



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

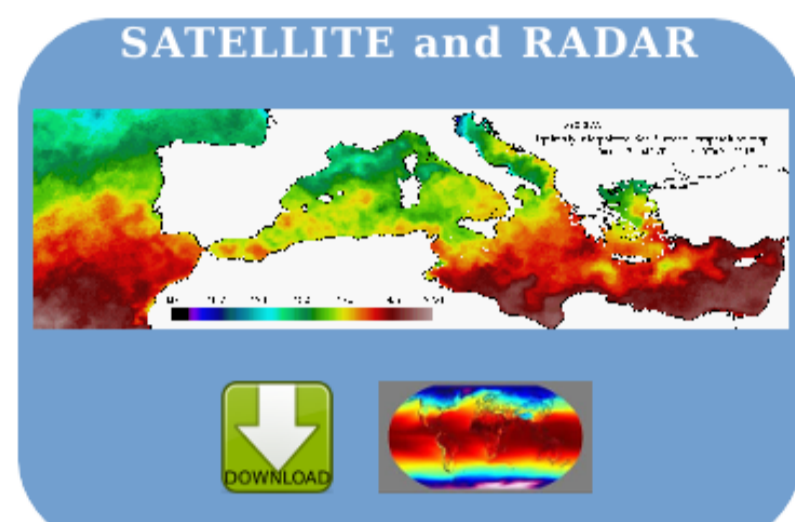


Elena Partescano, Alessandro Crise and Alessandra Giorgetti, OGS (Italy),
Lorenzo Corgnati and Carlo Mantovani, CNR - ISMAR La Spezia (Italy),
Vega Forneris and Cristina Tronconi, CNR- ISAC (Italy),
Massimiliano Drudi, INGV (Italy)
Contact: epartescano@ogs.trieste.it

INTRODUCTION



RITMARE SP5
DATA PORTAL SISTEMI OSSERVATIVI



<http://sp5-ritmare.cineca.it>

The Sub-Project 5 (SP5) of RITMARE (La Ricerca Italiana per il MARE) project aims 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 (<http://www.ritmare.it>).

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 Ocnasites 1.2, CF-1.6 and CF-1.4).

As a first step, we collected different Thredds 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.

COMPARATIVE ANALYSIS

1 - We collected information about each type of NetCDF (radar, satellite, glider, float, mooring and model).

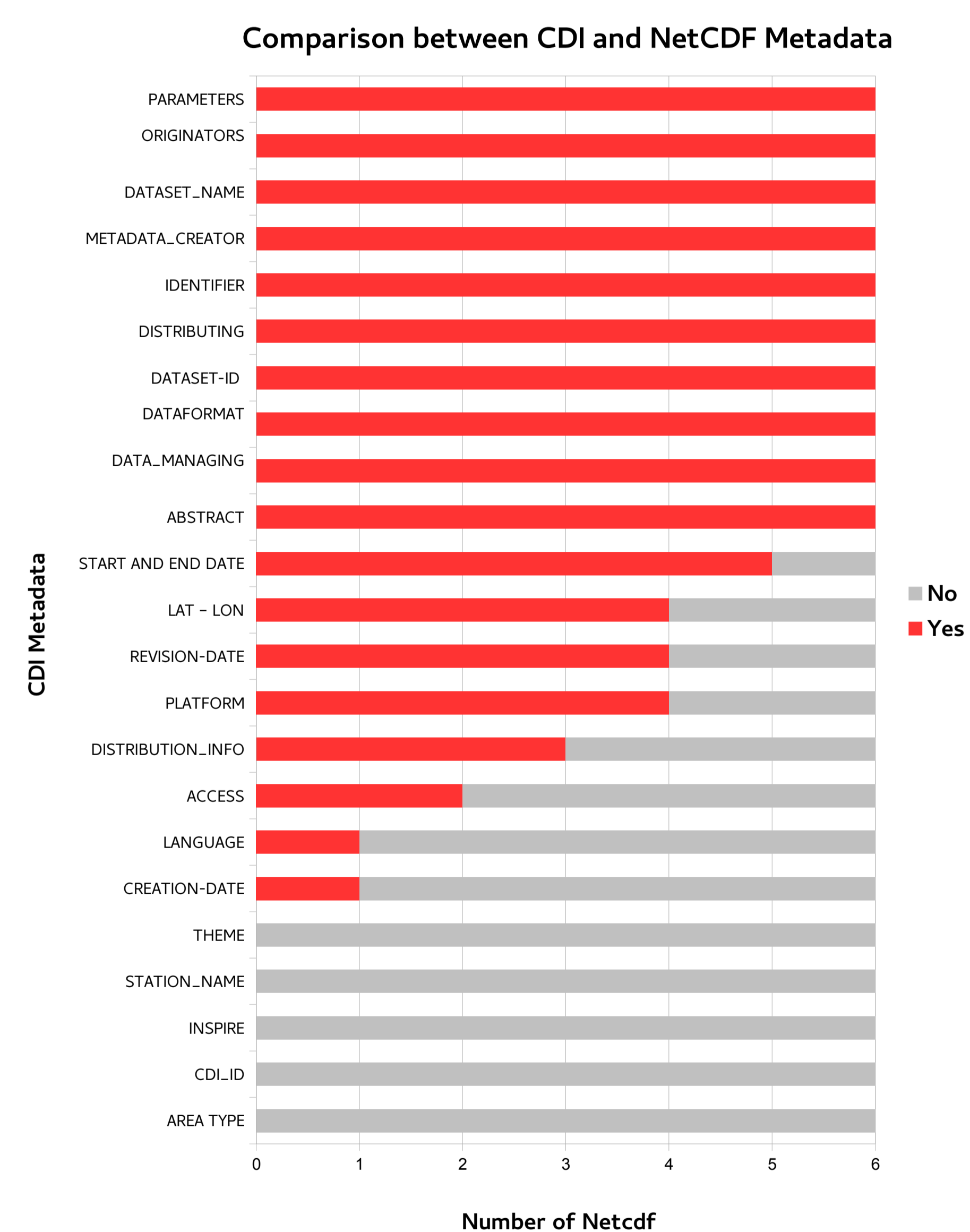
In particular, for every kind of NetCDF we collected information about: identity of the collating center, the data provider, the data type, the NetCDF version and convention, the project involved, the metadata standards, the list of standard services activated, the endpoint, availability, comment and suggestions. The second step was to analyse the set of metadata used, by all the communities taking in consideration diverse convention, mainly Ocnasites 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.

		Satellite	Radar	Models	Mooring	Glider	Float
WHAT	Data set name	✓	✓	✓	✓	✓	✓
	Data format	✓	✓	✓	✓	✓	✓
	Abstract	✓	✓	✓	✓	✓	✓
	Variables	✓	✓	✓	✓	✓	✓
WHERE	Lat/Lon/Region		✓		✓	✓	✓
WHEN	Start/End-Date	✓	✓	✓	✓	✓	✓
HOW	Datatype	✓	✓	✓	✓	✓	✓
WHO	Originator (institute)	✓	✓	✓	✓	✓	✓
	Originator (contact)	✓	✓	✓	✓	✓	✓
	Endpoint	✓	✓		✓	✓	✓
OTHER INFO	Other Info	✓	✓	✓	✓	✓	✓

2 - We compared 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, 5 fields used into CDI but absolutely absent in all kinds of NetCDF and 8 fields used in CDI and discontinuously in NetCDFs.



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 (Ocnasites 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.

Then, we analyzed 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.

