview of the H2020 ODYSSEA project with special emphasis on data management related issues. The goal of ODYSSEA is to set up an interoperable and cost-effective platform integrating data from observation and forecasting systems across the Mediterranean basin, building on key initiatives such as Copernicus, GEOSS, GOOS, EMODNet, ESFRI, Lifewatch, Med-OBIS and others.

A particular capacity of the first release (summer 2018) described in the paper is the integration, fusion, standardization, transformation and dissemination as OGC standard services of different data sets from different providers, locally acquired in-situ data and results from local high resolution models.

Applying advanced algorithms to homogenize and fuse large quantities of data, the platform will provide derived data services on-demand through a single public portal to a wide range of end-user groups and stakeholders.

Key words: interoperability, data fusion, OGC standard services, SOS, WMS, WFS, WCS, CSW

1 INTRODUCTION

ODYSSEA is a H2020 project that aims to develop, operate and demonstrate an interoperable and cost-effective platform that fully integrates networks of observing and forecasting systems across the Mediterranean basin, addressing both the open sea and the coastal zone.

The platform is prepared to deliver a wide range of services focused on the needs of users involved in coastal management such as navigation safety, port operations, water pollution prevention and response, eutrophication risk management, search and rescue missions, etc. allowing to exploit the added value of integrated Earth Observation (EO) technologies (satellite, airborne and ground based), Copernicus Marine Service and ICT to deliver customized and ready to use information. These services will provide an easy way to get in-situ data, local high-resolution forecasts and products and services (e.g. meteo-oceanographic conditions at specific locations, identification of optimum or critical working windows, support to sea pollution response actions, etc.) to a broad range of different users.

In the following sections, we present an overview of the procedures that are being adopted concerning the following key data issues: data integration, data fusion and data standardization and dissemination.

2 DATA MANAGEMENT

Taking in consideration that the platform will gather a large diversity of data sets, the issues of data management and data quality control assume a central concern. One goal is to ensure that data from different and diverse providers are readily accessible and useable by a wider community. In order to achieve this, the strategy is to move towards an integrated data system within ODYSSEA that harmonizes work flows, data processing and distribution across different systems.

The value of standards is clearly demonstrable. In oceanography there have been many discussions on processing data and information. Many useful ideas have been developed and put into practice, but there have been few successful attempts to develop and implement international standards in managing data.

2.1 The European context

Several initiatives exist in Europe for ocean data management, which are now coordinated under the umbrella of EuroGOOS. EuroGOOS is the network committed to develop and enhance the operational oceanography capacity of Europe as contribution to the intergovernmental Global Ocean Observing System (GOOS). The scope of EuroGOOS is wide and its needs are partially addressed by on-going developments in the scope of Copernicus, SeaDataNet and other EU initiatives.

Therefore, in order to improve quantity, quality and accessibility of marine information, to support decision making and to open up new economic opportunities in the marine and maritime sectors of Europe for the benefit of European citizens and the global community, it was agreed at the annual EuroGOOS meeting in 2010 that it is essential to meet the following needs (AtlantOS, 2016):

- ✓ Provision of easy access to data through standard generic tools; where "easy" not only means the direct use of data without concerns on data quality and processing but also that adequate metadata are available to describe how the data were processed by the data provider.
- ✓ To combine in situ-observation data with other information (e.g., satellite images or model outputs) to derive new products, build new services or enable better-informed decision-making.

These recommendations aim to reduce the duplication of efforts among agencies, to improve quality and reduce costs related to geospatial information, thus increasing data availability and making oceanographic data more accessible to the public, whilst fostering partnerships among stakeholders in the domain.

Along the last years, marine data integration initiatives such as the Copernicus in-situ data thematic centre (for access to near real-time data acquired by continuous, automatic and permanent observation networks), SeaDataNet (for quality controlled, long-term time series acquired by all ocean observation initiatives, missions, or experiments), EMODnet and several others have being making efforts to implement some relevant standards expected to facilitate the discovery and sharing of data among different users. These initiatives support both data providers, willing to share their observation data, and users, requesting access to oceanographic data (historic, real-time and forecasts).

ODYSSEA will follow similar guidelines, learning from these experiences and make improvements where they are required.

2.2 Data format and standards

Delivery of data to users requires common data transport formats which interact with other standards (vocabularies, data quality control). In practice, two major data types will be addressed: gridded data and in situ data. These data types are implemented in datasets from CMEMS (dynamic gridded data), EMODNet (static gridded data) and SeaDataNet (in situ data).

2.2.1 In situ data

In situ data include data from point time series, profiles, time series vertical profiles and trajectory profiles. For these, SeaDataNet procedures will deliver the main guidelines and NetCDF-CF will be the standard format to be adopted (SeaDataNet extension of the Climate and Forecast, 2018). Moreover, ODYSSEA will go one step further and use harmonized NetCDF files to feed an OGC SOS v2 service supported by the 52ºNorth software package in order to provide an interface to the users, via the Portal GUI. Additional interfaces with time series data will be implemented later by considering the usage of the OGC WFS protocol (GeoServer mapping engine) and download of NetCDF-CF files via FTP and HTTP protocols.

The 52º North SOS software implements the notion and concept of Observation which is based on the OGC/ISO Observation & Measurements model. We will briefly describe below the most relevant terms used by the implementation of the 52º North software package. An observation is the act of observing a phenomenon, that we call observable property. In other words, it is a measurement of a parameter's value, in case the latter is numeric. Furthermore, we say that an observation is registered by a procedure (e.g. tidal gauge) at a given location, whose term is feature of interest, and at a given time, called phenomenon time. Finally, the SOS allows for observations to be grouped in offerings, according to a certain criteria that the user may define. All these terms can be linked through a time series, uniquely characterized by the combination of procedure, offering, observable property and feature of interest, and that consists of a set of observations. The schema in Figure 1 summarizes and relates these concepts using database tables from the SOS containing only the most important columns (52°North Sensor Observation Service 4.x Database model).

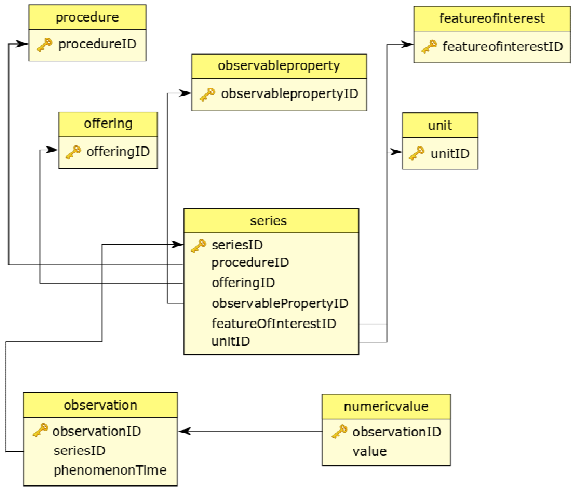

Figure 1 - Simplified SOS Database model

In summary, we have a harmonized format in which all data is stored, and an interoperable interface for data publication and consulting. Therefore, we must define a mapping between the former and latter

components, in order to guarantee the data flow from the data producer to the data consumer.

The SeaDataNet extension of the NetCDF CF convention consists of a set of:

- ✓ dimensions: INSTANCE, corresponding to the number of datasets, and MAXT, in the case of time series, which is the total number of time steps;
- ✓ variables, which may be coordinate (contain the spatial-temporal information), geophysical (measured parameters) or ancillary (as support), where each variable is characterized by a set of attributes (e.g. units);
- ✓ global attributes, concerning the entire file.

In the harmonized NetCDF file, there is a global attribute that points to the dataset metadata, in SensorML format, from which we can extract the procedure information to be inserted in the SOS, through the *InsertSensor* operation. For now, we will simply identify the data offering with the procedure, but one can organize the data as desired. The spatial coordinate variables, usually latitude, longitude and depth are mapped to the feature of interest, whereas each geophysical variable corresponds to an observable property and contains the corresponding unit. These five components constitute a series whose template can be inserted in the SOS using the *InsertResultTemplate* operation. Finally, multiple observations can be inserted "in batch" using the *InsertResult* operation, where the phenomenon time and (numeric) value of each observation correspond to the respective values of the temporal coordinate and geophysical variables.

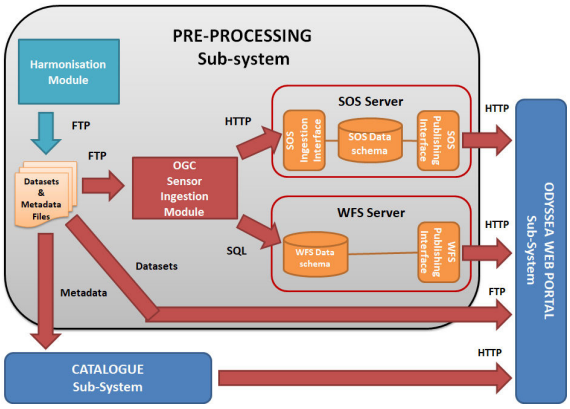

Figure 2 - Pre-processing sub-system

The workflow is split into 4 phases. In Phase I, all in situ data are ingested by the harmonization module which standardizes all incoming data. For Phase II, the user data provider is liable to contribute with descriptive information about the measurements, equipment used, sensor information and contacts. All this information is used to harmonize the data following Open Geospatial Consortium (OGC) directives for services and metadata.

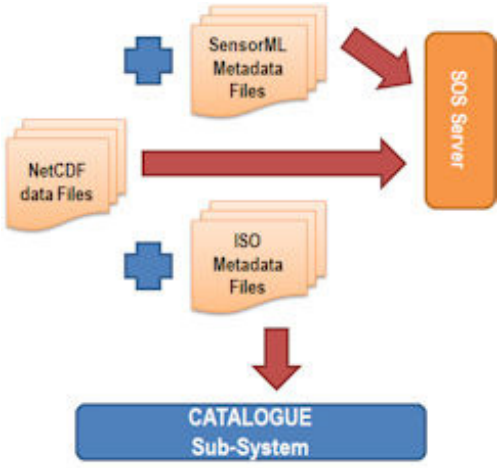

Figure 3 - SOS and WFS Database

In Phase III, the system performs three steps: the metadata produced by the data provider, following the European INSPIRE Directive, is published in the ODYSSEA catalogue. The processes executed in this phase create a link between the metadata and the data to be ingested; it also produces and publishes the data in Sensor Model Language (SensorML) and creates the SeaDataNet NetCDF (CF) using the NEMO Software. These files are ingested by an OGC module that reads the data and adds it to the SOS and WFS services in Phase IV.

2.2.2 Static gridded data

ODYSSEA is adopting the recently proposed SeaDataNet standards for marine chemistry (to support the EMODNet Chemistry pilot), bathymetry (to support the EMODNet Hydrography and Seabed Mapping pilots), geology and geophysics (to support the Geo-Seas project and the EMODNet Geology pilot) as well as marine biology.

2.2.3 Dynamic gridded data

For dynamic gridded data, the CMEMS standards will be followed. The files will be available in NetCDF CF format using the coordinated system WGS 84 (EPSG:4326). Access to the data is done via Motu, a robust web server which fills the gap between heterogeneous data providers and end users. Motu handles, extracts and transforms huge volumes of oceanographic data without performance degradation, and offers a web service interface which implements the OGC WCS standard.

2.3 Metadata, vocabulary and catalogues

The issue of metadata, vocabulary and catalogues is of primary importance for interoperability and easy discovery of data.

All files will conform to metadata conventions (http://cf-pcmdi.llnl.gov/) designed to promote the processing and sharing of files, and will make use of

the NetCDF CF API. These conventions define the metadata that describes the data with respect to each variable as well as the spatial and temporal properties of the data. This enables users from different oceanographic sub-domains to assess data comparability and facilitates the building of powerful applications with extraction, re-gridding, and display capabilities.

All data sets available on the ODYSSEA Platform have associated metadata supporting interoperability in accordance with the European INSPIRE Directive. In addition to this, for in situ data, SensorML is applied as an identification card of the sensor used in the observation.

One important prerequisite for consistency and interoperability is the use of a common vocabulary in all meta-databases and data formats. ODYSSEA follows the NERC Vocabulary Server (NVS2.0) hosted and managed by the British Oceanographic Data Centre (BODC).

The metadata information describing ODYSSEA services and products will be maintained in the catalogue application GeoNetwork, allowing the users (via GUI) and other systems (via OGC CSW Protocol) to search resources using temporal and spatial queries. GeoNetwork also provides the means to execute automated harvesting requests to other catalogues systems collecting relevant metadata records for ODYSSEA.

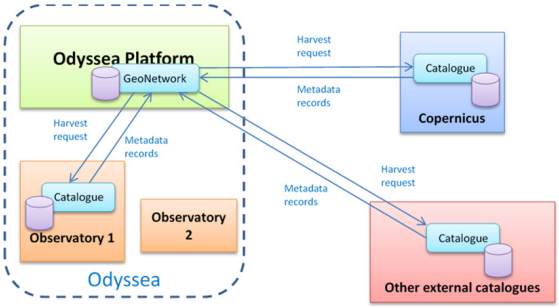

Figure 4 - ODYSSEA Catalogue

2.4 Data integration and fusion

Data integration and fusion are focused on improving the quality and robustness of the atmospheric boundary conditions of catchments/sewer (e.g., precipitation and evapotranspiration fields), hydrodynamic (momentum and heat fluxes) and wave (momentum fluxes) forecast models. As a standard procedure, high resolution local/regional prognostic models (e.g. WRF) will be downscaling forecasts produced by large scale solutions (e.g. ECWMF, NOAA GFS).

Additionally, the time series forecast service will provide some parameters (e.g. significant wave height, wind speed, air temperature) and put together forecasting solutions, allowing by this way to present graphically the uncertainty associated with the forecast models.

2.5 Data quality control

The data quality control has two sub-components: one focusing on the quality of observed in-situ data (e.g. tidal gauges, wave buoys, weather stations, etc.) and another in the modelled forecasts. In the first case, when uploading the data, the user will choose the validation performed before submission of the data. According to SeaDataNet criteria, each time observation must be classified from 0-9, corresponding to no quality control, good value, probably good value, probably bad value, bad value, changed value, value bellow detection, value in excess, interpolated value and missing value, respectively.

In the second component, model gridded data will be automatically compared with observations (e.g., buoys and CMEMS grid observation products) and the statistical analysis will be made available on a daily basis to the end users.

Acknowledgements

The ODYSSEA project is receiving funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N. 727277.

REFERENCES

AtlantOS, 2016. *Data Management Handbook*

Geoserver mapping engine, http://geoserver.org/

GeoNetwork, https://geonetwork-opensource.org/

SeaDataNet, 2010. *Data Quality Control Procedures*

SeaDataNet, 2017. *Data File Formats: ODV, Medatlas, netCDF. Deliverable D8.5*

SeaDataNet. (2018). *SeaDataNet data file formats: Data Transport Formats manual*

52ºNorth SOS project, https://52north.org/software/software-projects/sos/

52°North Sensor Observation Service 4.x Database model, https://wiki.52north.org/SensorWeb/SensorObservationServiceDatabaseModel

52º North SOS 4.x Documentation, https://wiki.52north.org/SensorWeb/SensorObservationServiceIVDocumentation