A step toward data interoperability: NetCDF metadata comparative analysis in RITMARE Italian Project

Elena Partescano, OGS (Italy), epartescano@inogs.it
Alessandro Crise, OGS (Italy), acrise@inogs.it
Alessandra Giorgetti, OGS (Italy), agiorgetti@inogs.it
Lorenzo Corgnati, CNR - ISMAR La Spezia (Italy), lorenzo.corgnati@sp.ismar.cnr.it
Carlo Mantovani, CNR - ISMAR La Spezia (Italy), carlo.mantovani@cnr.it
Vega Forneris, CNR- ISAC (Italy), v.forneris@isac.cnr.it
Cristina Tronconi, CNR- ISAC (Italy), c.tronconi@isac.cnr.it
Massimiliano Drudi, INGV (Italy), massimiliano.drudi@ingv.it

Introduction

The Ritmare (La RIcerca ITaliana per il MARE) project has been structured around the following three objectives: to support integrated policies for the safeguard of the environment (the health of the sea); to enable sustainable use of resources (the sea as a system of production); to implement a strategy of prevention and mitigation of natural impacts (the sea as a risk factor) (from http://www.ritmare.it).

In detail, the Sub-Project 5 (SP5) aims to strengthen and credit the Italian observing system in Europe by integrating, as far as possible, various national scientific, industrial and institutional components (especially ISPRA and Civil Protection) and also creating infrastructure edge products, offering at the same time to the domestic industry the opportunity to create new products and expand into new markets. Additionally, the subproject defines requirements, methodologies and technologies that ensure data quality and significance, defined as the ability to represent and interpret ecosystem characteristics and changes, and to analyze its trends (from http://www.ritmare.it).

To guarantee free and open access to data and metadata acquired by the observation sites, SP5 challenge has been to develop a system integrating and giving access to data acquired by satellites, moorings, radars, models, gliders and floats, trying to identify a technological solution that guarantees the interoperability between data collected by heterogeneous platforms.

The data used in SP5 derive from different marine communities and the data format used is mainly NetCDF format, but the conventions used are different (mainly Ocenasites 1.2, CF-1.6 and CF-1.4).

This format guarantees, by integrating different standard protocols and interfaces such as OpenDAP, WMS, WCS, an easy access and sharing of array-oriented data.

As a first step, we collected differentThredds Servers used as end-points to share data: gliders, floats and moorings (http://insitu-ritmare.cineca.it/thredds/catalog/Ritmare/catalog.html) developed in collaboration with CINECA; satellites and HF and X band radar (http://ritmare.artov.isac.cnr.it/thredds/catalog.html) made available by CNR-ISAC and CNR-ISMAR.

Then, we decided to investigate the different NetCDF formats and in particular to identify the common metadata used, to define a common set of metadata to guarantee an efficient data discovery.

In the case of radar data, the activities carried on in the RITMARE project led to the definition of data and metadata interoperable formats and QA/QC procedures to be adopted as European standards for coastal radar applications, thanks to the participation in the EuroGOOS HFR Task Team.

Comparative analysis

The first step was to collect information about each type of NetCDF (radar, satellite, glider, float, mooring and model). In particular, for every kind of NetCDF we collected information about: identity of the collating center, the data provider, the data type, the NetCDF version and convention, the project involved, the metadata standards, the list of standard services activated, the endpoint, availability, comment and suggestions. The second step was to analyse the set of metadata used, by all the communities taking in consideration diverse convention, mainly Ocenasites 1.2, CF-1.6 and CF-1.4.

As a result, we proposed 11 metadata as mandatory and common to all NetCDFs, answering questions such as what, where, when, how and who acquired data, to permit an efficient data discovery.

The last part of this work was to compare the main metadata catalogs used by SeaDataNet and EMODNet European marine community, the CDI (Common Data Index) with the metadata included in the NetCDFs used in Operative Oceanography (Coriolis, EuroGOOS and Argo). From this comparison we identified 10 common metadata, 6 fields used into CDI but absolutely absent in all kinds of NetCDF and 7 fields used in CDI and discontinuously in NetCDFs.

Another relevant element is that CDIs uses SeaDataNet convention (BODC vocabularies and EDMO catalogs), while the NetCDFs use Ocenasites 1.2, CF-1.6 and CF-1.4 conventions. This leads to the need of a format converter to compare the information coming from different metadata catalogs.

Conclusion

The aim of this work was to identify a technological solution that permits a data discovery using common metadata, present in data acquired by different platforms using different data conventions (Ocenasites 1.2, CF-1.6 and CF-1.4). The first step was to collect different end-points to share data: (http://insitu-ritmare.cineca.it/thredds/catalog/Ritmare/catalog.html) developed in collaboration with CINECA and (http://ritmare.artov.isac.cnr.it/thredds/catalog.html) developed by CNR-ISAC and CNR-ISMAR.

The second step was to analyze the metadata used in different marine conventions and identify a common set of metadata. After this, we compared the NetCDF metadata used in Ritmare project with the major marine metadata catalogs such as CDI.

The future step will be to find a technological solution, such as Geonetwork, that permits to link to THREDDS Data Server allowing an efficient data discovery directly on NetCDF files.

Acknowledgment

This research was supported by RITMARE project. We thank our colleagues from CNR-IREA, who provided expertise and assisted the research, and from CINECA for their technical support.