UTM-CSIC Data Service Architecture: from acquisition on-board to final dissemination

Juan Luis Ruiz, UTM-CSIC (Spain), jlruiz@utm.csic.es
Susana Diez, UTM-CSIC (Spain), sdiez@utm.csic.es
Xavier Rayo, UTM-CSIC (Spain), xrayo@utm.csic.es
Guillermo Garriga, UTM-CSIC (Spain), ggarriga@utm.csic.es

Overview

The Marine Technology Unit of the Spanish Research Council (UTM-CSIC) is responsible for the management of oceanographic research vessels. Within the UTM, the Data Service is in charge of managing spatial data acquired onboard oceanographic cruises and make these data as interoperable as possible in order to disseminate them so that they can be used and reused.

To achieve this goal an information architecture has been developed to control all the processes of data workflow described below. The data model can be consulted under http://data.utm.csic.es/lib/datamodel/

1. Data acquisition and events registration on-board

Underway data acquired by different vessel systems and sensors (position, heading, vessel speed, depth, weather, water temperature, salinity and other water physical parameters) are sent to the Continuous Oceanographic Data Acquisition system (CODA) which integrates them. The result is sent in real time to the next phase of the system, the Data Hub, and also in non-real time so that there is no loss of data in case there is a problem with the real time.

The Eurofleets Automatic Reporting System software (EARS) allows scientists and technicians working on research vessels to register events that take place during an oceanographic cruise, key aspect for the generation of metadata. Theses events are merged with some of the data acquired in CODA and sent in non-real time to the Data Hub.

2. Data Hub and web services on-shore

On-shore UTM Data Hub concentrates data and metadata of CSIC fleet through satellite connections using different communication protocols in real time (UDP) and non-real time (email, rsync ...) with the aim of feeding dissemination services and tools. In addition, the hub harvests data and metadata from other centers in order to be standardized and disseminated to portals of international projects such as Eurofleets or Seadatanet.

This information is stored in files offered through two types of web services, feeding end user applications. Those related to metadata, such as *getCruise* to retrieve cruise information and *getEvent* to retrieve events registered using EARS, and those involved in the management of time series of continuous acquisition parameters, such as *getLast* with the last registered datagram, *getSerie* that offers all the records between two dates and *getPos* that offers the position for a given time.

3. UTM-CSIC Data Portal. Data dissemination I

The UTM-CSIC Data Portal gathers a series of tools and services for data and metadata dissemination:

Metadata catalogue: Catalogue of Oceanographic Cruises (with more than 600 cruises) and related datasets, carried out since 1991 on research vessels managed by CSIC. The basic metadata of each oceanographic cruise is completed with documentation (cruise plan and report) and online resources such as DOI citation, cruise map and plot and download of underway data.

Geoportal: Web portal with an interactive display that allow users to navigate through a map to search and query datasets. Users are allowed to add their own data and connect to external WMS services in order to create their own maps. The query and search results are also linked to the metadata catalogue.

Quality Control Panel and Underway Plot: Web application to represent graphically the underway datasets acquired during Oceanographic cruises. Data graphics and maps are merged interactively and the user can navigate using the map or the graphics. Data quality control can be done from this application from both our served data or from local file from users.

Real time panel: Panel showing real time underway data acquired by different vessel systems and sensors managed via CODA (position, heading, vessel speed, depth, weather, water temperature, salinity and other water physical parameters).

SOS Client: Similar to the underway plot, in this case the SOS client shows historical series from underway data following the standards of SOS services.

Data Download: All raw data acquired during oceanographic cruises are available for downloading through the Data Download portal by registration, according to the privilege of each user. In case of restricted data, a formal request is required and cases are studied independently according to the current convention.

4. European Data portals. Data dissemination II

The UTM is a National Oceanographic Data Center (NODC) of the Distributed European Sea Data Infrastructure *SeaDataNet* (SDN). *Cruise Summary Reports (CSR)* are provided through a dedicated catalogue to be harvested by SDN and *Common Data Index (CDI)* through the Replication Manager and Import Manager, both developed in SeaDataCloud H2020 project. The related unrestricted data are available in the cloud to be downloaded through the SDN Portal.

During Transnational Access **Eurofleets** cruises on board the RV Sarmiento de Gamboa the Data Service provides the "European Virtual Infrastructure in Ocean Research" Eurofleets portal (EVIOR), with underway data and events information, sailing tracks and current position and metadata.