NETCDF-CF-OPENDAP: STANDARDS FOR OCEAN DATA INTEROPERABILITY AND OBJECT LESSONS FOR COMMUNITY DATA STANDARDS PROCESSES

Steve Hankin[1], Jon D. Blower[2], Theirry Carval[3], Kenneth S. Casey[4], Craig Donlon[5], Olivier Lauret[6], Thomas Loubrieu[7], A. Srinivasan[8], Joaquin Trinanes[9], Øystein Godøy[10], Roy Mendelssohn[11], Rich Signell[12], Jeff de La Beaujardiere[13], Peter Cornillon[14], Frederique Blanc[6], Russ Rew[15], Jack Harlan[16]

[1] *NOAA/Pacific Marine Environmental Laboratory, Seattle, USA, Email:steven.c.hankin@noaa.gov*
[2] *Environmental Systems Science Centre, University of Reading, Reading, UK, Email:j.d.blower@reading.ac.uk*
[3] *Ifremer, Brest, France, Email:Thierry.Carval@ifremer.fr*
[4] *NOAA/National Oceanographic Data Center, Silver Spring, MD, USA, Email:Kenneth.Casey@noaa.gov*
[5] *European Space Agency, Noordwijk, The Netherlands, Email:craig.donlon@esa.int*
[6] *CLS Space Oceanography Division, Ramonville, France, Email:olauret@cls.fr*
[7] *Ifremer, Brest, France, Email:Thomas.Loubrieu@ifremer.fr*
[\8] *Center for Ocean-Atmospheric Prediction Studies, The Florida State University, Talahassee, USA, Email:asrinivasan@rsmas.miami.edu*
[9] *University of Santiago de Compostela, Santiago de Compostela, Spain, Email:Joaquin.Trinanes@noaa.gov*
[10] *Norwegian Meteorological Institute, Email:o.godoy@met.no*
[11] *NOAA/Southwest Fisheries Science Center, Pacific Grove, CA, USA, Email:Roy.Mendelssohn@noaa.gov*
[12] *USA Geological Survey, Woods Hole, MA USA, Email:rsignell@usgs.gov*
[13] *NOAA/ Integrated Ocean Observing System (IOOS) Program, Silver Spring MD, USA, Email:Jeff.deLaBeaujardiere@noaa.gov*
[14] *Graduate School of Oceanography, University of Rhode Island, Narragansett, RI, USA, Email:pcornillon@gso.uri.edu*
[15] *Unidata / University Corporation for Atmospheric Research, Boulder, CO, USA, Email:russ@unidata.ucar.edu*
[16] *NOAA/ Integrated Ocean Observing System (IOOS) Program, Silver Spring MD, USA, Email:Jack.Harlan@noaa.gov*

ABSTRACT

It is generally recognized that meeting society's emerging environmental science and management needs will require the marine data community to provide simpler, more effective and more interoperable access to its data. There is broad agreement, as well, that data standards are the bedrock upon which interoperability will be built. The path that would bring the marine data community to agree upon and utilize such standards, however, is often elusive.

In this paper we examine the trio of standards 1) netCDF files; 2) the Climate and Forecast (CF) metadata convention; and 3) the OPeNDAP data access protocol. These standards taken together have brought our community a high level of interoperability for "gridded" data such as model outputs, satellite products and climatological analyses, and they are gaining rapid acceptance for ocean observations. We will provide an overview of the scope of the contribution that has been made.

We then step back from the information technology considerations to examine the community or "social" process by which the successes were achieved. We contrast the path by which the World Meteorological Organization (WMO) has advanced the Global Telecommunications System (GTS) – netCDF/CF/OPeNDAP exemplifying a "bottom up" standards process whereas GTS is "top down". Both of these standards are tales of success at achieving specific purposes, yet each is hampered by technical limitations. These limitations sometimes lead to controversy over whether alternative technological directions should be pursued.

Finally we draw general conclusions regarding the factors that affect the success of a standards development effort – the likelihood that an IT standard will meet its design goals and will achieve community-wide acceptance. We believe that a higher level of thoughtful awareness by the scientists, program managers and technology experts of the vital role of standards and the merits of alternative standards processes can help us as a community to reach our interoperability goals faster.

1. INTRODUCTION

It is generally recognized that meeting society's emerging environmental science and management needs will require the marine data community to provide simpler, more effective and more interoperable access to its data. There is broad agreement, as well, that data standards are the bedrock upon which interoperability will be built. The precise meaning of a "standard" and the path that would bring the marine data community to agree upon and utilize standards, however, are often

elusive. Assessments of the recent history of "standards processes", the human processes by which standards are developed, describe how traditional Standards Development Organizations (SDOs) such as ANSI, DIN, and ISO have been overtaken by market-driven consortia (also referred to as Standards Setting Organizations or SSOs) [1,2]. While successful IT standards are and will continue to be the foundation of data interoperability for the Web, the achievements of the consortium processes have been spotty[1]. The majority of IT standards published in recent years achieve prominence too briefly to realize the vision that they advance and are replaced in the marketplace by the next "cool idea".

Thus the view of the authors of this paper is that a high level of both pragmatism and scepticism needs to be brought to the ongoing debates about IT interoperability standards. Pragmatism implies that a "standard" should have demonstrated its ability to perform the task for which it is required before commitments are made to use it. Scepticism implies that simplicity, completeness, and readiness for the intended purpose should be weighed more heavily than unproven visions of superior interoperability in the future or the prestige of the organization issuing the standard.

In this paper we highlight two standards for marine data interoperability that we believe have passed the tests of pragmatic and sceptical evaluation: (i) netCDF, when joined with the Climate and Forecast (CF) metadata conventions and OPeNDAP as a network access layer; and (ii) the World Weather Watch (WWW) Global telecommunications System (GTS). Both of these technologies are providing high levels of interoperability for marine data today. As with all software, both have limitations, but are being incrementally enhanced along paths that address the most serious of those limitations.

2. DATA INTEROPERAVILITY VIA NETCDF-CF-DAP – AN OVERVIEW

2.1 NetCDF

NetCDF (network Common Data Form) is open-source software that supports the creation, access, and sharing of scientific data and metadata using machine-independent data formats. Developed and supported by UCAR's Unidata program[2] since 1989, netCDF's strengths include the simplicity of the underlying data model; flexibility and efficiency of data access; well designed Application Programming Interfaces (APIs)

that enables quick software development in popular programming languages; and an active user community. NetCDF has become a de facto standard for model outputs and climatological products in the atmosphere, climate, and ocean modelling communities. Its use for observational data is growing as we document in this paper.

Recently the developers of netCDF and HDF5 (Hierarchical Data Format), supported by a NASA award, have collaborated to create software that combines the desirable characteristics of each [3]. The HDF format has been widely used in the remote sensing community, so this merger promises to increase interoperability between these communities. The resulting software, netCDF-4, remains backwards compatible with current netCDF-3 programs and data, but contains performance enhancements and increased capabilities for encoding collections of observational data. This fusion has ushered in a period of IT experimentation throughout the netCDF community in which application software is adapted to utilize the new capabilities. Following this period we hope to see the successful adaptations of the new capabilities appear as extensions to the CF conventions, leading to an enhanced level of interoperability.

2.2 Climate and Forecast (CF) Metadata Conventions

The netCDF-3 data model is discipline-neutral – i.e. general enough to be applicable to many different disciplines of science and engineering. The CF metadata conventions[3] [4] describe how to encode ocean/atmosphere/climate datasets in netCDF – addressing space and time coordinates, units, standardized parameter names, sub-grid sampling methods and other concepts. CF-compliant data files are self-describing: no external tables are needed to interpret the file (contrasting with GRIB). CF includes a controlled vocabulary of parameter names -- 1100 at the time of writing – which are the result of inputs from many communities and disciplines. The metadata and the semantic meaning in CF files are readable by humans as well as easily utilized by programs. A particular strength of the netCDF-CF combination is the breadth of powerful scientific software that can read it[4].

CF is an unusually "open" standard. Any community member can propose enhancements and report problems, which are discussed publicly with the help of a Web-based issue-tracking system[5]. Conventions are developed only to address known issues. Instead of trying to foresee the future, features are added only as

---

[1] A notable exception is the IETF, which is discussed under Conclusions

[2] www.unidata.ucar.edu

[3] http://www.cfconventions.org

[4] http://www.unidata.ucar.edu/software/netcdf/

[5] http://cf-pcmdi.llnl.gov/discussion

required. While this high level of community involvement brings with it many advantages it can also at times slow the progress of CF. CF governance is based upon community-agreed procedures that define how new issues are to be opened, debated, tracked, and resolved. Three volunteer committees – a Conventions Committee, a Standard Names Committee and a Governance Panel – oversee the process. As the importance of CF continues to grow there will be an increasing need to find institutional support for tasks that will speed along the development of CF. Currently the British Atmospheric Data Centre Center and Lawrence Livermore Laboratory are providing levels of support, and it is expected that NOAA's National Climatic Data Center will soon be augmenting this. At the time of this writing the CF standard stands at version 1.4, released February 2009.

The CF conventions are extensible. The standard allows community-specific metadata that lies outside of the CF specification itself to be included. The CF conventions are gaining acceptance for the representation of in situ and satellite derived data through the addition of metadata that describe instrument types, calibrations, etc.. Currently the CF community and Unidata are working together to unify the representation of gridded and in situ observation data types under a Common Data Model (CDM) and its associated software libraries.
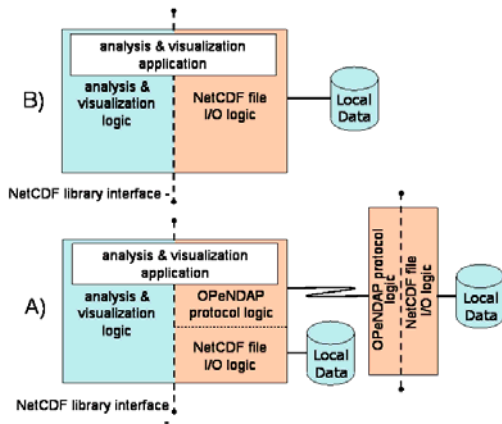

Figure 1. Illustrates how remote access to netCDF data becomes transparent to the application through OPeNDAP. B, before relinking; A, after relinking

2.3 OPeNDAP and server-side dataset configuration

The foundation of OPeNDAP (Open source Project for a Network Data Access Protocol) is the network data access protocol – the DAP. Conceived in the early 1990s as part of the DODS (Distributed Oceanographic Data System) effort, the DAP is a discipline-neutral protocol designed to facilitate access to data stored at remote locations on the Internet. The DAP allows users to request subsets of data; thus only the data of interest from potentially very large datasets need be transferred across the Internet. Hundreds of DAP servers currently serve thousands of data sets and many terabytes of data for a broad range of applications.

The netCDF-3 data model is (with minor extensions) a subset of the DAP data model. As a result many years ago collaborative efforts between OPeNDAP and Unidata achieved a tight fusion of netCDF with the DAP. As shown in Fig. 1 applications that have been programmed to read netCDF-CF data files can automatically read the same data across the Internet if "relinked" with the OPeNDAP-enabled libraries[6]. This technical detail has huge significance for interoperability. Common ocean science applications such as Matlab®, Ferret [5], GrADS [6] are instantly transformed into network data access tools when they are OPeNDAP-enabled.

The concept of a virtual file that is inherent in OPeNDAP has also made it possible to develop strategies to "reconfigure" data files, so that legacy data formats may appear through the Internet to be CF-compliant data. This approach can be applied to a broad range of ASCII and binary formats, including the GRIB file format mandated by the WMO. The data provider need only create a short configuration file in NcML (XML) format [7] in order to make the data appear to users as a fully CF-compliant dataset. No physical alteration to the legacy data files is required.

Many OPeNDAP servers also offer "aggregation" capabilities. Aggregation refers to the ability to present a large collection of individual files – e.g. files that may individually contain only single variables, time steps or horizontal grids – as a coherent "virtual dataset" with many variables occupying a shared 4-dimensional space-time coordinate system. Working with an aggregated dataset liberates the scientist/user from having to know which data files contain exactly the data that she needs. She is able to focus on the scientific questions.

The net effect of OPeNDAP is that it permits ocean science applications to read files in formats with which they were previously incompatible. and in a manner that is often more flexible than the original format. OPeNDAP can serve data from netCDF, GRIB, GRIB2, HDF4, HDF5, HDF-EOS, flat IEEE binary, many variations of ASCII, HYCOM binary and other file formats as well as relational databases. New formats can readily be added to well-tested servers that exist in

---

[6] Relinking is not required for applications written in Java, and is no longer required for any applications since the release of version 4.1 of the netCDF libraries.

Java (the THREDDS Data Server[7]), C++ (the HYRAX server[8]), and Python (the PyDAP server[9]).

3. APPLICATIONS OF NETCDF-CF-DAP FOR OCEAN DATA MANAGEMENT

3.1 Ocean Model Output Data

The trio of netCDF-CF-DAP has become a de facto standard for the output of ocean models. NetCDF files that follow the CF Conventions provide sufficient information to allow flexible extraction of 4D geospatial data in a uniform way. Users' client software can extract lat-long grids, vertical sections, time series or other subsets as needed. "Virtual" CF datasets configured and aggregated through NcML can be served at low effort via a variety of services, including DAP, WCS and WMS[10] using technologies such as the THREDDS Data Server from Unidata. The combination of these tools provides a powerful way for scientists to collaborate on model analysis [8] and was a foundation of success for data management during GODAE [9].

CF supports the complex types of grids that must be represented to capture the range of ocean model coordinate types. Grids may have a rectilinear or curvilinear coordinate system in the horizontal, and may have various vertical coordinate systems based upon fixed levels in depth or density or upon layers which expand and contract with the moving sea surface. In some cases the coordinate system may be time dependent and adaptive. The most complex of coordinate types remain as areas of active development -- targets for future versions of the CF conventions. These include unstructured (usually triangle-based) meshes and "gridspec" coordinate systems [10] that are composed of gridded tiles, connected along boundaries.

3.2 Satellite Data

The netCDF-CF-DAP combination has also become the de facto standard for several satellite-based level 3 and 4 ocean data sets and programs and in some cases level 2 native swath data sets. Three examples from across the spectrum of ocean remote sensing clearly illustrate the broad use and utility of this standard. First, the Group for High Resolution Sea Surface Temperature

(GHRSST[11]) mandates the use of CF-compliant netCDF throughout its international framework. GHRSST Regional Data Assembly Centers contribute SST data sets in near real time to a Global Data Assembly Center[12] at NASA's Jet Propulsion Laboratory, which passes them along after 30 days to NODC's Long Term Stewardship and Reanalysis Facility[13]. GHRSST augments CF with further metadata to support automated data distribution and long-term stewardship. The GHRSST and CF communities have collaborated in expanding the CF standard name list to include the specialized types of SST measured by orbiting sensors. Native L2 swath data, L3 gridded data and L4 analysis (merged) data products are all defined using netCDFn-CF. Over 25Gb of GHRSST data are produced each day and over 30,000 users have accessed the GHRSST netCDF-CF data.

A second example is the Jason-2 satellite altimetry mission. The four partners (NOAA, NASA, CNES, and EUMETSAT) agreed upon netCDF-CF as the base file format for the entire collection of mission data, simplifying the international exchange of the data and making the Jason-2 altimetry data stream easier to utilize than any prior altimeter mission. The third and most recent example comes from NOAA, where some data management teams have mandated the use of netCDF-4 with CF metadata for all their intermediate products. NOAA recognized the cost savings and efficiencies that can be gained by relying on netCDF-CF standards as part of a "one to many" approach with an enterprise file format converter for delivering the data in various output formats as preferred by different user communities.

3.3 JCOMM *in situ* networks

JCOMM, the Joint commission between IOC and WMO, coordinates the international implementation of the Global Ocean Observing System (GOOS). The challenge is not only to deploy and maintain the networks of observation platforms, but also to provide easy-to-use, integrated access to these data both in real-time and in delayed mode.

A strong data management architecture is essential to operate Global Data Centers for three of these networks: Argo profiling floats, OceanSITES long term moorings, and GOSUD surface underway observations from ships. All three of these systems have elected to base their data management on netCDF and CF conventions (plus extensions). In collaboration with European partners and other European laboratories, Ifremer also integrates

---

[7] http://www.unidata.ucar.edu/Projects/THREDDS/tech/TDS.html

[8] http://docs.opendap.org/index.php/Hyrax

[9] http://pydap.org/

[10] WCS (Web Coverage Service) and WMS (Web Mapping Service) are standards from the Open Geospatial Consortium (OCG -- http://www.opengeospatial.org/)

[11] http://www.ghrsst-pp.org

[12] http://ghrsst.jpl.nasa.gov

[13] http://ghrsst.nodc.noaa.gov

and distributes Gliders and Sea Mammals data with this architecture.

3.4 Argo float data

At the time that the Argo profiling floats data management architecture was initiated (2001), there were no off-the-shelf conventions suitable to encode Argo data. However, netCDF was emerging and offered accessibility to a variety of established scientific software tools. A set of metadata conventions for netCDF was designed and documented in the Argo User's Manual. With the emergence of the CF community the decision was made to extend the Argo format to be CF compatible and benefit from the wider availability of software tools. Moreover within the EU SeaDataNet project[14], collaboration have started to harmonize the CF and IOC GF3 parameter codes and implement gateways between the two vocabularies.

Argo floats observations are available from 2 OPeNDAP mirror sites (USGODAE server and Ifremer) and a long term archive (US NODC):

- http://www.usgodae.org/cgi-dods/nph-dods/ftp/outgoing/argo
- http://www.ifremer.fr/cgi-bin/nph-dods/data/in-situ/argo
- http://data.nodc.noaa.gov/opendap/argo
- http://www.usgodae.org/cgi-dods/nph-dods/ftp/outgoing/argo
- http://www.ifremer.fr/cgi-bin/nph-dods/data/in-situ/arg

3.5 OceanSITES data management

The OceanSITES program is the global network of open-ocean, sustained time series sites, called ocean reference stations. They provide fixed-point time series of various physical, biogeochemical, and atmospheric variables at locations around the globe, from the atmosphere and sea surface to the seafloor.

OceanSITES uses netCDF-CF to make it easier to share data and to aggregate observations from multiple sites, while ensuring that the data can be created and understood by the basic netCDF utilities.

A single OceanSITES data file contains time series measurements performed at different vertical levels on a mooring (oceanic, meteorological parameters or derived variables associated with the site, and complete location, time, and provenance information). All OceanSITES observations are publicly available. More information about the project is available on-line[15].

3.6 Underway data management

Atmospheric and oceanographic observations from underway ships are becoming widely available to the research and operational communities [11]. A number of underway data stewardship initiatives are taking advantage of netCDF as the primary data storage and transport format. Two projects – the Shipboard Automated Meteorological and Oceanographic System Initiative [12] and the Global Ocean Surface Underway Data Project[16] – evolved out of the data assembly center system established during the World Ocean Circulation Experiment (WOCE), prior to the development of the CF standard. Both GOSUD and SAMOS are committed to using netCDF and both are migrating towards the CF conventions.

GOSUD is in the final stages of approving a CF-based netCDF file format[17]. The current SAMOS data format has minor deviations from CF regarding units[18] (e.g., wind directions referenced to vessel heading). These minor differences will be reconciled with the CF units conventions. The use of netCDF-CF will help to facilitate the desired expansion of underway observations.

3.7 High Frequency (HF) Radar Data Network

HF radar is recognized as a cost-effective solution to augment in situ coastal ocean surface current measurements and to provide increased spatial and temporal resolution. In the United States the Integrated Ocean Observing System (IOOS) has for several years been supporting an effort to integrate an existing, loosely-organized network of over 100 radar sites, distributed (unevenly) along the U.S. coast line. Through the IOOS-funded effort consistent data standards and best practices are being applied at all sites, and the data are brought together into consolidated near-realtime coastal current grids. These grids represent the integration of both HF radar observing assets and other IOOS surface current measurements.

The gridded fields are managed as standard netCDF-CF datasets and described with ISO-19115 metadata. Since October 2009 NOAA's National Data Buoy Center has served these data via OPeNDAP, providing both DAP and OGC/WCS access. The radial velocity fields from individual sites, a lower level product that is of use for specialized applications involving data assimilation and optimal interpolation, are not presently available

---

[14] http://www.seadatanet.org/

[15] http://www.oceansites.org

[16] http://www.ifremer.fr/gosud/

[17] http://www.ifremer.fr/gosud/doc/gosud-dm-user-manual-08-064.pdf

[18] http://samos.coaps.fsu.edu/html/docs/samos_netcdf_manual.pdf

through netCDF-CF-DAP. Plans are underway to develop netCDF representations for this data. by the German Federal Maritime and Hydrographic Agency (Bundesamt für Seeschifffahrt und Hydrographie)

4. THE GLOBAL TELECOMMUNICATIONS SYSTEM (GTS)

The WMO Global Telecommunications System (GTS) predates the Internet by several decades. The GTS is the communication and data management infrastructure of WMO World Weather Watch. It is described through a number of WMO Manuals (e.g. No. 2, 9, 49, 306, 386) which also contain metadata of the data transmitted. The structure of the GTS is outlined in Fig 1. GTS hardware and software standards were developed in support of operational meteorology, which remains its primary focus -- providing the communication capabilities required to guarantee efficient collection and distribution of observations and processed products. GTS supports WMO programmes beyond operational meteorology based upon directives from the WMO Executive Council.

The primary focus of the GTS is the delivery of real-time data from the many far-flung locations at which they are measured or generated, to recipient sites, whose primary activity historically has been operational weather forecast. The task of winnowing the myriad messages flowing on the GTS in order to identify data of possible interest falls to the recipient. (GTS is often referred to in the vernacular as a "firehose" of data.) Each formatted message on the GTS is identified by a standardized header. Winnowing of messages is generally accomplished by software filters that examine these headers.
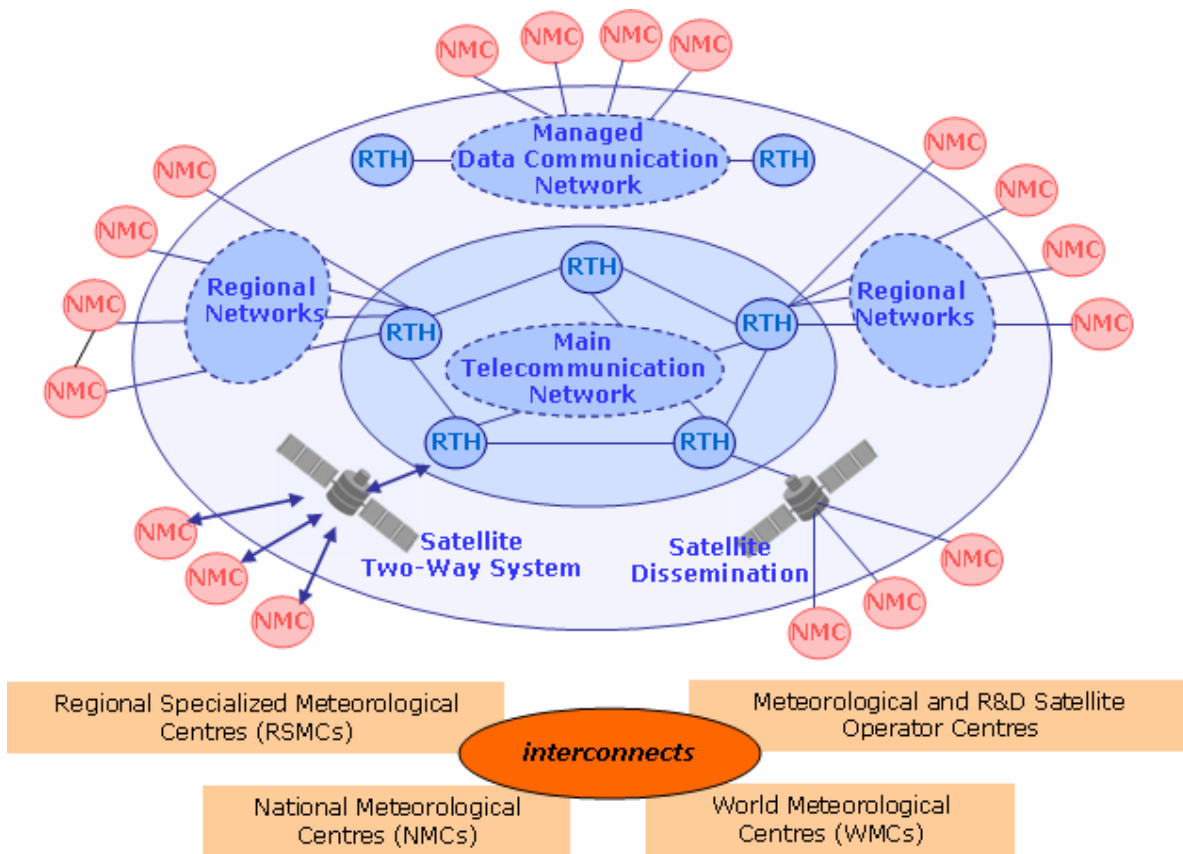

Figure 2. Structure of the WMO Global Telecommunication System[19]

---

[19] http://www.wmo.ch/pages/prog/www/TEM/GTS/index_en.html

The earliest GTS data format standards were alphanumeric code forms, typical of information technology of the time (1951). Each type of data -- a ship observation, time series report, vertical profile, etc. -- was associated with its own message type. Each message type had a distinct format. To meet ever more demanding operational needs Table-Driven Code Forms such as BUFR (Binary Universal Form for the Representation of meteorological data) and GRIB2 (GRIdded Binary, Edition 2)[20], were introduced in 1998 and 2000, respectively. In 2002, the WMO Commission for Basic Systems (CBS) approved the wholesale migration from alphanumeric codes to BUFR for observational data. For maritime codes, this process should be completed by 2012. JCOMM is overseeing efforts to ensure that the metadata encoded into BUFR will be far more detailed and complete than it has been in the past, and address the expanding needs of diverse international user communities, including long term climate analysis. Both BUFR and GRIB are rather complex file formats, reflecting the operational requirements of GTS. Bandwidth is a major consideration for GTS communications, so GTS file formats (in contrast to NetCDF binary and in starker contrast to most XML encodings) are designed to maximize transmission efficiency of entire files.

GTS is a "data push" system operating on a private network. The data that flow are determined by the providers of the data. Data exist on the GTS only for the moment that they are flowing. The Internet on the other hand is a "data pull" system in which the recipient instigates the flow of data. Internet servers hold the data for arbitrary periods of time on behalf of their potential users. WMO is currently developing the WMO Information System (WIS) which will utilize GTS push as well as push and pull through the Internet. WIS will allow WMO to meet top-level objectives as described in the WMO Strategic Plan, promoting the use of standards, focusing on a wider user community, and implementing data discovery, access and retrieval services, among other characteristics.

Operational communities like WMO must be technologically cautious, as reliable services are a strict requirement. When compared to the standards-development procedures followed by the CF or OPeNDAP communities, the steps that must be followed to advance a change in the GTS standards are highly formalized and slow. Even changes that follow

---

[20] The WMO's Expert Team on the Assessment of Data Representation System (ET/ADRS) has been recently established to discuss alternative data formats for use in real time international data exchange. The recommendations of the ET/ADRS urge WMO to cooperate with the netCDF and CF communities and develop a model of use and governance of these standards.

"Fast Track" procedures (those not requiring software modifications) require a 4-step process of (i) validation by multiple committees; (ii) approval for pre-operational stage by the same; (iii) endorsement by the CBS; and (iv) endorsement and validatation by the Executive Council.

5. CONCLUSIONS

The central question for any standards process should always be, "Does it produce standards that achieve their intended goals?" By this criterion, the WMO standards process is highly successful. The GTS supports a vast infrastructure of global weather observations and forecasting. It has continued to function at the highest level of operational stability, even as changes have been introduced. The WMO standards process is highly controlled -- "top down". Although opportunities for exploiting superior technologies are in some cases missed, so too are the technological blind alleys that would have compromised the operational missions of the WMO. Adopting standards is fundamentally about managing risk [1]. Building interoperability through incrementally enhancing established standards should be understood as an approach that minimizes risk.

The netCDF-CF-DAP standards processes, with greater speed and far less formality, also continue to achieve their intended purposes. NetCDF, CF and OPeNDAP are all products of the scientific research communities in oceanography, meteorology and climate sciences. The growth in the usage of these standards is attributable to their ability to meet the needs of the research scientists and scientific software developers in these fields. These standards exemplify the "bottom up" process by which de facto standards grow -- competing successfully in the marketplace by meeting the needs of their users. The recognition of these standards continues to grow. The trio of netCDF-CF-DAP was elevated to the status of "Recommended Standard" for gridded data interoperability by the IOOS Data Management and Communications (DMAC) Steering Team in December 2008 [13]. NASA has endorsed both netCDF and OPeNDAP as standards (ESDS-RFC-011, and ESDS-RFC-004, respectively [14]). The formal process to advance these technologies through the OGC standards process was initiated at the OGC Technical Meeting held in Mountain View, California in December 2009.

The contrast between the netCDF-CF-DAP and WMO standards processes is less stark than it may at first appear. While the final adoption processes differ greatly – a community consensus versus a top down mandate – both the netCDF-CF-DAP and WMO standards processes share the following key characteristic: changes to the standards are thoroughly vetted through testing in situations of realistic complexity before they are adopted. This is arguably a defining characteristic of most highly successful IT

standards processes. It is a key characteristic that has guided the much esteemed Internet Engineering Task Force (IETF) that has enabled it to produce so many of the interoperability standards that we depend upon today for web browsing, file transfer, and email, amongst others. The contrasting process of developing standards through committee processes often results in standards that have had inadequate testing in situations of realistic complexity [2]. This weakness has been the downfall of many standards.

While netCDF-CF-DAP and GTS demonstrably meet their design goals, they fail to address many of the important data management requirements that we recognize today, such as retaining detailed background metadata; cataloguing and discovery of datasets; provenance and version tracking; etc. These omissions reflect the recent, rapid expansions in perceived requirements. Technologies that address these emerging needs are under development, and some of them show great successes in other communities. Responding to society's pressing needs for improved marine data interoperability, there is a strong temptation to adopt these promising technologies as soon as possible.

The authors of this paper believe that the pragmatic and sceptical approach leads to the best investments over the long term. Making commitments to unproven standards in the face of rapidly changing technologies is a form of gambling: investing today's limited resources in a technology lottery in hopes of a "big win". The history of IT standards has shown us time and again that there is seldom a big win from taking this approach. Rather there is a pronounced risk of confusion and setbacks through making premature commitments [2,15]. In parallel with this cautious approach to community standardization, there should be vigorous experimentation, testing and refinement of promising new technologies. Such research and development efforts should be conducted in project-specific contexts, and should be allowed to grow in scope only after a careful evaluation of their merits.

The authors of this paper believe that interoperability will not be achieved until the entire ocean science community shares in the responsibility. Program managers and scientists play roles in defining the path to interoperability that are as important as those who develop the data systems. All community leaders must apply the "look before you leap" principles of pragmatism and scepticism that are articulated in this paper. And all community leaders have a role to ensure that data management – especially community-wide infrastructure efforts -- is funded sufficiently to meet growing expectations. In particular the ocean observations community should provide its share of financial support to the continued advancement of

netCDF, CF, OPeNDAP and the associated software tools upon which we depend.

6. REFERENCES

1. Cargill, C. and Bolin, S. (2004). Standardization: A Failing Paradigm, http://www.chicagofed.org/news_and_conferences/conferences_and_events/files/cargill.pdf (paper presented at the Standards and Public Policy Conference, Federal Reserve Bank of Chicago, May 13-14)

2. Henning, M. (June 2006). The Rise and Fall of CORBA, ACM QUEUE, *http://www.zeroc.com/documents/riseAndFallOfCorba.pdf*

3. Hartnett,E.J. and Rew, R.K. (2008). Experience with an enhanced netCDF data model and interface for scientific data access, *24th IIPS, American Meteorology Society, New Orleans, Louisiana*, http://www.unidata.ucar.edu/netcdf/papers/AMS_2008.pdf

4. Jonathan G. (2003) The CF metadata standard, http://cf-pcmdi.llnl.gov/documents/other/cf_overview_article.pdf\

5. Hankin, S.C., Harrison, D.E., Osborne, J., Davison, J., and O'Brien K. (1996) A strategy and a tool, Ferret, for closely integrated visualization and analysis. *J. Visual. Comp. Animat.*, 7, pp149–157

6. Doty, B. and Kinter, J.L. III (1992). The Grid Analysis and Display System (GrADS): A practical tool for Earth science visualization. *Eighth International Conference on Interactive Information and Processing Systems (Dallas TX, 5-10 January, 1992*

7. Caron, J. (2008) The NetCDF Markup Language, *http://www.unidata.ucar.edu/netcdf/ncml/*

8. Signell, R.P. Carniel, S., Chiggiato, J., Janekovic, I., Pullen, J., and Sherwood, C.R. (2008). Collaboration tools and techniques for large model datasets. *Journal of Marine Systems* Volume 69, Issues 1-2, pp154-161, doi:10.1016/j.jmarsys.2007.02.013

9. Blower, J.D., F. Blanc, Clancy, M., Cornillon, P., Donlon, C., Hacker, P., Haines, K., Hankin, S.C., Loubrieu, T., Pouliquen, S., Price, M., Pugh, T.F., and Srinivasan, A. (2009). Serving GODAE Data and Products to the Ocean Community, *Oceanography*, 22(3), 70-79.

10. *Gridspec: A standard for the description of grids used in Earth System models*, accessed on 21 December 2009 at http://www.gfdl.noaa.gov/~vb/gridstd/gridstd.html

11. Smith, S. R., and coauthors (2010). Automated Underway Oceanic and Atmospheric Measurements from Ships. In these proceedings (Vol. 2).

12. Smith, S. R., and coauthors (2010). The Data Management System for the Shipboard Automated Meteorological and Oceanographic System (SAMOS) Initiative. In these proceedings (Vol. 2).

13. Standards package for the representation and transport of gridded data: netCDF+CF+OPeNDAP+aggregation, accessed on 21 December 2009 under Recommended Standards at http://ioosdmac.fedworx.org/

14. *Earth Science Data Systems Standards Process Group*, accessed on 21 December 2009 at http://www.esdswg.org/spg/docindexfolder

15. Hogan, M. and Radack, S. (1997). The quest for information technology standards for the global information infrastructure, *StandardView, Association for Computing Machinery (ACM)*, Volume 5, Number 1, March