COMMISSION FOR BASIC SYSTEMS

World Meteorological Organization
Working together in weather, climate and water

netCDF CF conventions for WMO data exchange

19-20 September 2019

CF workshop
Agenda Item [1]

13 September 2019

Submitted by: Øystein Godøy, Mariano Masiokas, GCW Project Office – WMO Secretariat/ WMO GCW

NetCDF/CF in the context of GCW: advantages and bottlenecks

Introduction[1]

The World Meteorological Organization's Global Cryosphere Watch (GCW) is an international mechanism to foster international coordination and partnerships with the goal of providing authoritative, clear, and useable data, information, and analyses on the past, current and future state of the cryosphere. The scope of GCW activities includes cryosphere observations and data access, including best practices, assessments and product development, support to prediction, and research.

The GCW stakeholders are WMO Members and partners delivering services to users, to decision and policy makers, and in support of the development of climate adaptation and mitigation policies. The goal of the GCW framework are achieved by linking research and operations, and scientists and practitioners, through sustained and mutually beneficial partnerships between research and operational institutions on matters related to cryosphere.

One of the key priorities of GCW for the next 4 years is to facilitate the standardization, access to, and quality management of current and past cryosphere data, information, and products , including by fostering collaboration through partnerships between operational and scientific communities to expand the availability and access to cryosphere data and information. These are important to fostering the development and publication of user-driven value-added cryosphere products based on in-situ, space-based, and airborne observations, as well as models.

The primary focus of this document and presentation is on the element of Access to and utilisation of observations and products, in the context of the GCW framework. The types of data handled through WMO GCW data management ranges from traditional meteorological sensor observations, through manual observations, outline products and gridded products like satellite products etc. Data available through the GCW portal are provided by a variety of institutions with different data management

[1] Standard text from the WMO GCW website.

infrastructure, systems, expertize, and raison d'etre (e.g. glaciology data used for institutional, national, and international programs e.g. WGMS, GLIMS, etc)

In most regions cryosphere data is scarce and observing programs expensive to run. Access to data from the heterogeneous community of data providers is the only feasible mechanism to meet the needs of data users (modellers, prediction, etc).

GCW Data Management

Introduction

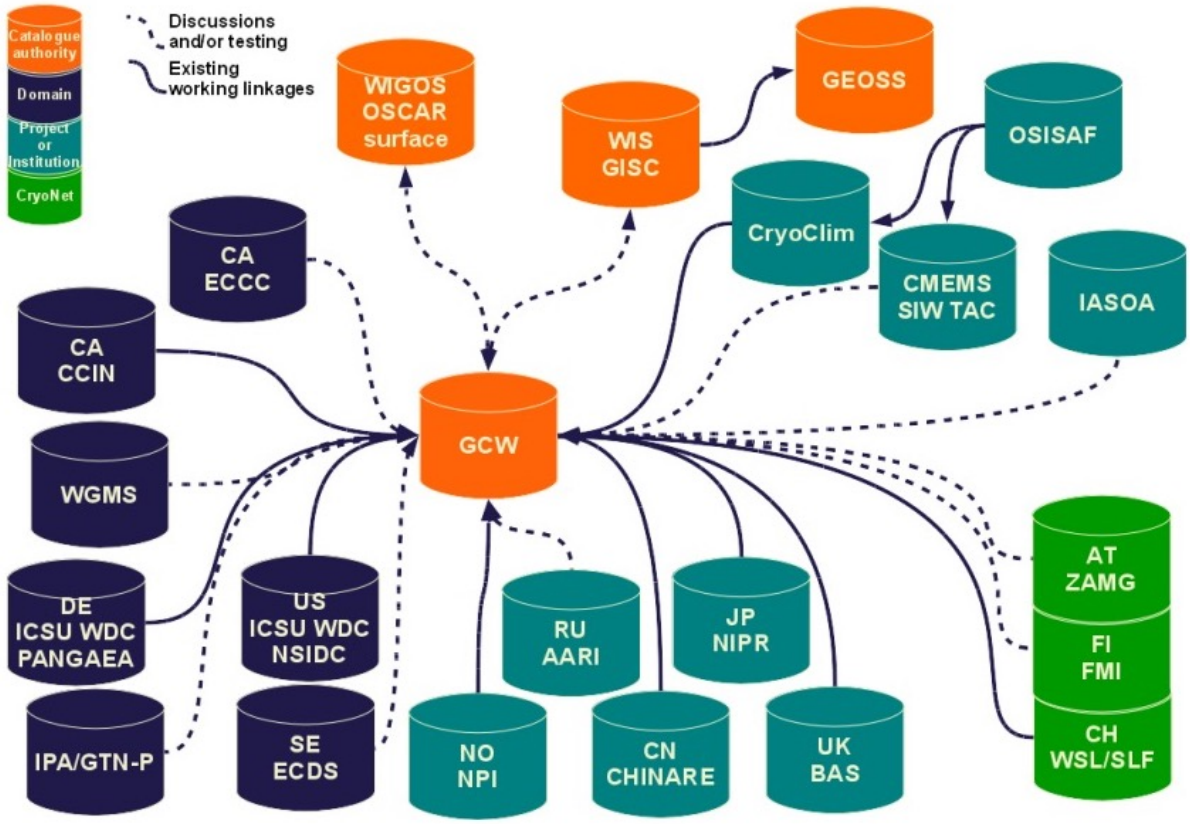

Figure 1: Overview data centre interactions of WMO GCW data management. Solid lines indicate existing linkages; dashed lines indicate ongoing discussions and/or testing.

GCW data can be accessed through the **GCW Data Portal**. The Portal is a web interface that contains information about datasets (through discovery metadata), but not the data itself. Instead, it links to datasets that are stored at partner data centres.

The diversity of data, current metadata and data semantics, formats, and protocols, as well as types of organisations contributing are significant barriers to the development of a data catalogue or portal. Standardization, timeliness, and interoperability of data and data systems are critical.

Data from many contributors to GCW use own formats, aligned with specific applications. A common data and metadata model is needed, and GCW has adopted NetCDF for this, as is a widely used data representation format, adopted already by many communities.

The primary advantage of using NetCDF/CF in the context of WMO GCW is that it links with the generic data model of UNIDATA - Common Data Model. This links with both file format and the data exchange through OPeNDAP. As OPeNDAP allows data streaming, there is no need to move files around in order to establish services on top of the datasets served in the GCW data portal.

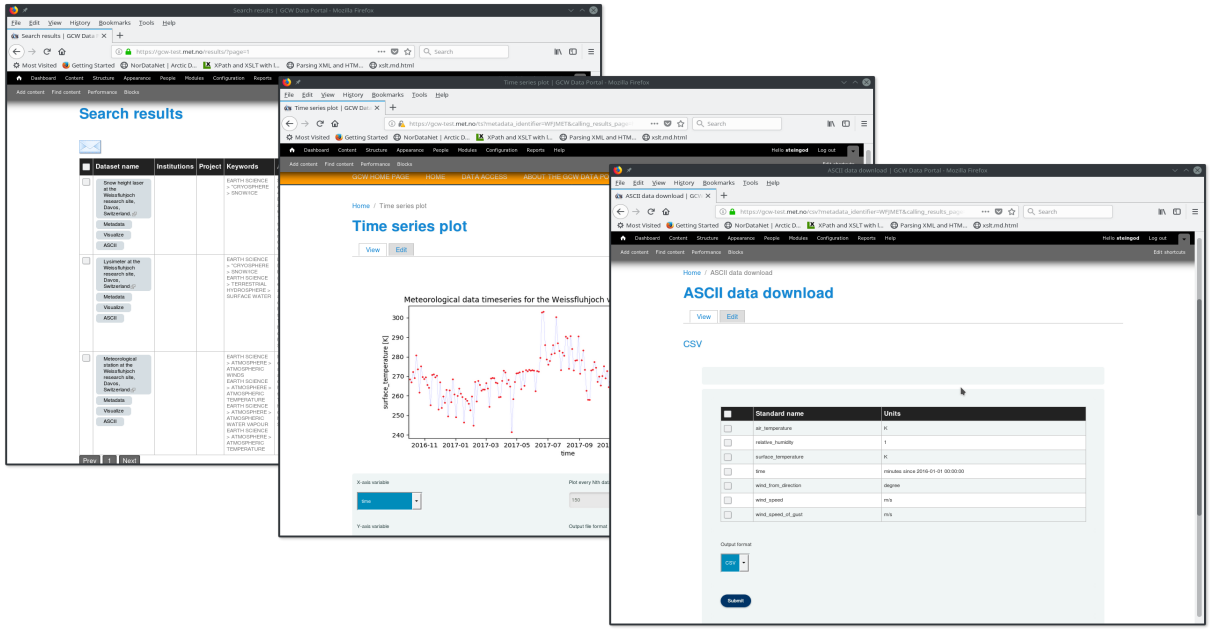

Figure 2: Illustration of how NetCDF/CF datasets are visualised and served through the GCW Data Portal.

Advantages

NetCDF/CF provides FAIR compliant data documentation standard which simplifies integration of data across a distributed data management system composed by both operational NMHSs and independent research stations. All data are mapped to the same generic data model using a harmonised semantic framework for annotation which is beneficial when combining data from different sources. There are also emerging work focusing on linked data approaches which are considered beneficial in a longer term. The ability to easily stream data using OPeNDAP is a major benefit which reduces development and maintenance costs for distributed systems. The ability to create virtual datasets e.g. along the temporal dimension is highly appreciated.

Bottlenecks or gaps

Given the wide variety of data collected by GCW partners, quite frequently CF standard names are missing. A coordinated approach and faster process of getting standard names accepted would be appreciated. Furthermore, linking CF standard names and WIGOS keywords using a semantic framework would be highly beneficial and improve functionality for data providers. The CF standard is lacking

standardised approaches for vector data like annotated polygons, lines and points (e.g. KML, Shapefiles). Adding support for this type of data would close a gap. Finally, the degrees of freedom using the CF standard complicates building services on top of NetCDF/CF served data. Development of community profiles for fixed stations, moving stations (surface and profile observations) as well as for gridded data mapping to OGC WMS etc would be benficial for further development of services. Better handling (more user friendly) of auxiliary information would benefit from support of CMD2 allowing hierarchies. Mapping towards DAP4 should be addressed. Finally, for large datasets on well defined map projections, geographical positions should be possible to leave in order to reduce storage sizes.

Recommended text

In achieving the GCW goals on standardization, access to, and quality management of current and past cryosphere data, information, and products, the initiative of WMO investigating the opportunities and costs of developing regulations, based on the CF conventions, is a welcome and necessary step, to enable interoperability in the exchange of NetCDF data for WMO purposes in the WMO Information System (WIS). Without it, the goals of GCW are not fully feasible.

It is recommended that in the process of establishing the terms of reference of the associated process the following are taken into consideration: (1) the harmonization and/or semantic mapping of terminology and semantics between WIGOS, WIS, and CF convention, with appropriate governance, to ensure the right balance between expediency and usability; establish consultative process for standardised approaches for newer data types, e.g. vector data like annotated polygons, lines and points (e.g. KML, Shapefiles), identification of WMO profiles for various data sources (e.g. profiles for fixed stations, moving stations (surface and profile observations) as well as for gridded data (including facilitating mapping to OGC WMS). (2) While the existing WMO data representation system is addressed to a relatively small and known community, the CF Convention provides a framework of interoperability used by a much broader and diverse community. Hence, the governance process needs to address the degrees of freedom possible and the user friendliness, to ensure that the usability of data remains the core goal. (3) Based on the experience to date, additional specific recommendations will be made, including that mapping towards DAP4 should be addressed, and that for large datasets on well-defined map projections, geographical positions should be possible to leave in order to reduce storage sizes.

The GCW community is a very keen stakeholder of the proposed process and wishes to remain closely engaged in the process of defining the framework, and addressing specific needs.
