Improving the Copernicus Climate Data Store metadata scheme with the RDA metadata standards repository

The challenge

The RDA Adoption project is part of a wider C35_512 initiative, which is concerned with evaluation of the quality and the provision of quality assurance information to users of the data stored in the Copernicus Climate Data Store (CDS). The CDS repository contains several hundred terabytes of various types of data (satellite and in-situ observations, climate model outputs from many types) for which we present data quality summaries and recommendations at the level of variable and dataset.

Apart from providing recommendations on metadata standards, the biggest challenge we have encountered was the elaboration of metadata Key Performance Indicators (KPIs) for the Copernicus Climate Change Service (C3S) data and metadata services. One of these KPIs was developed following the recommendations of RDA. This particular KPI involves the monitoring of INSPIRE compliance of the data discovery services, which for this project is based on the Open Geospatial Consortium (OGC) Catalog Service for the Web (CSW).

The RDA outputs adopted

The conventions for CF (Climate and Forecast) metadata are designed to promote the processing and sharing of files created with the NetCDF API. These conventions are gaining acceptance and have been adopted by a number of projects and groups as a primary standard.

The conventions define metadata that provide a definitive description of the data represented in each variable, and the spatial and temporal properties of the data. This enables users of data from different sources to decide which quantities are comparable, and facilitates the building of applications with powerful extraction, regridding, and display capabilities.This is the link to the RDA metadata repository for CF metadata can be found here on GitHub. http://rd-alliance.github.io/metadata-directory/standards/cf-climate-and-forecast-metadata-conventions.html

The Common Information Model, a metadata standard used by the climate research community and presented in the RDA metadata standard repository (http://rd-alliance.github.io/metadata-directory/standards/cim-common-information-model.html) has also been used as a recommendation in the CDS, especially regarding the Coupled Model Intercomparison Project (CMIP) climate models.

The adoption process

The adoption process began at the same time of the Adoption grant (October 2019) and it is almost complete. In addition to the adoption of some of the standards available in the RDA metadata repository, we've added some extra details on tools (CF checker, CMOR, etc.) and extensions of well-known metadata profiles (World Meteorological Organization (WMO) Core Metadata Profile: https://wis.wmo.int/2006/metadata/WMO%20Core%20Metadata%20Profile%20(October%202006)/documentation.htm, INSPIRE directive metadata regulation: https://inspire.ec.europa.eu/).

Our team is composed by four people, focused on climate data and metadata. We work following an agile methodology (SCRUM) where we define tasks and plan them in three-week sprints. This iterative process allows us to dynamically plan and adapt to deviations on the masterplan.

Find out more at:

www.rd-alliance.org/recommendations-outputs

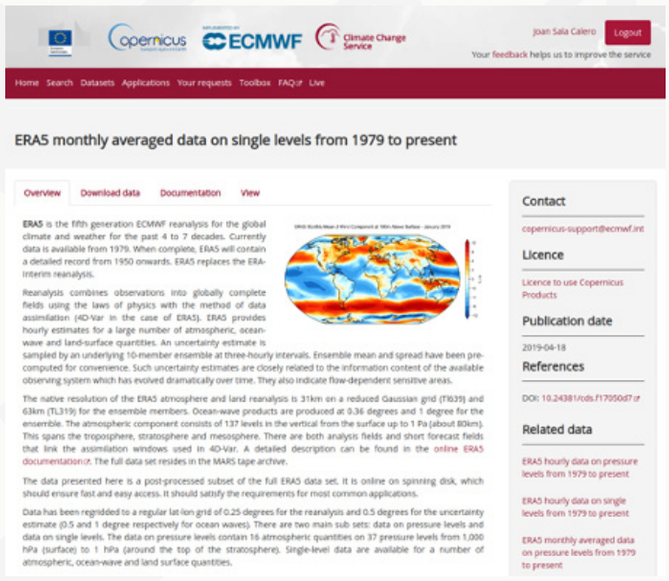
Visit rd-alliance.org

or write us at enquiries@rd-alliance.org

The impact of the adoption

The end beneficiaries of this work will be the users of the Climate Data Store as they will be able to get robust information on which metadata standards are included in the data they are using, making their usage easier and increasing their trust in the data provided. The second beneficiaries will be the CDS managers who, thanks to the work done in this proposal, will get a more trusted and documented data repository.

The impact of the work developed in the grant is significant since the data service has around **35,000 registered users** from over **60 countries** with presence in all continents. The data volumes are also significant since the service provides around **50TB of daily data** to the users.



Example of user interface for CDS record

Lessons learned

Due to the variety of meteorological data inside the CDS, we realized that we needed to start with the standardization and data checks of the most relevant data (Seasonal forecasts and Reanalysis). We will extend our experience to the other datasets in a next phase. The objectives of the C3S_512 tender fitted very well with the RDA Europe adoption grant, meaning that we had all the resources available to perform our tasks.

About Copernicus

Barcelona Supercomputing Center - Centro Nacional de Supercomputación (BSC-CNS) is the national supercomputing centre in Spain. We specialise in high performance computing (HPC) and manage MareNostrum IV (and soon V), one of the most powerful supercomputers in Europe. With a total staff of more than 650 R&D experts and professionals, our research focuses on four fields: Computer Sciences, Life Sciences, Earth Sciences and Computer Applications in Science and Engineering. Within the Earth Sciences department, the Computational Earth Sciences (CES) group we develop, manage and maintain a common data service framework to collect, standardize and distribute climate and atmospheric data.

The research leading to these results is part of the Copernicus Climate Change Service (C3S) (Framework Agreement number: ECMWF/COPERNICUS/2018/C3S_512_BSC), a programme being implemented by the European Centre for Medium-Range Weather Forecasts (ECMWF) on behalf of the European Commission.

Contact: Joan Sala Calero

Find out more at:

www.rd-alliance.org/recommendations-outputs

Visit rd-alliance.org

or write us at enquiries@rd-alliance.org