



**Inter-Program Expert Team on Data  
Representation Development (IPET-DD)  
Zagreb, Croatia 18-20 February 2019**

**IPET-DD-1  
Agenda Item [7]  
20 February 2019**

**Submitted by: Øystein Godøy/Norwegian  
Meteorological Institute**

## **Reasons for using NetCDF and importance of conventions**

### **Introduction**

NetCDF provides the possibility to generate self-describing, machine independent data encoding. In order to achieve fully self-describing and machine interpretable data, conventions on how to encode data as well as the semantics required are also needed. This document complements a presentation providing an overview of considerations made in the context to [WMO GCW](#) to ensure interoperability across the heterogeneous community contributing to GCW.

### **WMO GCW data management**

The intention of the GCW Portal is to be the entry point to datasets describing the cryosphere and form the information basis for the assessment activities of the Global Cryosphere Watch. It offers a web interface that contains information about datasets through discovery metadata provided by the data providers (or host data centre). These discovery metadata are harvested on a regular basis from data centres actually managing the data on behalf of the owners/providers of the data.

The GCW Portal utilises interoperability interfaces to metadata and data in order to provide a unified view on the datasets that are relevant for GCW activities. The GCW Portal will facilitate real time access to data through Internet and WMO GTS as requested by the user community. This requires a certain level of interoperability at the data level in addition to at the metadata level. On GTS, WMO formats (BUFR and GRIB) are required and the GCW Portal can transform data into these formats in the dissemination process, provided contributing data centres are following the required standards for documentation and interfaces to data.

The GCW data management user community (data providers and consumers) is a heterogeneous community ranging from national meteorological and hydrological services through universities to independent research institutions. In the WMO context, the GCW user community is very similar to the WMO GAW user community. GCW needs to bridge between the traditional WMO communities and the

science community. The situation for data providers in the context of GCW is that many have existing systems, but that these are not interoperable. Some doesn't have any structured data management.

The data management component is an enabling service in the sense that it identifies relevant datasets and their locations and provides an interface that can be used in the evaluation of GCW data and products. The portal will support simple visualization (generation of maps or diagrams like time series) and transformations such as reformatting and re-projection of data, *if the data are served through the appropriate interfaces and forms*. The GCW Data Portal takes the role as a mediator between the data provider and the data consumer hiding some costs for the user community (Figure 1).

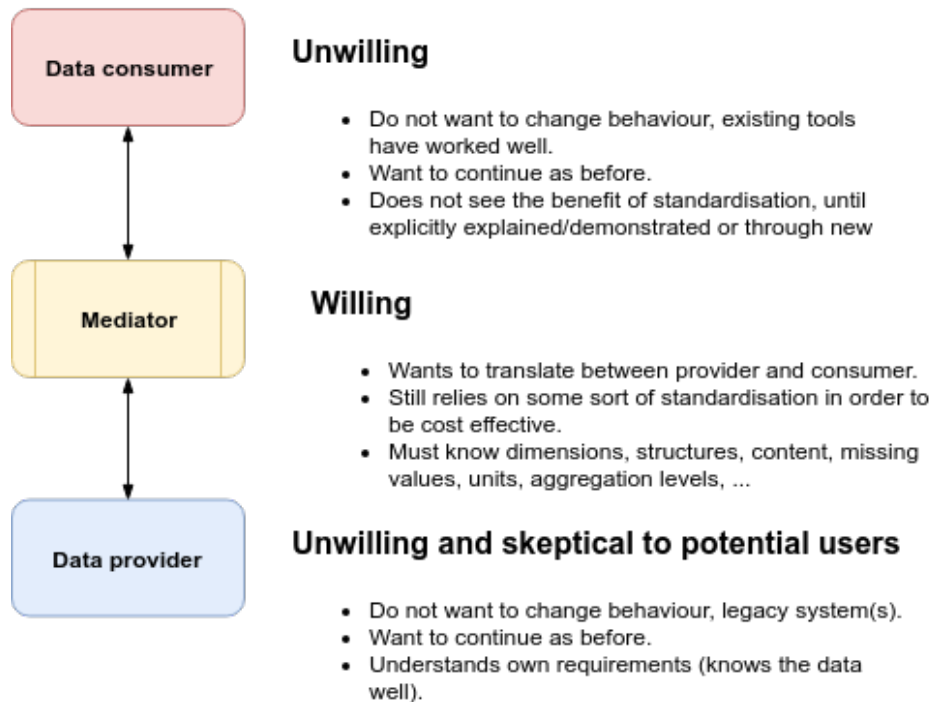
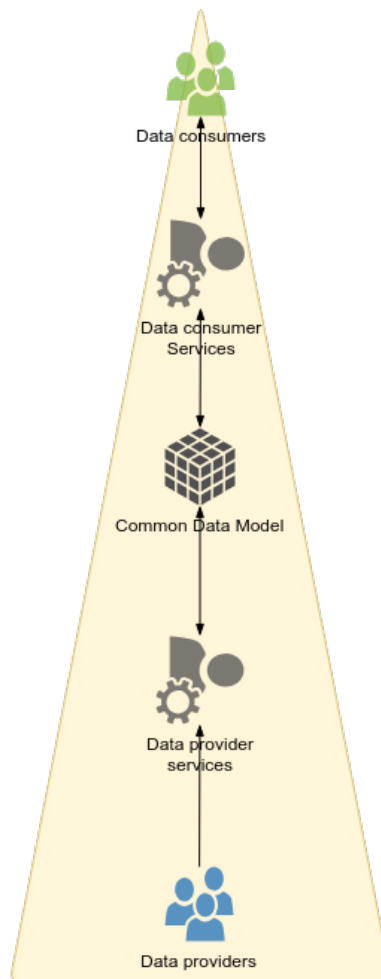


Figure 1: The GCW Data Portal is a mediator between user communities.

GCW data management shall integrate datasets and provides access to data and information on past, present, and future cryospheric conditions. To achieve these results, the data portal must be attached to real-time and near-real-time data management systems and to data archives. While interfacing with existing data management systems, GCW respects partnership and ownership. GCW itself will rely on distributed data management technologies and partners (e.g. CryoNet stations) to establish the GCW catalogue. This process will create a unified interface to datasets in an otherwise fragmented terrain. No information on data (discovery metadata) will be kept in the GCW catalogue without an agreement with the data producer/data owner.

In the GCW context, at least two types of metadata are relevant. One is “discovery” or index metadata identifying general characteristics of a dataset, including what was measured where and when, potential restrictions on data use, data custodians, and the available interfaces to the actual dataset. This is the type of metadata that will be exchanged within GCW. Another type, “use” metadata, is required when a

user has accessed a dataset and begins to use it. Such metadata typically include a specification of variables, units used, how missing values are encoded, and other details on the contents of the dataset. The ingested discovery metadata will be harvested from project specific, national, and international catalogues. Successful exchange of metadata will involve some degree of adaptation of systems on either side. However, in order to establish a sustainable system, the number of standards the GCW portal has to support cannot be too many. Furthermore, the actual data also has to be standardised to support integration of data among data providers in support of data consumers (Figure 2).



*Figure 2: Standardisation across existing data repositories in a cost efficient manner.*

### **NetCDF/CF in the context of WMO GCW**

NetCDF in itself doesn't solve any interoperability issues between data providers and data consumers in the GCW context. However, when the [Climate and Forecast Conventions](#) (CF) and the [Attribute Convention for Dataset Discovery](#) (ACDD) are added, NetCDF files are standardised in manner which

allows building services on top of the data. Such services can be visualisation, transformation (e.g. subsetting, reformatting, combination of data etc).

The primary issue in WMO GCW is not that data is made available as NetCDF, but that data are encoded according to CF and ACDD and made available using [OPeNDAP](#). This way existing data can be integrated in GCW data management through adapters, removing the need for a full reimplement of data centres, that map non standardised terminology into standardised terminology and structures (Figure 3). The original data can be NetCDF, or it can be spreadsheets or relational databases. The mapping into a standard (CF) and application of the data streaming protocol OPeNDAP allows services on top of the data.

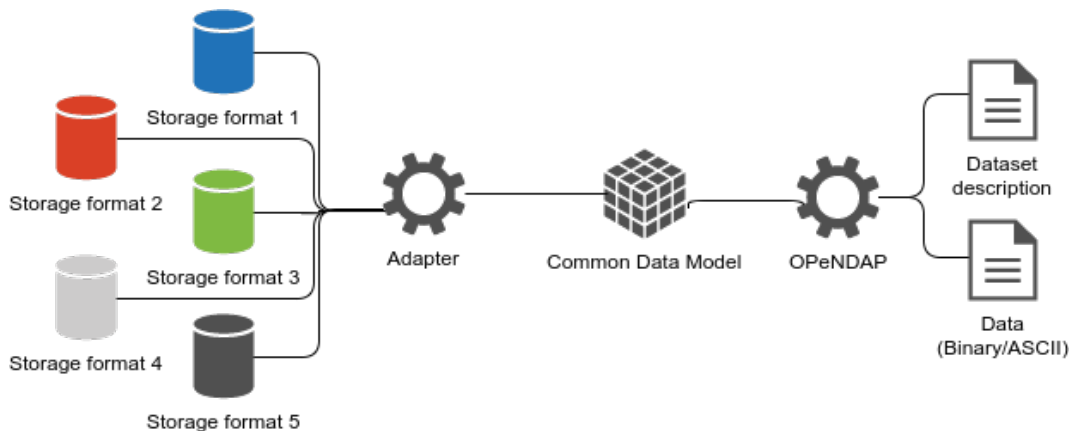


Figure 3: Mapping existing non-standardised data to standardised data using adapters.

### Benefits of using NetCDF/CF and ACDD

WMO table driven formats are primarily used within the WMO community. The process of reading, notspeaking of creating these table driven formats is difficult and in most scientific communities considered too difficult and time consuming. On the other hand, NetCDF is widely used in science communities, especially within climate modelling. The CF convention has support for gridded data, timeseries, profiles etc (Figure 4) which simplifies the process of combining data from multiple sources.

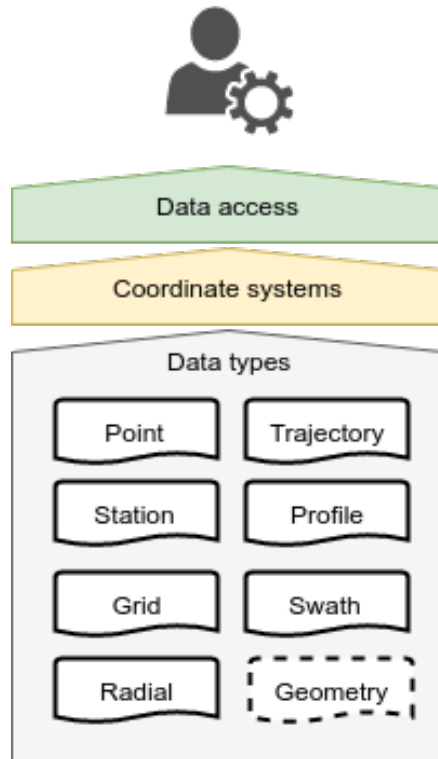


Figure 4: Several types of data are supported or in the process of being supported through CF conventions.

## Lessons learned

The CF conventions are too wide, narrower profiles are necessary to be able to build services on top of the data. The scientific communities not accepting NetCDF is often not using any standards at all, but if supported with tools they are usually positive. Although CF is not focusing on the semantic web approach, the approach using standard names etc is useful in that context as well and part of several initiatives exploiting this approach for better discovery and use experiences across communities. The major insufficiencies of CF currently are lack of a standardised way to handle shapes (e.g. similar to KML, Shape etc) and hierarchies in data (grouping). When it comes to standardisation efforts, too few people and communities are engaged. It is not beneficial for the geoscientific communities to compete, but rather to join efforts and focus on extending the scope for standards that have sufficient momentum.

-----