

DESIGNING COHERENT NETCDF FORMAT FOR JCOMM IN-SITU NETWORKS

T. Carval, S. Pouliquen and Argo, Gosud, OceanSITES Data Management Committees Team
<http://www.argo.net> <http://www.oceansites.org> <http://www.gosud.org>

Argo, Gosud and OceanSITES data management

In 2001, when we started designing the Argo Data management architecture for **profiling floats data**, there were not any on the shelf data framework that we could have used to encode Argo data.

However, **NetCDF** was emerging with a variety of scientific software able to manipulate them. An implementation of NetCDF was designed and documented in Argo User's Manual.

BODC/ GF3 vocabulary was meeting our requirement for temperature, salinity and oxygen observations. We could agree pretty rapidly on a common vocabulary/units and set of attributes.

In 2003, within GOSUD, for **underway data** acquired from vessels, we had to address the same issue, with a bit more variable, it seemed clear to the GOSUD data management team that format and standard used needed to be coherent with Argo as the parameters measured and used jointly by the operational oceanography community as well as the research.

In 2005 the same issue was raised within the OceanSites data management team for **time series reference sites**. The problem was then a bit more complicated as the number of parameters was more important and was mixing both ocean and atmospheric variables, profile and time series data as well as multidimensional data such as currents from ADCP.

We decided then to approach the **CF community** and together with them we extended the Argo/Gosud format to meet OceanSites requirements and identified actions to upgrade Argo and Gosud format to be CF compliant in order to benefit from the tools designed for CF compliant files. Moreover within the EU FP6 SeaDataNet project collaboration started between CF and GF3 community to converge of vocabularies and implement gateways between the two vocabularies.

Therefore we have presently with OceanSITES format a framework able to handle other oceanographic dataset that is pretty handy for users.

We plan to use this format within the MyOcean project to **integrate in-situ data at an European level**.

NetCDF and CF conventions

NetCDF (network Common Data Form) is a set of software libraries and machine-independent data formats that support the creation, access, and sharing of array-oriented scientific data.

The Argo, Gosud and OceanSITES data management architectures are based on an implementation of NetCDF and its main conventions : CF and COARDS. The 3 User's manual describe the NetCDF implementation :

- <http://www.ifremer.fr/coriolis/cdc/argo/argo-dm-user-manual.pdf>
- <http://www.ifremer.fr/gosud/doc/gosud-dm-user-manual.doc>
- http://www.oceansites.org/docs/oceansites_user_manual_ver_1_1.pdf

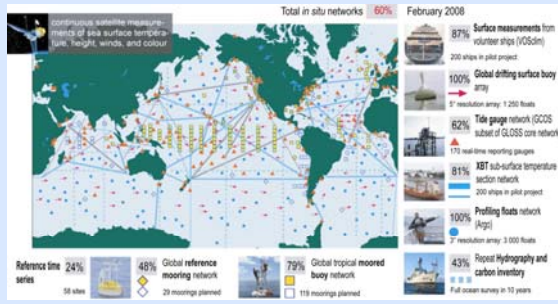
As explained on Unidata NetCDF web site, the conventions for climate and forecast (CF) metadata are designed to promote the processing and sharing of files created with the NetCDF API. The CF conventions are increasingly gaining acceptance and have been adopted by a number of projects and groups as a primary standard. The conventions define metadata that provide a definitive description of what the data in each variable represents, 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 building applications with powerful extraction, regridding, and display capabilities.

The CF conventions generalize and extend the COARDS conventions.

For more information on CF, COARDS, NetCDF, Udonits, and ISO8601 see:

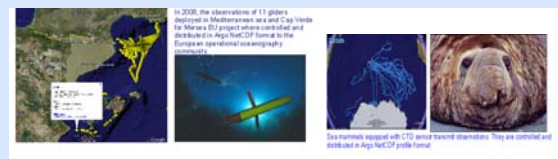
- **NetCDF:** <http://www.unidata.ucar.edu/software/netcdf/docs/BestPractices.html>
- **Udonits:** <http://www.unidata.ucar.edu/software/udunits/>
- **CF:** <http://cf-pcmdi.llnl.gov/>
- **COARDS:** http://www.ferret.noaa.gov/noaa_coop/coop_cdf_profile.html
- **ISO8601:** http://en.wikipedia.org/wiki/ISO_8601

Argo, Gosud, OceanSITES : Global data management architecture within JCOMM



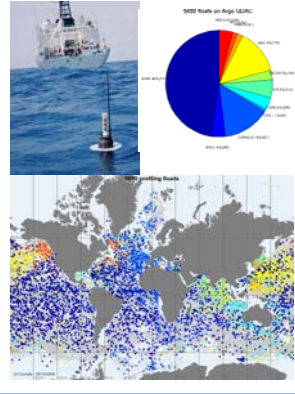
JCOMM, the Joint commission between IOC and WMO, coordinates the international implementation of the global ocean observing system. The challenges is not only to deploy and maintain these networks but also to provide an easy integrated access to these data both in real-time and in delayed mode.

A strong data-management architecture, mainly based on NetCDF and CF conventions is essential to operate Global Data Centers for three of these networks : Argo profiling floats, Gosud surface underway observations from ships and OceanSITES long term moorings. In collaboration with European partners and other European laboratories, Coriolis also integrates and distributes Gliders and Sea Mammals data with this architecture.



ARGO

Argo is an international program that aims at deploying 3000 floats in the whole ocean on a 3°x3° grid cycling to 2000m every 10 days for about 4 years. Deployment started in 2000 and the 3000 float goal was reached in November 2007. The goal is presently to sustain this network by deploying 800 float per Year. US-GODAE and Coriolis, the two Argo Global Data Centers provide both FTP and WWW access to the data.



GOSUD

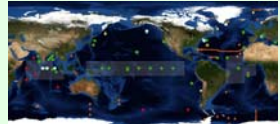
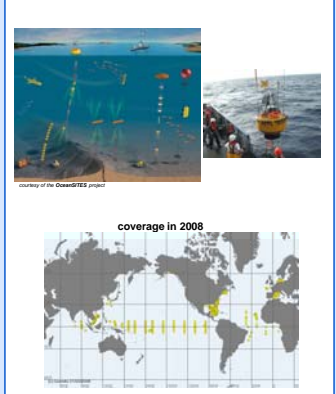
GOSUD is an international program that aims at collecting Thermosalinograph data from research vessels and to transmit them in real-time. The program started in 2002 with French research vessels and VOS coordinated by IRD and was extended to European vessels within Mersea. US-NODC and Coriolis, the two Global Data Centers provide both FTP and WWW access to the data.



OceanSITES

OceanSITES is an international program that aims at maintaining fixed point observatories in reference areas. It also aims at improving access to these reference data. The EuroSITES sites are part of this network.

Coriolis is setting up one of the two Global Data Centers for Argo and providing both FTP and WWW access to the data. It has first integrated the TAO/TRITON/PIRATA networks and the European Sites operated within MERSEA. It should be extended to more USA sites in 2008.



OceanSITES data exchange formats

The OceanSITES program is the global network of open-ocean **sustained timeseries sites**, called ocean reference stations, being implemented by an international partnership of researchers. OceanSITES provides fixed-point timeseries of various physical, biogeochemical, and atmospheric variables at different locations around the globe, from the atmosphere and sea surface to the seafloor. The program's objective is to build and maintain a multidisciplinary global network for a broad range of research and operational applications including climate, carbon, and ecosystem variability and forecasting and ocean state validation.

All OceanSITES data are publicly available. More information about the project is available at <http://www.oceansites.org>.

OceanSITES uses the NetCDF (network Common Data Form) system, a set of software libraries and machine-independent data formats. **Our implementation of NetCDF is based on the community-supported Climate and Forecast (CF) specification**, which supplies a standard vocabulary and some metadata conventions. OceanSITES has several more restrictions than the CF standard. These are intended to make it easier to share in-situ data, to make it simpler for the GDACs to aggregate data from multiple sites, and to ensure that the data can be created and understood by the basic NetCDF utilities.

An OceanSITES data file contains measurements such as temperature and salinity, continuously performed at different levels on a platform (e.g. mooring), as well as meteorological or other parameters recorded at the site, derived variables associated with the site, and complete location, time, and provenance information.

- The requirements are drawn almost exclusively from the NetCDF Style Guide:
- Units are compliant with CF/COARDS/Udonits ;
 - The time parameter is encoded as recommended by COARDS and CF.
 - Parameters are given standard names from the CF table
 - Where time is specified as an attribute, the ISO8601 standard is used.

Argo data management architecture

Argo data-management is based on Argo NetCDF files exchanges (profiles and trajectories data files, metadata and technical data files).

