World Meteorological Organization

United Nations
Educational, Scientific and
Cultural Organization

Intergovernmental
Oceanographic
Commission

WORLD METEOROLOGICAL ORGANIZATION

INTERGOVERNMENTAL OCEANOGRAPHIC COMMISSION (OF UNESCO)

JCOMM DATA MANAGEMENT PLAN

Prepared by the Members of the Data Management Coordination Group

2015

WMO/TD-No. 1426

JCOMM Technical Report No. 40, Revision 2

[page left intentionally blank]

WORLD METEOROLOGICAL ORGANIZATION

INTERGOVERNMENTAL OCEANOGRAPHIC COMMISSION (OF UNESCO)

JCOMM DATA MANAGEMENT PLAN

Prepared by the Members of the Data Management Coordination Group

NOTES

WMO DISCLAIMER

Regulation 42

Recommendations of working groups shall have no status within the Organization until they have been approved by the responsible constituent body. In the case of joint working groups the recommendations must be concurred with by the presidents of the constituent bodies concerned before being submitted to the designated constituent body.

Regulation 43

In the case of a recommendation made by a working group between sessions of the responsible constituent body, either in a session of a working group or by correspondence, the president of the body may, as an exceptional measure, approve the recommendation on behalf of the constituent body when the matter is, in his opinion, urgent, and does not appear to imply new obligations for Members. He may then submit this recommendation for adoption by the Executive Council or to the President of the Organization for action in accordance with Regulation 9(5).

Chairperson, Publications Board
World Meteorological Organization (WMO)
7 bis, avenue de la Paix
P.O. Box No. 2300
CH-1211 Geneva 2, Switzerland

Tel.: +41 (0)22 730 84 03
Fax: +41 (0)22 730 80 40
E-mail: Publications@wmo.int

IOC (OF UNESCO) DISCLAIMER

CONTENTS

[page left intentionally blank]

JCOMM DATA MANAGEMENT PLAN

1. THE VISION AND OBJECTIVES OF JCOMM

The stated vision of the Joint WMO-IOC Technical Commission for Oceanography and Marine Meteorology (JCOMM) is to coordinate, regulate and facilitate, at the global level, a fully integrated marine observing, data management and services system that uses state-of-the-art technologies and capabilities; is responsive to the evolving needs of all users of marine data and products; and includes an outreach programme to enhance the national capacity of all maritime countries. JCOMM aims to maximize the benefits for its Members / Member States in the projects, programmes and activities that it undertakes in their interest and that of the global community in general. For information about JCOMM see http://www.jcomm.info.

The long-term objectives of JCOMM according to the JCOMM Strategy document (currently under review; excerpt from the January 2014 version below) are:

(i) **To enhance the provision of marine meteorological and oceanographic forecasting and analysis services** in support of the safety of life and property at sea and in coastal areas; contribute to risk management for ocean-based economic, commercial and industrial activities; contribute to the prevention and control of marine pollution, help to sustain healthy and productive oceans, develop integrated coastal area management services for recreational activities and the safety of coastal settlements and activities ;

(ii) **To contribute to the development, enhancement and delivery of climate services related to the marine atmosphere and coastal and deep oceans**, based on the core competencies within the Commission in marine meteorology and oceanography, as a contribution of JCOMM to the Global Framework for Climate Services (GFCS) and to coordinate and enhance the provision of the data, information, products and services required to support climate research and the detection and prediction of climate variability;

(iii) **To coordinate the enhancement and long-term maintenance of an integrated global marine meteorological and oceanographic observing and data management system**, containing both in situ and remote sensing components and including data communication facilities, in the most cost-effective and efficient way, as part of the IOC-WMO-UNEP-ICSU Global Ocean Observing System (GOOS) and the World Weather Watch (WWW), and in support of the GFCS, the WMO-IOC-ICSU World Climate Research Programme (WCRP), and the WMO-IOC-UNEP-ICSU Global Climate Observing System (GCOS), and other major WMO and IOC Programmes. This system is contributing to the WIS and the International Oceanographic Data and Information Exchange (IODE) Ocean Data Portal (ODP), and will be complying with the requirements of the WMO Integrated Global Observing System (WIGOS);

(iv) **To manage the evolution of the services through the selective incorporation of advances in meteorological and oceanographic science and technology**. The incorporation will be facilitated by the active participation to international system development research groups, such as the GODAE OceanView, and engaging experts from these communities in the JCOMM expert teams;

(v) **To promote and facilitate the equitable participation of all WMO Members and IOC Member States** in all activities of, and benefit from all products and services provided by, JCOMM. to work to enhance capacity development in the field of marine meteorology and oceanography information and services and to ensure that all countries are allowed to benefit from and contribute to these advances, and to contribute to the work of JCOMM in general;

2. PURPOSE AND SCOPE OF THIS PLAN

JCOMM was established by the World Meteorological Organization (WMO) and UNESCO's Intergovernmental Oceanographic Commission (IOC) in 1999 to be their major advisory body on all technical aspects of operational marine meteorology and oceanography. Prior to this, marine meteorological and oceanographic observations, data management and service provision programmes were internationally coordinated by two separate bodies: IOC working jointly with WMO through the Committee for the Integrated Global Ocean Services System (IGOSS), and WMO through its Commission for Marine Meteorology (CMM). The IGOSS dealt with real-time oceanographic data (defined as data collected within the last 30 days) and managed physical oceanographic variables only – hence the perception on the oceanographic side that JCOMM deals in real-time data only. On the meteorological side, CMM's work covered the complete time frame from real-time to delayed-mode (data not distributed in real-time and usually of higher resolution and quality) and built and maintained archives of marine data. This current JCOMM Data Management Plan (DMPlan) must address issues relevant to both real-time data handling as well as managing delayed-mode data in archives.

The present structure of JCOMM (in 2012) has three Programme Areas (PAs), one for Observations (OPA), one for Services and Forecasting Systems (SFSPA) and one for Data Management (DMPA). The groups in the OPA and SFSPA are focused on activities supporting a type of data (or sometimes an observing technique), or on a service such as safety or emergency support. Each of these has its own history of managing data and information prior to its incorporation into JCOMM. When JCOMM was formed, a decision had to be made about how to organize the cross-cutting activities of data management. The choice was made to place data management in a separate PA to recognize that managing the data and information of JCOMM is an important activity equal to acquiring and delivering data and services. The potential weakness is that the activities of the DMPA may not be strongly linked to the day-to-day data management activities in the various groups of the other PAs. The challenge of the DMPA is to work within the requirements of the activities within the OPA and SFSPA and still achieve the broad goals of JCOMM. This plan will adopt an approach that looks for commonalities across all of these systems and exploits these to improve interoperability.

A main goal of this DMPlan, therefore, must be to explain how data management can be conducted under the present structure to promote the long-term objectives of JCOMM.

JCOMM deals in a variety of data within the broad domains of oceanography and marine meteorology. Both meteorology and physical oceanography have a strong history of data exchange, and it is these types of data that are normally considered part of JCOMM activities. Biological or chemical variables have a history of data exchange within oceanography but only in delayed-mode, and only for a limited number of variables. Only recently have these kinds of data been exchanged in real-time, such as in the International Ocean Carbon Coordination Project. The coastal module of GOOS has defined common variables to be exchanged and more than half of these are outside of the physical oceanographic domain. In this context, JCOMM must position itself to handle this broader range of variables.

New observing technologies are being developed at a fast pace. In addition to the improvements in measuring traditional physical oceanographic variables, such as temperature and salinity, there are sensors being built that can provide immediate and reliable measurements of chemical and biological components in the ocean. These data can be transmitted instantly through satellite systems. New offshore cabled networks allow for the streaming of data of all kinds, from television images, acoustics or more conventional oceanographic and meteorological observations with 2-way communication to the sensors. Open ocean moorings with both meteorological and oceanographic measurements available immediately are being deployed. All of these data will be challenging for JCOMM to both coordinate and manage.

Computer modeling of the atmosphere has been an important activity for many years. In the last few years, modeling of the ocean has increased substantially. Now, ocean and atmospheric models are starting to be fully coupled. Ice modeling is being injected into these coupled models and there are some developments to include biological and chemical components in the ocean as well.

Model results are valuable for forecasting and in hindcast studies as well. Models permit us to identify data gaps, propose strategies to fill in those gaps, and to predict conditions where data are sparse. When these results are reformulated as products, they can become particularly useful in decision making, disaster mitigation and a host of other uses.

Models can be operated in research, operational or reanalysis-mode. Models run for research purposes are constantly being checked, results verified against observations, model characteristics altered and so on until such time as the newer version is determined to be an improvement on an existing operational model. Results from research model runs are of use to the research community primarily.

In contrast, the characteristics of operational, and to an even larger extent, reanalysis models are held stable. Results from these models are the basis for many products because they use the physics in the models to fill in information in oceanic and atmospheric areas where observations are not available. These results are valuable to save and can be used in a similar way as are historical observations.

Satellite observations are also of interest to JCOMM. Satellites provide the synoptic and broad scale views that are unattainable from in situ observing systems. They are a complement to the in situ systems in that they provide surface conditions on broad spatial scales at an instant of time. There is already a well developed international system for managing satellite data (see http://www.ceos.org/, the home page of the Committee on Earth Observation Satellites (CEOS). The DMPlan needs to take into consideration the level of interoperability that is required with CEOS and how this can be attained.

Metadata must also be considered as an important component of JCOMM. Metadata is a term used to cover a wide range of information. It may be information that describes the contents of archives (such as what data they contain, over what time and space scales) down to detailed information about characteristics of the instrumentation, placement of sensors, or characteristics of the models. The terminology to talk about these different kinds of metadata has sometimes been confusing in the past, so this document will provide examples and guidance in this regard.

Metadata are important for a number of uses. Just as the scope of metadata is wide, so its uses are broad. For example, information about the contents of archives is used in cataloguing systems so that potential users can locate data of interest. Information about instrument characteristics or sampling schemes is important in comparing measurements from different instruments to ensure that systematic differences are taken into consideration.

The DMPA is not alone in addressing issues of managing oceanographic and meteorological data in the international arena. On the oceanographic side, the Intergovernmental Data and Information Exchange (IODE, http://www.iode.org/) committee of IOC has operated for many years managing many different kinds of data including types common with JCOMM. The difference has been that IODE has mostly concentrated on data that arrive in the data system with significant time delays, some that may be up to years while the management of the real-time data was left to IGOSS. The IODE is a close partner in managing the oceanographic data and is a co-sponsor of many of the data management activities of relevance to JCOMM.

A more recent initiative of WMO, thus far advanced largely through its Commission for Basic Systems (CBS), is its WMO Information System (WIS) (see web link for further details: http://www.wmo.int/wis). This is an overarching approach and a single coordinated global infrastructure for the collection, distribution, retrieval of, and access to data and information of all

WMO and related programmes. JCOMM, as a technical commission being co-sponsored by WMO, is a contributor to the WIS.

Both oceanographic and meteorological data contribute to the holdings within the World Data System (WDS) (see http://www.icsu-wds.org/). It is expected that the WDS will ultimately archive all of the data collected and managed by JCOMM. Therefore, the JCOMM data management activities need to provide data and information to the WDS members and work with them to build a complete global data system.

The creation of the Global Climate Observing System (GCOS) and all of its components views JCOMM as an important contributor. The 2010 update of the GCOS Implementation Plan (GCOS-138) (see http://www.wmo.int/pages/prog/gcos/Publications/gcos-138.pdf) cites 25 specific actions requiring the support of JCOMM as an 'Agent for Implementation'. Of these actions, some relate directly to the data systems. The DMPlan must provide the direction that will ensure these expectations are met.

Details on how JCOMM should link into these various programmes are provided later in this document. The initial sections of this plan discuss what activities JCOMM should undertake to ensure that the data collected under its Programme Areas are well coordinated and managed.

Finally, it is obvious that the data management component of JCOMM is broad and has to make many connections both within and outside of the two parent organizations of WMO and IOC. This DMPlan provides the broad outlines and recommendations by which the DMPA will help attain the vision of JCOMM. This plan will not provide the details of how the recommendations will be met; rather, this is the subject of an implementation plan that must be built from the agreed plan. As technology advances, and as the implementation develops, there will be changes in emphasis or new capabilities not anticipated by this DMPlan.

3. ORGANIZATION OF THE PLAN

An examination of the long-term objectives of JCOMM and expectations for action by JCOMM, such as from the GCOS Implementation Plan (GCOS-IP), require that:

- there exists a functioning system of reliable and regular observations at sea;
- the data and information come to processing centres in a timely way;
- notifications of hazardous conditions are issued to mariners or nations in time to take action to avoid potential harm;
- data collected by JCOMM activities be maintained over many years such that climate variability, trends, and prediction can be studied and advanced;
- information be maintained about the observing practices so that older data may be compared to more recent data;
- there be standardization in such areas as data formats, content, naming conventions, processing procedures, etc., to ensure interoperability;
- data version control be addressed; and
- data management activities and experiences are made available equally to all WMO-IOC Member / Member States.

To organize the discussion, this document will divide the tasks into main areas and make recommendations in each. The major data management themes that are used are:

- Data and Information Exchange – includes issues of transporting data to and between archives or processing centres;

- Data Processing – includes issues of data quality assessment, version control, content, etc.;

- Access – includes issues of finding, browsing, and moving data and information to users;

- Coordination and Linkages– includes issues of how activities in the different PAs need to link together as well as the links between JCOMM and other organizations; and
- Communications – including the dissemination of information about JCOMM data management, training materials, performance measures, reports, etc.

Interoperability is a central issue for JCOMM because of the many contributors to data collection, processing and distribution. This is an issue that cuts across all of the sections of the DMPlan. A heavy emphasis is evident here on the adoption of standardized procedures in all areas of data management. It is only through the adoption of standard practices that the required level of interoperability will be attained.

4. DATA AND INFORMATION EXCHANGE

This section deals with the various aspects of moving the observational data collected at sea to the appropriate archive or data distribution centres. Most of the data coming ashore from instruments do so in data structures driven either by the instrument manufacturers or by telecommunications demands. Once ashore, the data are converted to formats used for data exchange in real-time (up to 30 days old for oceanographic data, a few hours old for meteorological data) or a less timely exchange referred to as delayed-mode. Real-time data exchange normally uses the GTS (Global Telecommunications System) with relatively few and well controlled formats. In delayed-mode, there are many formats and the communications channel is increasingly the Internet.

4.1 From Collectors to the Shore

Operations at sea are strongly challenged by the size of the world's oceans, by harsh conditions, by scarce power for instrumentation and by constrained communications capabilities. Because of these, measurements at sea are difficult to obtain and often are limited in geographic extent and to short periods of time. In moving data ashore (often via satellites), a premium is placed on compact data formats that squeeze the most information into the smallest message length. Consequently, the data streams that come ashore are strongly linked to the types of instruments used and so manifest a wide variety of data formats. Individual processing systems have been developed to manage these data streams.

While there is some hope that limitations on communications bandwidth will ease in the near future, the trend is to make even more varied observations at sea. So, although bandwidth will increase, the quantity of measurements will also. Without the adoption of some standard for reporting from platforms at sea, there will continue to be a variety of data formats.

The WMO and IOC are establishing an International Forum of Users of Satellite Data Telecommunications Systems (Satcom Forum). . The future Forum is meant to provide an international mechanism, covering the wide user base that exists within the co-sponsoring Organizations, to address remote data communication requirements – including tariff negotiations as needed – for automatic environment observing systems using satellite data telecommunication systems (Satcom systems).

Both the WMO and IOC have sought for decades to standardize communications of data, largely in the context of reporting data over the GTS. An avenue that has not been widely explored is to use these same standards, or others, for reporting directly from the instruments at sea.

Whilst it will be difficult to find acceptance for new standards for already running systems and procedures, it would be a beneficial step forward if all new sensor and submission developments would be based on standards such as defined by the Sensor Observation Service (SOS). The offered sensor data comprises descriptions of sensors themselves, which are encoded in the Sensor Model Language (SensorML), and the measured values in the Observations and Measurements (O & M) encoding format. These formats are open standards and specifications of the same name defined by the Open Geospatial Consortium (OGC).

Recommendation 4.1: JCOMM should encourage instrument manufacturers to standardize the formats of the data and information coming from instruments used at sea.

4.2 Using the Global Telecommunication System (of WWW) (GTS)

The GTS (see http://www.wmo.int/pages/prog/www/TEM/GTS/index_en.html), is one of the means of data exchange used by oceanographers and meteorologists. For WMO, it is the transmission mechanism of choice for operational, time-critical data exchange. This is still true in the development of the WIS. Being co-sponsored by WMO, JCOMM will need to adopt this same view.

Historically, Traditional Alphanumeric Codes (TACs) have been used on the GTS to report oceanic and atmospheric marine data. These have had strongly regulated formats and contents (e.g. FM-13 SHIP) with the advantage of a commonly understood protocol for naming variables, for units of reporting and additional information to be sent with the observations. However, in order to accommodate the needs of and to represent the rapidly evolving observing system a series of Table Driven Codes (TDCs) have been developed within the WMO. These TDCs are similarly strongly regulated but also self describing, flexible and expandable.

For the *in situ* observational data the preferred method of transmission on the GTS in TDC is the Binary Universal Form for Representation of meteorological data (BUFR, FM-94. See http://www.wmo.int/pages/prog/www/WMOCodes.html). To facilitate the migration to BUFR the TACs have been frozen and there has been a phased transition with parallel distribution in both TDCs and TACs on the GTS. This parallel distribution is due to end for marine data in November 2014, with only TDCs permitted to be transmitted over the GTS after this date. It should be noted that whilst BUFR is the preferred method of transmission on the GTS it is also possible to represent the contents of BUFR messages in text form using the Character form for the Representation and EXchange of data (CREX, FM-95).

As an example of the flexibility of BUFR, a BUFR message defines the data to be reported at the start of the message using a set of descriptor tables followed by the data itself. The descriptor tables allow most, if not all, physical meteorological and oceanographic parameters to be reported together with metadata on the observations. This metadata can include parameters such as the time period over which an observation has been made, the direction of an oceanographic profile (i.e. whether the observations were made on the up or down cast), the height or depth of the observation and the platform making the observation. It is also possible to report limited information on the instrument types used to make the observations.

In order to simplify and reduce the size of the description section of BUFR messages it is possible to nest the data descriptors. For example, BUFR allows the definition of descriptors that in turn define a sequence of other descriptors (I.e. BUFR Table D sequences) for common observation types. These Table D sequences are similar in concept, but not as generic, to the Extensible Markup Language (XML) ‘bricks’ previously proposed and can be used to build a template for reporting different observations. For example, the equivalent to FM-13 in BUFR can be described by a single BUFR descriptor that in turn references other BUFR descriptors. The BUFR templates can also include guidelines or regulations on how those Table D sequences should be used (e.g. see http://www.wmo.int/pages/prog/www/WMOCodes/TemplateExamples.html). However, it should be noted that there is no requirement to use the predefined sequences or templates and that it is possible to report observations from new observing platforms quickly and easily using the existing descriptors.

The set of tables that describe the contents of a BUFR message, essentially a controlled vocabulary, are strongly regulated. Changes to the code tables have to be proposed to the WMO Commission for Basic Systems (CBS) via the Inter-Programme Expert Team on Data Representation, Maintenance and Monitoring (IPET-DRMM), with any approved changes subsequently validated before they can be implemented. The units used for the parameters reported are generally restricted to the SI (International System of Units see http://physics.nist.gov/cuu/Units/index.html) units, but with a number of exceptions for commonly used non SI units. A full list of the permitted units can be found in the BUFR code table C6 http://www.wmo.int/pages/prog/www/WMOCodes/WMO306_vI2/VolumeI.2.html). An approximate timescale for proposing changes to the BUFR tables, through to the changes being implemented is a minimum of 1 year. For complicated changes or templates longer time scales should be expected.

Within JCOMM, and through collaboration with the WMO[1], a number of BUFR templates to replace the currently used TACs have been developed and validated for marine data using the current BUFR tables. Examples include the templates for XBT (Expendable Bathythermograph) and CTD (Conductivity, Temperature, and Depth probe) data, available from http://www.wmo.int/pages/prog/www/WMOCodes/TemplateExamples.html. At present, these templates are limited to the old TAC FM code forms. Consequently, these templates will need expanding to represent new technologies, parameters and observing system components as and when they start being used. In this regard, JCOMM must remain a partner in the process of migrating from TACs to TDCs. In the simplest case, JCOMM must actively work to build capability for exchanging data in TDCs, including BUFR. The work will entail such tasks as:

- maintaining, developing and validating appropriate BUFR templates to represent the needs of the oceanographic community;

- continuing to encourage ocean and meteorological centres to develop capacity to both read and write TDCs; and

- consideration of how TDCs may be used to acquire data from instruments and platforms at sea.

Recommendation 4.2a: DMPA lead the development of the detailed plan to change GTS data reporting from TACs to TDCs.

BUFR allows the discipline based definition of the code tables, with the tables grouped into a series of Master Tables. For meteorology, the underlying BUFR tables are grouped into Master Table 0 (MT0). MT0 also allows for the reporting of the physical oceanographic parameters and the current BUFR templates for oceanographic data make use of MT0. In order to expand the parameters it is possible to report and to better represent the needs of the oceanographic community Master Table 10 (MT10) has been developed for oceanographic data. This includes classes for the representation of biological, chemical and geophysical elements as well as the standard meteorological and physical oceanographic elements. However, due to the lack of capability to process BUFR within ocean centres BUFR MT10 has not been in operational use and has remained in the development stage. However, with encouragement from JCOMM to get both

---

[1] The templates have been developed through the DMPA and the OPA (through SOT and DBCP) with advice from the IPET-DRC (now IPET-DRMM)

ocean and meteorological data reported in BUFR, MT10 should be revisited and evaluated against current needs (for example, one can question the need of MT10 if we can open up a another oceanographic class within BUFR Master Table 0 (MT0)).

Recommendation 4.2b: The DMPA in association with the appropriate WMO committee should evaluate MT10 for its relevance to present and future needs.

The advantage of BUFR, and other TDCs, is the set of tables (termed “classes” in BUFR) that designate variables and attributes, in a machine-readable form. These tables constitute the vocabulary of BUFR. While it is certainly advantageous in that the meanings are well defined, the construction of BUFR causes some unpleasant side effects. One such is that a BUFR variable is characterized by the number of bits used to express the value. The same variable, such as sea temperature, may have more than one BUFR variable assigned, since sea temperature can be recorded to 1, 2 or 3 decimal places. Each needs a different number of bits to express the value and so each gets a different BUFR designator. A second issue is that because the data are in binary, and different computer operating systems have different ways of handling binary data, different software routines are needed for the different operating systems. So, there are a number of versions of BUFR encoding and decoding software and this challenges them all to produce identical results. These shortcomings, along with the strengths of BUFR need to be considered in discussions of metadata and vocabularies described later. It should be noted that similar problems will exist for most other data formats.

Recommendation 4.2c: Enhanced interaction between JCOMM and CBS or other appropriate WMO committees is needed to expand the scope of TDCs to more fully incorporate JCOMM considerations, and the archival and exchange of historical and delayed-mode data in its originally reported form

4.3 Using the Internet

Data are also exchanged using other telecommunication systems, notably the Internet. For these exchanges, there is no standard for naming variables and attributes, no universally agreed structures or formats, and in fact no real order at all, beyond the broad constraints of standards such as the Hypertext Transfer Protocol (HTTP) and the File Transfer Protocol (FTP). Use of the Internet is very widespread and this lack of order makes the exchange of data a complicated process because every data provider must give detailed information on their formats, contents, processing steps, etc., and receivers need to build software capability to handle the wide variety from different data providers.

4.3.1 Network Common Data Form (netCDF)

There are preferred practices that are starting to emerge for data exchange using the Internet. The use of netCDF (see http://www.unidata.ucar.edu/software/netcdf/) for the exchange of in situ ocean data is increasingly prevalent. Its use started in earnest during the World Ocean Circulation Experiment, WOCE, in the 1990s and is now a part of the Argo (see http://wo.jcommops.org/cgi-bin/WebObjects/Argo, and OceanOBS’09 Community White Paper (CWP) from Pouliquen *et al* on Argo Data Management[2]), OceanSITES (OCEAN Sustained Interdisciplinary Timeseries Environment observation System , see http://www.oceansites.org/, and OceanOBS’09 CWP on OceanSITEs[3]) and GOSUD (Global Ocean Surface Underway Data Pilot Project, see http://www.ifremer.fr/gosud/) programs. Today’s version of netCDF is most suitable for data that has some regularity in one or more of the horizontal, vertical or time coordinates, but it can be used even when this is lacking (Note that this limitation is being addressed in a new version to be issued soon). The weakness of netCDF is that there is no single standard for naming variables or attributes. There are common practices including the Climate and Forecast (CF - see http://www.cgd.ucar.edu/cms/eaton/cf-metadata/) conventions, but the use of netCDF for data

---

[2] https://abstracts.congrex.com/scripts/jmevent/abstracts/FCXNL-09A02a-1710033-1-CWP-1710033.pdf

[3] https://abstracts.congrex.com/scripts/JMEvent/abstracts/FCXNL-09A02a-1779458-1-CWP2B09.pdf

exchange would be greatly enhanced with the adoption of a standard vocabulary. Never-the-less, netCDF is in wide enough use that provision of data in this format should be considered by JCOMM.

Recommendation 4.3.1a: JCOMM to stay informed on netCDF maintenance and developments, and to support the widespread use of netCDF as a data exchange format.

Recommendation 4.3.1b: JCOMM to encourage usage of the CF convention for variable naming in netCDF and stay informed of CF updates to meet JCOMM contributors' needs.

Recommendation 4.3.1c: deprecated (merged with 4.3.1a).

4.3.2 Extensible Markup Language (XML)

XML is yet another way to structure data and information for exchange. To date, its main use has been in exchanging low volume data, though perhaps at high frequency. It is a very popular structure because of its flexibility, readability and the wide availability of software to parse messages and extract content. With the development of Service Oriented Architecture models, XML will only gain popularity.

The flexibility of XML is also one of its weaknesses. The meaning of the XML tags must be understood by both the sender and receiver of the message. This means that each tag must be defined; in effect the vocabulary must be established between sender and receiver, before the messages are exchanged. Until this vocabulary is defined, this is no better solution than using any other format with arbitrary names for variables.

Still, the commercial acceptance of using xml and hence the broad availability of software makes this an attractive option to consider for both data and metadata.

Through participation of IODE in the Ocean Data Interoperability Platform and other working groups, it has been noted that the BODC has been very proactive in its participation in these venues and its efforts to both expand the suite of vocabularies delivered through the NERC Vocabulary Server (http://www.bodc.ac.uk/products/web_services/) as well as to promote the use of the established code lists such as the BODC Parameter Codes (P21). BODC is currently providing access to the hosted vocabularies through a web portal and via a suite of web services, delivering results in formats including RDF XML. The BODC has also provided synthesized (filtered) views of the parameter codes list in support of specific domains, expanding the use of the vocabularies without creating new ones.

The use of JavaScript Object Notation (JSON) as a mechanism for encoding data for transit via the web has gained significant momentum and this should be considered as this encoding form is inherently more efficient than XML and is used regularly in data driven application use cases (e.g. read/parse/visualize data).

Recommendation 4.3.2a: Deprecated.

Recommendation 4.3.2b: DMPA to monitor the development of vocabularies and provide input as needed.

4.3.3 Other Formats and Data Structures

Of course, netCDF is not the only format used for exchanging data on the Internet (see OceanOBS'09 Community White Paper (CWP) from Hankin *et al* on NetCDF-CF-OPeNDAP[4] Standards for Ocean Data Interoperability and Object Lessons for Community Data Standards

---

[4] Open-source Project for a Network Data Access Protocol

Processes – Hankin *et al*, 2009, see website[5]). Other data structures are in use, but there is no coordination between these formats and so no standards have developed. Still, JCOMM must recognize that netCDF is not the sole way that users will wish to exchange data and therefore must continue to keep abreast of formats used and developed that offer broad scale appeal. It will be necessary as well to create mappings from one set of naming and format conventions to another, but at least with fewer formats, the mapping process will be easier. This should include, for example, mapping of BUFR names to netCDF CF conventions.

Recommendation 4.3.3a: JCOMM must recognize that other formats and data structures besides netCDF will have appeal and encourage activities that broaden their use and standardize their content.

At present, there is a distinction made between data that are exchanged in real-time and those exchanged in delayed-mode. Typically, the GTS and its suite of data formats are used for real-time, and a host of formats, netCDF being one, for exchange in delayed-mode. This distinction is artificial in that it is only because of limiting bandwidth of communications channels that the full resolution data cannot be sent as soon as the observations are taken. Artificial or not, there are a wide variety of exchange formats for delayed-mode data using whatever communications channels are available.

For example, standardized alphanumeric International Maritime Meteorological (IMM) formats were introduced by WMO around 1951 for the exchange of delayed-mode (e.g., keyed logbook data) from Voluntary Observing Ships (VOS). These were upgraded in 1982, and continue to be upgraded and used in the Marine Climatological Summaries Scheme (MCSS) today. The most recent development in this area is the International Maritime Meteorological Archive (IMMA) format, which is widely used by the research community to access VOS and many other marine data stored in the International Comprehensive Ocean-Atmosphere Data Set (ICOADS). Like the earlier IMM formats, IMMA is a relatively simple alphanumeric format, but with a more flexible and extensible design including a critical archive feature that permits attachment of the original (raw input) data. The IMMA format has been repeatedly revised and adapted to the latest requirements. Currently IMMA0 format (ICOADS 2010) is still in use for storing historical and contemporary marine data. Extensive improvements are now being implemented in prototype form and planned for final adoption as IMMA1 in 2014. A variety of changes are being implemented to bring IMMA into closer agreement with recent IMM Transfer Format IMMT. Furthermore the implementation of a near-surface oceanographic attachment is in development as a new component of the IMMA format. The continuous adaptation of existing agreed upon formats used in DACs, GDACs and CMOCs, including details of past data codes and formats is envisaged to be an important component of the new Marine Climate Data System (MCDS).

The delayed-mode formats have operated outside of the more formalized system governing evolution of the TDCs and freezing the form of the TACs. While that system helps assure the rigid levels of standardization required for mature operational applications, even small format changes in the TDCs therefore can take some time to be implemented. In contrast, the IMM (and some other delayed-mode formats) can be modified relatively easily (e.g., by JCOMM without required cross-Commission agreement), and this flexibility can have important advantages, at least in the research arena, in that the formats can be more immediately responsive to new developments in technology, and in data and metadata management.

Recommendation 4.3.3b: JCOMM work with partners to encourage the continuing evolution of exchange formats to more robust and stable forms, while at the same time assuring that sufficient flexibility and agility can be preserved for the archival of JCOMM's delayed-mode data and metadata.

4.3.4 Marine Climatological Data

---

[5] https://abstracts.congrex.com/scripts/jmevent/abstracts/FCXNL-09A02a-1707296-1-Hankin-cwp4c06.pdf

Marine climatology data (see OceanOBS'09 CWP from Woodruff *et al* on Surface in situ Datasets for Marine Climatological Applications[6]) are traditionally managed through the Marine Climatological Summaries Scheme (MCSS), established in 1963. While the MCSS is currently dealing essentially with Voluntary Observing Ship real-time (GTS) and delayed mode (collected through a network of Contributing Members, and two Global Collecting Centres (GCCs) in the UK and Germany) data, there is a need to consider other and new sources of ocean data (e.g. drifting and moored buoys, XBTs, Argo, rigs and platforms, tide gauges, satellite, etc.) to fully address the WMO and IOC applications requirements for appropriate marine-meteorological and oceanographic climatological data (met-ocean climate data), and particularly address those for long term climate monitoring (Global Climate Observing System – GCOS), seasonal to inter-annual climate forecasts, the Global Framework for Climate Services (GFCS), and ocean climate requirements of the Global Ocean Observing System (GOOS).

A modernization of the MCSS is therefore underway, and a Vision for a new Marine Climate Data System (MCDS) has been adopted by JCOMM through Recommendation 2 (JCOMM-4) to address those requirements.

The MCDS Vision is to formalize and coordinate the activities of existing systems, and address gaps to produce a dedicated WMO-IOC data system operational by 2020 in the view to have compiled coherent met-ocean climate datasets of known quality, extending beyond the GCOS Essential Climate Variables (ECVs). These will be of known quality collected from multiple sources to be served on a free and unrestricted basis to the end users through a global network of less than ten WMO-IOC Centres for Marine-Meteorological and Oceanographic Climate Data (CMOCs) covering specific JCOMM data domains. Data, metadata and information will be fully interoperable with the WMO Information System (WIS) and the IOC/IODE Ocean Data Portal (ODP), will be compatible with, and contribute to the High Quality Global Data Management System for Climate (HQ-GDMSC) that is being developed by the WMO Commission for Climatology (CCl).

This system is expected to improve timescales for met-ocean climate data availability, facilitate the exchange of historical met-ocean climate data sets between countries, and thereby increase the amount of ocean observations eventually made available to the relevant end user applications. Furthermore, integrated data and metadata will be available containing comprehensive dataset information e.g. historic details on current and past data codes and formats.

The data management structure will be standardized, well defined and documented for existing and new data across JCOMM activities and state of the art marine climate and statistical products will be easily accessible.

The development of the MCDS requires using state of the art integrated and standardized international systems for the improved data and metadata-flow and management of a wide range of met-ocean climate data. This includes integrating collection, rescue, quality control, formatting, archiving, exchange, and access of *in situ* and satellite sources. This system will be based on improved quality management, documenting processes and procedures, using higher level quality control, added value data processing, including bias correction, and comparison of the observations with satellite and meteorological and oceanographic model gridded fields.

It is expected that the relevant data and associated metadata will be of known quality, and extend to products that satisfy the met-ocean climate data requirements for climate monitoring, forecasting, and services.

Recommendation 4.3.4: JCOMM with partner organizations, including in particular the IODE will develop a strategy and implementation plan to realize the Vision for the MCDS, and address the recommendations from OceanOBS'09 (i.e. CWP from Woodruff *et al*[6]).

---

[6] https://abstracts.congrex.com/scripts/jmevent/abstracts/FCXNL-09A02a-1727870-1-cwp4c18.pdf

5. DATA PROCESSING

Members / Member States of WMO and IOC participating in JCOMM maintain and support their own national archives. The strategies and resources (i.e., people, hardware, software) needed are driven by national requirements and funding. It is not worthwhile for JCOMM to try to dictate the details of how this archiving takes place. However, JCOMM can provide valuable coordination in recommending data management practices that standardize how data are handled and thereby improve the preservation of the data and its usability. This section discusses the processing functions that impact the fidelity of archived data.

5.1 Data Versions

The raw observations coming from an instrument may be considered to be the first version of the data, and the highly processed data that are exchanged to be another version. There may be many versions representing the processing steps between these two and there may be other versions after data are available for exchange. Versions are generated by the calibrations applied to the data collected, by value adding processes such as quality control, by smoothing and filtering, etc. It is important to be able to distinguish between these versions especially after the versions have reached archives or are exchanged.

In dealing with data versions, the satellite community speaks of “levels” of data. The levels are indicative of the amount of processing that the data have undergone. So, level “0” is assigned to the data as delivered directly from the instrument sensor, while level “3” are gridded data processed from a single type of sensor (one satellite sensor or one in situ network).

The conventions used in The Global Ocean Observing System (GOOS) Prospectus 1998 (see annex 4 of http://ioc-unesco.org/index.php?option=com_oe&task=viewDocumentRecord&docID=171) are based on the definitions used by the atmospheric research community since the Global Atmospheric Research Programme (GARP). These are very similar to those of the satellite community with level “3” being gridded products and level “4” being model results.

To date there has been no history of the use of such schemes in oceanography although this is likely to change very quickly with the development of operational oceanography and global modeling. However, there are subtleties that are not expressed in these simple schemes that can have a strong impact on usage of the data. For example, for bandwidth or other reasons, the full resolution data coming from oceanographic sensors at sea are not immediately returned to shore and distributed, such as on the GTS. Instead, the TACs provide a way to distribute a degraded copy (in both vertical resolution and in precision of measurement). These low resolution forms enter and reside in archives until the higher resolution forms arrive in delayed-mode, thus leading to the potential archiving of both resolutions. And, though it should be relatively simple to recognize that these derive from the same observation, this is not always straightforward. The delayed-mode data may have corrections to positions or times, or calibrations carried out on the original measurements. All of these change the content and make the matching of the quickly arriving, low resolution data to the more slowly arriving, corrected, high resolution form a difficult process.

In recognition of this issue, a pilot project was initiated within the Ship Observation Team (SOT) in the OPA that uses a unique identifier attached to both the real-time and delayed-mode versions of the same original data. The matching is then done through examining these identifiers, not through looking at any of the data or information about the data. The scheme has shown value in this use and should be considered as a candidate technology in addressing the versioning issue.

Another subtlety is associated with the archiving of the data. It is common in both oceanography and marine meteorology archive centres to see the same data arrive more than once. The first time the data might arrive in real-time and in a lower resolution form as already described. The next time, the higher resolution or more complete delayed-mode form arrives (e.g., VOS reports containing data and metadata fields in IMM format that cannot be represented in FM-13). After some work by scientists, errors in the data may be fixed, calibrations carried out and so on and the data are again sent to the archive centre. It may be that some years later, the same data are submitted yet again because whoever was responsible was not sure they submitted the data so they send it again to be sure, or some data that were formerly missing have been recovered and the data are resent. This can happen for complete collections or sometimes just for selected components. These all represent versions of the data.

Recommendation 5.1: DMPA needs to consult JCOMM PAs to get a full description of the versioning issue, to develop a strategy to manage versions, and to implement the strategy.

5.2 Data Quality

Assessing the quality of data is a complicated process that uses knowledge of oceanography and meteorology, and knowledge of the area and time in which the data are collected. The degree of scrutiny is dependent on the use of the data. Data collected and distributed in real-time, such as in model assimilation, must be handled rapidly and usually this means being checked by automated procedures. It is accepted that such procedures cannot bring to bear all of the knowledge that an experienced person would, but as long as the number and type of errors that pass through the procedures do not adversely affect the results, this is acceptable. Assessing the quality of delayed-mode data often falls into the domain of scientists who have their own experience and tools they use.

Because of the accepted constraints of moving data in real-time, it is easier to get agreement on standardized procedures for carrying out quality control for this exchange. Users accept that not all errors will be caught. Getting the same acceptance for standardized delayed-mode procedures is more difficult. Often, the problems are more difficult to detect and require a broader set of corroborating evidence.

It is becoming increasingly common for ocean data to be reported or distributed with quality indicators attached. Processing centres add value to the original data by passing the data through test procedures and then using flags to indicate the quality of the data. These typically are placed on each observation.

There are a number of places where standardization of practices would benefit users. At present, a large impediment to a user wishing to take advantage of the flags is that the test procedures are not well documented nor are the descriptions readily available. Moreover, there are so many different procedures applied in different ways, even if a user accepted the tests as valid, combining data from many sources would mean determining that all groups used functionally similar tests. A standard set of tests to be applied as a minimum set would greatly improve the process of exploiting flags set by different groups. This standard set would need to respect the differences between real-time and delayed-mode data.

Recommendation 5.2a: DMPA should encourage the development and wide spread implementation of a standard suite of data quality testing procedures.

In the process of carrying out tests on observations, a decision must be made about how to report the results of the tests. There is wide agreement that flags will be used, but there is no agreement on what they will report. In one scheme, flags are used to indicate if an observation (or perhaps group of observations) passes or fails a test. The result is that a single observed value may have a collection of flags attached to it, one for each test performed. The advantage of this scheme is that there is no interpretation of the “goodness” of the data, simply a statement of success or failure of a test on an observation. Users are therefore able to decide for themselves which tests they consider appropriate and so take into consideration the test results when viewing the data.

The disadvantage of the pass / fail flagging is that it does not help a user who does not have enough knowledge to decide the importance of this or that test failure. Another scheme for indicating data quality has been developed for them. In this scheme, the same suite of tests may be run as for the first, but depending on the results, a value judgment is made about whether the data are acceptable. So, each observation receives a single flag that indicates if the measured value is considered good or not (or some degree of uncertainty). The advantage of this strategy is that, at least in oceanography, there is still much energy devoted to visual inspection of data and it is the technician operating the quality control process who makes the final decision. The disadvantage is that the results are not strictly reproducible since it relies on operator judgment.

There are arguments to support both schemes. In situations where there is an abundance of data, algorithms can be used to decide on the inclusion or exclusion of data to be used. Where data are scarce, each observation is precious and work is undertaken to use everything that can be used. It is time to reconcile these different approaches to ensure that users are best served.

Recommendation 5.2b: DMPA should resolve the differences in how the quality of data is indicated to best serve user needs.

The objective of data quality control is to ensure the data consistency within a single data set and within a collection of data sets, and to ensure that the quality and errors of the data are apparent to the user, who has sufficient information to assess its suitability for a task. Data quality flags provide the user of the data with clear information about actions taken to change the original data. The procedure for flagging data values to indicate their quality, reliability, or checks which have been carried out, or altering values after checking, filling in data gaps, etc., can vary from project to project, and between different laboratories and data centres.

The JCOMM/IODE Ocean Data Standards project has published *Recommendation for a Quality Flag Scheme for the Exchange of Oceanographic and Marine Meteorological Data*. (IOC Manuals and Guides, 54, Vol. 3.) which recommends a Quality Flag Standard to define a common set of quality flags to be used by data centres and projects. This two-level quality flag scheme will facilitate the exchange and integration of multi-disciplinary oceanographic and marine meteorological data. The first, or primary, level defines the data quality flags only, while the secondary level complements the first level by providing the justification for the quality flags, based on quality control tests or data processing history.

A related question that also needs consideration is how to manage the quality flags that come with data. One solution is for the receiver to keep these flags and add their own. After data change hands a few times, the resulting series of quality flags may well be confusing to a user. The other solution is for the receiver to use the attached flags as guidance, but if additional Quality Control (QC) is done, to overwrite the incoming flags with the results of this action. The user sees only the one set of flags, but there may be some valuable information lost.

Recommendation 5.2c: JCOMM to encourage the use of the IODE Quality Flag standard to indicate data quality and to provide a scheme for the exchange of data between data centres and programmes.

These recommendations are important. There are substantial resources expended in assessing the quality of data, and these are consumed over and over again by each group receiving data. Until there is consistency in what procedures are applied, acceptance on a broad scale of the correctness of the procedures, and a standardization of how results are reported, there will be no resource savings.

5.3 Duplicates

Data versions are created in the processing of data from original forms to forms in an archive and made available to users. Duplicates, and near duplicates may also be created as well. This can happen as the data are transferred to various archive centres and exchanged in projects around the world. Because of processing that takes place at each of these centres, the data and information that is archived may not be identical to the original. This can be the result of format conversions, trimming away of some of the information when data are sent from one place to another, errors in transcription, etc. Similar problems can occur in an individual archive where the same data arrive from two different sources. If these arrive at separate times, have different content, or are processed by different people, the fact that they are duplicates may go unnoticed.

If someone wants to assemble all of the available data, they may contact different archives asking for their holdings. In return, they will receive these multiple copies, some of which will be obvious as exact duplicates and some which will be inexact duplicates that are not so obviously derived from the same original. These duplications may have significant impacts on analysis and so are undesirable and need to be eliminated.

The detection of exact duplicates is relatively straightforward, and can be taken care of by algorithmic means. But finding inexact duplicates is not so simple. Inexact duplicates can arise, for example, if position precision is degraded, or if a value derived by extrapolation is inserted at the surface. The longer that data have existed, the more likely it is that that they have been through more transformations and exchanges and so the more likely that inexact copies exist in a number of places. Though there are examples of software operating today that are reasonably good at detecting such duplications, none of the schemes are fool proof.

A possible solution would be to employ unique identifiers such as Digital Object Identifiers (DOIs) as briefly described under versioning as a way to help find duplicates. The idea would be to attach a unique identifier to the original data as collected. As the data went through processing, were delivered to and passed among archives, the unique identifier would always accompany the data; that is it would never be removed by any process along the way and would never be altered. Then, as newly arrived data came to an archive or user, they need only check for a duplication of the identifier, and not have to devise elaborate rules to decide if the new data are exact or inexact copies of something already present.

Another solution would have national archives maintain data collected by their own nationals, but would provide these data to all others. That is, the original or reference copy would be maintained at a single location. All other copies would be recognized as copies and if there were any differences, the original version would be considered the true version. Weaknesses of this are that not all nations have the same infrastructure to fully support this model, and there would need to be some resolution of where data would go when the collection activity is multi-national.

Recommendation 5.3a: DMPA develop a methodology to address how to identify exact and inexact duplicates in contemporary JCOMM data.

The unique identifier approach can be effective both within a single archive and across archives that exchange data. It could be implemented in an incremental way, such that as an archive adopted the practice of employing unique identifiers, and it would accrue benefits by simplifying its duplicates detection process. In order for this to propagate into all of the archive systems in the world, there would need to be close cooperation of JCOMM with existing archives of IODE, WMO and ICSU.

Recommendation 5.3b: JCOMM consider developing a comprehensive system to uniquely tag data from all of its programmes and employ this to detect data duplications.

5.4 Contents

Each data system, whether dealing in real-time or delayed-mode data, has its own scheme for storing the data and the information about the data. The methods that are used are strongly influenced by the available computer infrastructure. The result of these varying approaches is that it is difficult to compare data and information from different sources. For example, it may be that one source provides a lot of detailed information about the instrumentation employed whereas another source provides no such information. The differences are not usually as extreme as this example, but these are differences that can be remedied to provide a more consistent and therefore, interoperable collection of data when assembling them from different sources.

There are some examples in operation now that have gone part way to standardization. For example, there are two vocabularies represented by BUFR tables and the CF conventions. As noted earlier, a mapping between these vocabularies would ease the inter-comparison of data reported in each.

However, this is really only a first step. As an example, it is common practice for an archive centre to keep information about the origins of the data they receive. It should be possible to develop a set of standard attributes to be recorded when known about these origins. Then, when data are requested, this information can be reported in a standard way.

5.5 Processing history

It is almost always true that when data arrive at any data centre they are transformed into some other internal data structure. But an important consideration in the archive and data preservation process is the accurate preservation of all of the originally reported data. Errors can frequently occur in translations of data between different formats. Keeping a copy of the data as originally received is a safe way to guard against transcription errors.

The transformation process can include actions such as converting data from one format or units to another (e.g., to SI units in BUFR, from originally reported units), applying quality test procedures, ingesting data into archives or models, corrections when possible, and so on. The processing stages can become very complicated with many decision points that cause changes in processing depending on the kind of data involved and its origins. Each of these steps that transforms the data or adds to it (as a quality assessment procedure adds quality flags) could be recorded as a processing history. This strategy can be very useful in finding and fixing problems generated in the course of routine processing. It can also be very useful in explaining anomalies detected in data.

There are a few programmes within JCOMM now where retaining a processing history has become standard practice (e.g., the Global Temperature and Salinity Profile Project). These programmes should be examined to determine the value of creating and preserving a processing history. JCOMM can then use this advice as a basis for recommending appropriate usage on a broad scale.

Recommendation 5.5: DMPA explore the value of preserving a processing history and recommend broad adoption if appropriate.

5.6 Metadata

Metadata (e.g. instrument/platform characteristics, tests performed and failed, origins of the data stream, data processing history, information about the data-sets) are critical for the users of the data. The OceanOBS'09 Community White Paper (CWP) on Metadata Management (Snowden *et al*, 2009, see website[7]) introduced the following levels of metadata:

---

[7] https://abstracts.congrex.com/scripts/jmevent/abstracts/FCXNL-09A02a-1745513-1-CWP4C13.pdf

(i) **collection/discovery level** describing data-sets,
(ii) **provenance (or lineage) level** describing the processing history, and
(iii) **instrument/platform level** describing the instruments and platforms used to make the observations.

The CWP explains that "*Without an active effort to manage the metadata describing ocean observations, much of the current research and operations expenses may be considered wasted in the worst case or suspect at best, because there will be no useful access to the data that results (where useful means: discoverable; recoverable; manipulable (understandable syntax); deconstructable (understandable provenance); and meaningful (understandable semantics). This will make the conclusions developed from that data scientifically marginal and less useful than they could be, because there will be no way to evaluate or test them by re-analyzing the data.*" Recommendations from the CWP includes (a) automation, (b) using unique identifiers, (c) proper planning at early stages of data collection, (d) open access, (e) timely access, (f) use of web services, (g) real-time distribution, (h) documenting quality control, (i) adoption of standards, and (j) distributing the system appropriately.

There are a number of initiatives related to metadata that are currently underway. One that has a substantial international subscription is the Marine Metadata Interoperability Project being run from the Monterey Bay Aquarium Research Institute (see http://marinemetadata.org/). Their work is mostly focused on ontologies (the science of describing the kinds of entities in the world and how they are related) and vocabularies.

There is also a project within JCOMM to define the kinds of metadata that should accompany measurements that are distributed in real-time or in delayed-mode. Generally, these deal with characteristics of instruments, data quality, etc. This has been pursued by the Expert Team on Data Management Practices (ETDMP) and is strongly linked to both the WIS developments and those of SeaDataNet (see http://www.seadatanet.org/).

There is the Dublin Core Metadata Initiative (see http://dublincore.org/). At the risk of over-simplifying this, the metadata considered here grew from the domain of library science and is strongly related to describing document origins and contents.

The International Organization for Standardization (ISO) is also well known for developing a broad collection of standards including for metadata (see http://www.iso.org/iso/en/ISOOnline.frontpage).

Another good example is IODE OBIS project that has adopted the Ecological Markup Language and associated vocabularies (e.g. WoRMS for species).

Each of the activities above tackles the issue of defining standards for recording metadata. Most have a particular purpose in mind, and this drives the content to be described. However, it is evident that metadata comprises a wide range of information. One attempt to address and categorize this range has been made by the U.S. Data Management and Communications Expert Team on Metadata (see http://dmac.ocean.us/index.jsp). This Expert Team divides the categories of metadata into "consumer use", "data management", "discovery", "access", "transport", and "archive". Specific metadata items exist in more than one of these groups.

At present, the term metadata is used in many ways, with the interpretation being provided by the context of the use. But this is confusing, and it would be far better to take the approach of developing categories of metadata, and to define the content to suit the purpose. One would then speak of discovery metadata, or transport metadata and both the purpose and content would be clear.

Recommendation 5.6a: DMPA examine existing metadata initiatives to develop a categorization that aligns with the purpose of the metadata.

Recommendation 5.6b: DMPA use the metadata categorization to develop a plan in which metadata initiatives align with its work and become engaged in these activities.

While the above activity is going on, there is a fairly well described class of metadata that is used for discovery. The information in this class is sufficient for a potential user of data to identify the data collections that exist in their area of interest, in a time frame of interest, in a scientific domain of interest, and perhaps even with variables of interest. This class of information appears in the FGDC (U.S. Federal Government Data Committee, see http://www.fgdc.gov/) standard used within the U.S., in the ISO-19115 standard (see http://www.isotc211.org/) and in the GCMD (Global Change Master Directory, see http://gcmd.nasa.gov/) and there are others. The objectives of these are to build standard records that can be stored in an electronic catalogue and that can be searched to find data of interest. This work is far enough advanced that JCOMM can usefully participate and in so doing will fulfill one of the recommendations covered in the section on Access that follows.

Recommendation 5.6c: Deprecated.

There exist a number of sources of information about the characteristics of platforms and instruments which are used to acquire oceanographic and meteorological data. For marine meteorology, there is WMO Publication 47 (Pub 47): International List of Selected, Supplementary and Auxiliary Ships (see http://www.wmo.ch/pages/prog/www/ois/pub47/pub47-home.htm). More recently, the Chinese Oceanographic Data Centre has established an electronic, on-line data base of instrumentation information about ocean data buoys, platforms, and other automated Ocean Data Acquisition Systems (ODAS - see http://www.odas.org.cn/ for information about this system). Both of these sources reflect the need to have information about the ways that observations at sea are collected. Such information is crucial in helping to explain topics such as systematic changes in observations from one platform to another or in compensating for changes in observation methods when looking at long time series. These sources are but two examples of the kind of additional information that is needed to interpret observations.

Gathering information of this kind and keeping the information up to date is not simple. It must rely on the individuals in countries whose job it is to service the platforms or instruments to ensure that changes are recorded quickly. However, there is also the role of an international body to be sure that the information is readily available, preserves past information as well as reflecting the most recent information.

The META-T project was completed in 2010 and allowed to learn the following lessons:

(i.) The concept of a single point of access to all JCOMM metadata that operates independently from the platform data management process is not sustainable, especially in a low funding environment.

(ii.) Organizing the metadata development effort around a geophysical variable, temperature in this case, was less effective than organizing around platforms.

(iii.) JCOMM is organized into panels that have common platforms and data processing systems. Exploiting this organizational infrastructure will likely be more effective, at least initially, than creating a separate metadata service that aims to integrate across panel activities and is developed independently. On the other hand, such cross-panel metadata services (e.g. the ODAS Metadata Service, ODASMS) may also still end up having important downstream integrating and permanent archive roles.

(iv.) Data content and data representation standards are enabling technologies that can serve to integrate the platform-focused activities of the panels.

(v.) Determination of the platform metadata that should be collected requires input from a broad range of ocean observation data stakeholders. This group should span domains from short timescale weather forecasting to climate science. As with many other elements of the observing system, metadata management should be seen as cross-disciplinary and not serving only one domain area. This requires, for example, recognition on the part of operational weather forecasters of the climate requirements and a willingness from the climate community to work within the confines of existing operational weather forecasting community infrastructure/resource constraints.

(vi.) A key strategy recommended by Meta-T is to include as much metadata as is practically available at the time of GTS encoding in the BUFR templates. Therefore, of primary importance to the overall management and distribution of data and metadata, is the design of BUFR templates.

(vii.) A new JCOMM DMPA Task Team focused on exploring web based technologies could leverage the DMPA Task Team on Table Driven Codes (TT-TDC) efforts and extend them with web-services that provide deeper functionality than the GTS can currently deliver.

Recommendation 5.6d: JCOMM to encourage all agencies keeping information about instruments, platforms, etc., to place this information on-line and keep it up-to-date. JCOMM to develop a strategy for managing the international suite of instrument metadata sources so that they are easily found and used in line with the OceanOBS'09 recommendations (i.e. Snowden *et al* CWP[7]).

Recommendation 5.6e: Deprecated (merged with 5.6d above)

5.7 Model Data

Computer modeling is an activity carried out in many Member / Member States actively involved in JCOMM activities, with uses ranging from research to product development. The results are closely linked to how the observational data are assimilated into the model and how the computations are carried out by the software. Numerical models can produce large volumes of data since they can provide a continuous, quantitative representation of atmosphere or ocean variability in the four dimensions of space and time.

The results of models are valuable to others because they take limited observational data and perform a kind of interpolation / extrapolation to provide results where observations are poorly sampled in space and / or time. Models can be used to hindcast or reconstruct past variability; nowcast or provide the state of the system by combining observations, dynamics and other empirical information; and forecast conditions in the future. The resulting value-added fields and products are used by others directly or as inputs to other kinds of models.

Research models are run to explore scientific issues, are constantly being improved or changed, and have results that are generally of immediate use to only a small audience. Operational models are run on a routine schedule, have characteristics that are fixed for considerable periods of time and hence can be readily documented, have undergone some degree of observational validation, and provide products that are of wider use and distributed to clients on a routine basis. Re-analysis fall into a third model category, in which tightly constrained versions of operational models are run over long retrospective periods to produce the most homogeneous results for climate research. The shared characteristics of operational and reanalysis models define the key attributes that determine if model results have value to archive. JCOMM should consider the results of such models as data assets and consequently they should be managed appropriately.

Recommendation 5.7a: JCOMM to work with the modeling community to define the characteristics that determine which outputs should be archived.

The volume of data produced by a model may be an issue. In addition, it will be important to devise an appropriate indexing scheme so that subsets of the outputs can be quickly identified and accessed.

Recommendation 5.7b: JCOMM to work with relevant modeling groups to develop cost-effective strategies for the storage and archival of operational model outputs and products.

The model characteristics are of great importance as they impact what data and information to archive, and for how long they should be archived. In addition to this, the data assimilation schemes, observational inputs, computational algorithms and generally the important internal operations of the model need to be documented so that comparisons may be made between models and observations, and reliability can be assessed.

Recommendation 5.7c: Appropriate model characteristics will be archived with model results.

Models change and improve so that older versions of models are retired and newer versions come into operation. Each time there is a change to an operational model, the value of retaining output from the earlier version should be assessed.

Recommendation 5.7d: JCOMM will collaborate with model developers to decide the long-term value of preserving outputs of retired versions of models.

5.8 Specialized data centres)

The Specialized Oceanographic Data Centres (SOCs) are a legacy from IGOSS. They were formed by member agencies volunteering to carry out an activity to meet a particular need within IGOSS. The SOCs were of different kinds with different foci, such as managing data or on capacity-building activities. The list of SOCs included ones for real-time ocean profile data, sea level, and surface drifters. With the development of the MCDS, SOCs have been deprecated, and replaced by Data Acquisition Centres (DACs) or Global Data Assembly Centres (GDACs).

The Responsible National Oceanographic Data Centres (RNODCs) were a creation of the IODE. These included RNODCs for the Southern Ocean, for surface drifters, for MARPOLMON (Marine Pollution Monitoring Programme), for IGOSS, for the Western Pacific (WESTPAC), the JASIN Project (Joint Air-Sea Interaction Experiment), ADCP (Acoustic Doppler Current Profilers) and for the Indian Ocean (INDO). In the last review of IODE, it was determined that the RNODC system was working only in a few cases and were abolished by IODE in 2005 (Resolution IODE-VIII. 2). In 2013 the IODE Global Data Assembly Centres (IODE GDACs) were established as structural elements of IODE (Recommendation IODE-XXII.13). IODE GDACs will receive and assemble marine meteorological and/or oceanographic data, in real or delayed-mode, from the appropriate data streams and check the are consistent through quality control procedures defined by the international standards and methods established by IODE, WMO or JCOMM as appropriate.

As part of the development and implementation of the JCOMM Marine Climate Data System (MCDS), and its data-flow, Data Acquisition Centres (DACs) and Global Data Assembly Centres (GDACs) are also being established. The MCDS development by JCOMM is being undertaken in close cooperation with the IODE in order to build on potential synergies and avoid overlap and duplications.

Recommendation 5.8: JCOMM and IODE seek to harmonize the work of IODE National Oceanographic Data Centres (NODC)s and GDACs with the Marine Climate Data System (MCDS).

6. ACCESS

Finding data of interest in the world of distributed archives and data sources is not easy and it can be very difficult for the occasional user. However, the ability to make observations in the marine environment is changing rapidly, and it is becoming much easier for new data sources to appear that are outside of the traditional data collection communities. Therefore, it is important to have some way to find all of these data sources. Once data are found, they must be available to users. If there are products generated from the data, these products must also be made available. This section discusses how to provide access to the data and information held in JCOMM.

6.1 Discovery

Currently, there is a high level of importance assigned in the international data management community to the construction of catalogues that describe the data held in the large archive centres. This development is advancing through the acceptance of standards for describing data holdings as embodied in such catalogues as the GCMD (Global Change Master Directory), through the use of the FGDC (US Federal Government Data Committee) metadata standard and the ISO model. There is much work currently being done to develop domain specific "profiles", such as ones for meteorological or for oceanographic data, within ISO standards. Constructing catalogues with the same (or fields that have a 1-1 mapping) contents and linking catalogues by using standards such as ISO-23950 allows queries to cross from one catalogue to another. This achieves interoperability, without the requirement of centralizing the catalogue. These also help to address the "data discovery" problem.

Another strategy is to exploit existing commercial search engines, such as Google, as the way to locate data. To do this requires placing appropriate information in the parts of static web pages that describe the data so that the search engine web crawlers can locate the information and index it. To be effective, there needs to be an agreed standard for how data will be described, perhaps similar to what appears in ISO profiles.

Both of these strategies have merit and perhaps cater to different user communities. These ideas need to be further explored and tested.

Recommendation 6.1a: JCOMM to promote the use of existing ISO standards for data discovery metadata to support interoperable catalogue services and registries.

Recommendation 6.1b: Deprecated

6.2 Browse

After potentially useful data collections have been identified, some further exploration usually is required to determine if the archive has the specific data of interest. This is often necessary since the data discovery information may not be detailed enough to answer all questions that a user might have. Generally, the tools needed to support this browse capability are closely tied to the archive in which the data reside. This can be remedied to a degree by adopting certain technologies. For example, OPeNDAP Open-source Project for a Network Data Access Protocol (OPeNDAP) (see http://www.opendap.org/) also provides browse capabilities for data collections available in netCDF or a few other formats and using certain standard structures. This technology provides a looser connection to archive formats.

Web service technology, as embodied in the standards promoted by the Open Geospatial Consortium (OGC) (see http://www.opengeospatial.org/) such as for web map services, may be of use. This strategy would identify a set of data browse services that every archive centre should comply with and which would deliver a standardized response. This provides a very loose connection to archive structures since each archive would write the connection software needed to provide the standardized service.

There are other technologies being explored including those embodied in the WMO Information System (WIS), which will be described later in this document.

Recommendation 6.2: JCOMM explore the implementation issues of existing or proposed methods for supporting browse functions.

6.3 Data Delivery

Delivering data to users is a fundamental objective. There are several ways to deliver required data to end-users – on request and on regular basis (subscription). While implementing delivery mechanisms one should consider potential data volume to be transferred, which means that appropriate compression technigues must be applied. User should be informed (by email or on web page) that following delivery has been completed (or failed). Data delivery services in general should support SMTP (email) and FTP delivery. Basic data download by end-user is not considered as quaranteed delivery process. IODE ODP has implemented following delivery services and supports delivery of the data provided by ODP stakeholders.

A further consideration is the national policies on providing basic data and higher-level products to users. Both the WMO and IOC have policies regarding access to data. It is important that each Member / Member State follow these policies to provide as free and open access to the data as they can. Each Member / Member State needs to analyze their capabilities and national policies and determine for themselves what can be supported.

Recommendation 6.3a: Each Member / Member State of WMO and IOC participating in JCOMM needs to examine its ability and willingness to provide all of its data holdings on-line. Each will determine what level of support it can bring to bear.

The WMO Information System (WIS) has a role to play in providing access to data. The role of JCOMM in WIS will be discussed later in this document.

Not all data users will be satisfied by the functionality provided by WIS. This means that JCOMM will also need to participate in other distribution schemes for providing data and information to clients. In general terms, it is important to be sure that what is built is part of a larger project and where possible, exploits standards such as those promoted by OGC or ISO bodies. There must be agreements on a small but common set of data exchange formats that all sites will offer.

Recommendation 6.3b: Deprecated (moved to section 7)

Because much of the real-time data are collected through JCOMM programs, and because JCOMM through cooperation with IODE and WMO has access to the historical record, an argument can be made for developing at least two sorts of products from the archives. The first is to build climatologies that can be used by all Members / Member States. Such climatologies are very valuable in testing whether newly arrived data appear to have unusual values and so may, in fact, be in error. They are also valuable in assessing if present conditions are warmer or cooler than average. At the moment, there are a number of climatologies in existence and depending on which are used, and results of comparisons and conclusions about differences can vary.

Climatologies should be built with the active collaboration of appropriate members of the scientific community. Such collaboration ensures that there are sound principles behind the choices made in consolidating and averaging data and hence increase the acceptance of the results.

A second product is to construct specialized archives. A good example would be to build an archive of all of the instrumented wave elevation and wind data where the waves are extreme. Such an archive would be invaluable to wave modelers since there are few such data to be had, and it is in the extreme events that the differences of the models show most clearly. There are other examples of where such extreme event data would be useful as well, including episodic events such as El Nino, harmful algal blooms (HABS), sudden deepening of storms, etc.

Just as for climatologies, the building of such archives needs to be done in close cooperation with the appropriate scientific group. In the case of extreme waves, the group exists in the SFSPA. In some cases, it may be sufficient to build appropriate data mining tools that can find and extract the data required from global archives. In other cases, it may be necessary to search out such data from the various agencies around the world that hold them and do all of the consolidation and standardization to bring the data into a single archive.

Recommendation 6.3c: DMPA encourage the development of a new Marine Climate Data System (MCDS), including the compilation and adoption of a standard climatology, of the creation of specialized archives, and other products that have wide spread applicability to Members and Member States.

Not to be forgotten is the importance of the information that describes the instruments used to make the observations, the ways the observations were collected, whatever processing they may have passed through and so on. These metadata are extremely valuable in helping to interpret the observations and are especially important when looking at long time series where instrumentation may have changed. These metadata should accompany the data so that the user has the full information required to make maximum use of the data.

Recommendation 6.3d: JCOMM needs to ensure that all information required for the correct interpretation of data be included when data are delivered to clients.

6.4 Data Access Policies and Security

The first issue has to do with the data and information access policies of Members / Member States of WMO and IOC participating in JCOMM. Both relevant WMO (see http://ftp.wmo.int/pages/about/Resolution40.html) and the IOC (see www.iode.org/policy) have data policies that have been constructed with careful consideration of the views of Members / Member States. Different countries will have more explicit policies that will apply to exchanging data, or delivering data to clients as envisaged in this DMPlan. Clearly, JCOMM data management activities need to operate within these intergovernmental and national policies.

A draft WMO Policy for the exchange of data and products to support the implementation of the Global Framework for Climate Services (GFCS) is under elaboration by the WMO with the goal to have it adopted by the WMO seventeenth Congress in 2017. JCOMM was invited to take part in its development, and the policy, if and when adopted will also have to be followed by JCOMM.

The second issue concerns guarding the integrity of data holdings from malicious individuals who break into computer systems and do harm. This is a serious issue even when the data are freely available. Each country providing access to its data or information holdings needs to protect these assets in conformance with national practices.

7. COORDINATION AND LINKAGES

As a Technical Commission co-sponsored by WMO and IOC, JCOMM has many connections among its internal programmes as well as to organizations and groups outside. The JCOMM DMPlan must recognize these partners and the distributed nature of the international data system.

Data collection is carried out as a national activity, whether or not it is part of an international programme. There is a diversity of observational, processing, and distribution capacity among nations. This DMPlan needs to function within this context, and encourage, and support increasing these capacities for Members / Member States of WMO and IOC participating in JCOMM.

To meet JCOMM objectives will take an increased level of cooperation and coordination by all Members / Members States and other international organizations. Strong partnerships need to be forged with these other programmes. This section explores the implications of these connections and discusses some of the needs to support or enhance the cooperation.

7.1 Within JCOMM Activities

The OPA encompasses many of the observation programs at sea. It includes programs using surface drifters with the Data Buoy Co-operation Panel (DBCP), Volunteer Observing Ships (VOS) with the VOS Scheme, Ships Of Opportunity (SOO) with the SOO Programme (SOOP) Implementation Panel (SOOPIP), the international tide gauge network (GLOSS, the Global Sea Level Observing System), and others. Many of these systems have been in place for a number of years and have built procedures for managing their own data and information streams. Some, such as VOS, are heavily reliant on facilities at WMO for managing information about the fleet. Others, such as the DBCP, recognizing the limitations of their initial systems, are now helping to build new ones to hold information about their platforms (e.g., ODAS metadata).

The JCOMM Services and Forecasting Systems Programme Area (SFSPA) also has observing programs incorporated into their work. For examples, the Expert Team on Waves and Coastal Hazards Forecasting Systems (ET-WCH)and the Expert Team on Sea Ice (ETSI) both deal in observations made by others. Their focus has been on coordinating activities so that data collected by one organization are easily available to another. The SFSPA also includes groups with a strong focus on products to support such activities as safe operations at sea or responding to accidents.

The DMPA has activities that connect directly to some of the observation programs within the other PAs. But this is not true for all OPA and SFSPA activities. The interaction between the data managers of the different groups has been through informal discussions only with the result that there is only a small degree of commonality.

Recommendation 7.1a: JCOMM to ensure regular exchanges of information and ideas on how data are managed between the groups in OPA, SFSPA and DMPA.

It was recognized that in creating JCOMM there needed to be links to the satellite community and the data that are so acquired. JCOMM has satellite rapporteurs in each of the programme areas. Their responsibilities are to understand the activities in the PA, to understand activities taking place in the satellite community, to help make linkages where appropriate and to bring to the attention of one or the other community actual and proposed activities that impact operations.

Data management in the DMPA is focused on *in situ* observations. There is no intention to duplicate the data management activities that are employed in the satellite community. However, it is important to build bridges to that community so that data handled by JCOMM and data acquired by satellite operators can easily be combined and compared. In particular, JCOMM Programme Area is invited to collaborate with the JCOMM cross cutting Task Team on Satellite

Data Requirements (TT-SAT). The TT-SAT is particularly tasked to develop a strategy and process document with recommendations on how to (i) maintain JCOMM short term forecasting applications requirements, with emphasis on the integrated use of *in situ* and remotely sensed data for data assimilation systems in forecasting services with focus on accuracy and timeliness requirements; (ii) develop integrated satellite products for multiple marine meteorological and oceanographic users; and (iii) promote consistent quality control between *in situ* and remotely-sensed data, together with appropriate feedback mechanisms.

Recommendation 7.1b: JCOMM must consider interoperability issues with satellite data providers so that satellite and in-situ data are easily compared.

Data management activities across JCOMM PAs will be furthered by the introduction of standard practices in many facets of their work. This ranges from simple things such as the adoption of common naming conventions for variables, consistent units of measurement, selection of common formats for delivery of data to clients, and mandatory metadata content to describe data holdings throughout JCOMM PAs. There is much work to be done in the domain of standards. JCOMM should take the approach of adopting an existing practice as the standard as the first choice when this is available. If no existing practice meets the minimum needs for JCOMM, the second consideration should be given to make appropriate adaptations to an existing practice. Finally, and as a last resort, JCOMM may need to devise its own standards, though this should not be done without careful consideration.

Recommendation 7.1c: JCOMM should collaborate with the IODE with regard to the Ocean Data Standards and Best Practices Project.

The IODE, together with JCOMM, has developed a project, Ocean Data Standards and Best Practices (ODSBP) Project, to achieve broad agreement and commitment to adopt key standards related to ocean data management and exchange to facilitate exchange between data centres and contributing programmes. This project has established an internationally recognized process for submitting proposed standards and for their acceptance by the ocean and marine meteorology community.

Recommendation 7.1d: JCOMM to identify and propose relevant standards to the Ocean Data Standards and Best Practices (ODSBP) Project. Recommendation 7.1e: Deprecated.

The IODE has established the IODE Clearing House Service for Data and Information Management Practices which will be a repository for ocean data standards and best practices documentation. The JCOMM Catalogue of Standards and Best Practices, established as a contribution to the WIGOS Pilot Project for JCOMM, will be integrated into IODE Clearing House and used to disseminate relevant standards and practices.

Recommendation 7.1f: JCOMM to contribute documents describing best practices and related standards to the Clearing House Service.

Members / Member States will have varying abilities to respond to adopting recommended standards. JCOMM will need to ensure an appropriate implementation procedure is in place. The speed of implementation of standards may be enhanced by an appropriate use of capacity-building activities.

Recommendation 7.1g: Consideration must be given to how to implement the standard across WMO-IOC Members / Member States as rapidly as possible. Documented use-cases should be developed and consideration should be given to how capacity building resources may be used.

Coordination must also take place with the other programmes in the IOC, WMO, as well as regional and national activities. Some of this will be ensured by members of the DMPA and other PAs being participants of the various activities. A challenge to the DMPA will be to keep abreast of these activities, and to select those in which to participate actively and those that bear watching only. Because of the wide variety of programmes, the DMPA needs to adopt a reporting process whereby Members / Member States hearing of significant activities to JCOMM can report these. Equally, DMPA needs to look ahead to select priority activities and use this as a basis for gauging where member resources are to be invested.

Recommendation 7.1h DMPA establish a reporting process that has members informing the group of significant activities in other programmes.

Recommendation 7.1i: DMPA set priority activities each intersessional period and use this as the guidance to selecting activities for its Members / Member States.

Within the OPA, work is being done to develop and provide regular reports with a concise view of where the observing system stands in meeting Global Climate Ocean Observing System (GCOS) objectives. The target audience for this report is senior members of governments who have the ability to influence budgets. OOPC is also working with OCG on variable-based system metrics, and will be working with JCOMMOPS and the Observing System Monitoring Centre (OSMC) to develop and regularly report on these metrics.

Recommendation 7.1j: DMPA in collaboration with OPAencourage the completion of quarterly reporting of system-wide metrics on both a network and variable basis, following the model used by OPA. Lead responsibility: DMCG and OCG chairs

The data systems in PAs have more detailed measures of how they are meeting their requirements. Some of these are formalized in an annual reporting mechanism, while others are less formal.

The objective would be to find elements across all data systems that gauge the success of the programmes to meet the overall objectives of JCOMM. It would provide a means for data managers of the various systems to see how they compare to others, to identify weaknesses and to show quantitative improvements as corrective actions are taken. Developing this list is an activity that could be coordinated by DMPA but requires the support of data systems operated by OPA and SFSPA.

Recommendation 7.1k: DMPA collaborate with appropriate members of OPA and SFSPA to develop a set of data system performance metrics and implement a standard reporting of these results.

Recommendation 7.1l: DMPA must keep aware of other and continuing projects to improve the access to data and where possible both participate in the projects and adopt procedures that improve access to JCOMM data.

7.2 With IODE Activities

The IODE began many years before real-time transmission of oceanographic data was practical. Its focus, therefore, is on acquiring the data collected after the cruise or data collection activity takes place, carrying out some degree of quality assessment and building national archives to ensure the data are preserved. Data managed within the IODE system generally are of scientific quality and therefore suitable for investigations into climate studies. Many of the data come from scientific researchers who contribute the data to their national data centres. These national centres, of which there are about 60, come together under IODE to exchange information and to build the current data exchange system. The IODE centres handle a variety of data, including a wide range of physical, geological, chemical, biological, and even some meteorological observations. In scope, the types of data managed by IODE are broader than currently managed by JCOMM.

However, there is overlap in both the kinds of data managed by JCOMM and IODE and the time-scales on which those data are handled. Depending on national organization, there can be a high degree of cooperation between IODE and JCOMM; this cooperation is vital. The full suite of oceanographic and meteorological measurements is large and diverse, and the work needed to manage the data is substantial. Where there is a high degree of overlap of interests in types of data, it is important to consider streamlined operations. In this regard, the IODE and JCOMM share the Expert Team on Data Management Practices (ETDMP) and coordinate its activities. Likewise, certain data management programmes of IODE, such as the Global Temperature and Salinity Profile Project (GTSPP), are jointly supported by JCOMM and IODE. There are other examples, and it is important to identify and recognize these joint programmes.

Recommendation 7.2b: Data management programmes of joint interest to both JCOMM and IODE be formally recognized and supported by both organizations.

IODE Members maintain a number of archives that are of direct interest to JCOMM and vice-versa. It is important for JCOMM to gain easy access to data maintained by IODE (of course, the reverse is also true for IODE having access to JCOMM archives). The comments made earlier about processing and access all apply here and should be taken into consideration. For example, confusion may arise where real-time data are handled by one organization, but the delayed-mode by another. In such a case, there may be differences in labeling the origins of the data, in the resolution, in processing, etc. This means a high degree of cooperation will be needed to ensure data can cross organizational boundaries without confusion of content. There is little doubt that this will mean the adoption of interoperability standards (a ready mapping of one standard to another) or the adoption of the same standards.

Recommendation 7.2c: IODE and JCOMM cooperate to ensure easy access and clearly described content of respective and interoperable data streams and archives. Lead responsibility: OCG and DMCG chairs.

7.3 With Other IOC Programmes

There are a number of other programs and projects within IOC including GOOS (the ocean component in GCOS), GODAE Oceanview, the Ocean Observations Panel for Climate (OOPC), the Group for High Resolution Sea Surface Temperature (GHRSST), Argo, etc. JCOMM is involved in many of these.

JCOMM figures prominently in the 2010 update of the GCOS Implementation Plan (GCOS No. 138), being mentioned in 23 actions of the 138 actions of the Plan. These span the range of JCOMM activities in all Program Areas. Those with direct mention of data management functions include:

- ***Action O3***: Improve number and quality of climate-relevant marine surface observations from the VOS. Improve metadata acquisition and management for as many VOS as possible through VOSClim, together with improved measurement systems.
- ***Action O7***: Continue the provision of best possible SST fields based on a continuous coverage-mix of polar orbiting IR and geostationary IR measurements, combined with passive microwave coverage, and appropriate linkage with the comprehensive *in situ* networks noted in O8.
- ***Action O17***: Establish an international group to assemble surface drifting buoy motion data, ship drift current estimates, current estimates based on wind stress and surface topography fields; prepare an integrated analysis of the surface current field.
- ***Action O20***: Document the status of global sea-ice analysis and reanalysis product uncertainty (via a quantitative summary comparison of sea-ice products) and to prepare a plan to improve the products.

- ***Action O28***: Develop projects designed to assemble the *in situ* and satellite data into a composite reference reanalysis dataset, and to sustain projects to assimilate the data into models in ocean reanalysis projects.
- ***Action O31***: Monitoring the implementation of the IOC Data Policy.
- ***Action O32***: Develop and implement comprehensive ocean data management procedures, building on the experience of the JCOMM Pilot Project for WIGOS.
- ***Action O33***: Undertake a project to develop an international standard for ocean metadata.
- ***Action O34***: Undertake a project to apply the innovations emerging from the WMO Information System, and innovations such as OPeNDAP to develop an ocean data transport system for data exchange between centres and for open use by the ocean community generally.
- ***Action O35***: Plan and implement a system of regional, specialized and global data and analysis centres for each ocean ECV.
- ***Action O36***: Support data rescue projects.
- ***Action O38***: Develop plans for, and coordinate work on, data assembly and analyses.
- ***Action O39***: Develop plans and pilot projects for the production of global products based on data assimilation into models. All possible ECVs.
- ***Action O40***: Undertake pilot projects of reanalysis of ocean data.

Implementation of the recommendations of this plan will address all of these actions.

Recommendation 7.3a: JCOMM and DMPA move quickly to adopt a data management plan and to further develop an implementation plan based on the DMPlan as rapidly as possible.

The GODAE project (see http://www.godae.org/) has developed a number of products, intercomparisons, and common output strategies for ocean forecast information. The work that has been done is important and directly relevant to implementing recommendations that are part of this strategy. Similarly, the GHRSST (see wwww.ghrsst.org) project has much to offer in demonstrating, among other products, how in situ and satellite observations can be used together. Argo (see argo.jcommops.orgshows how an international system can collect, manage and distribute data to support operational oceanography requirements. All of this experience is important and needs to be captured to build an effective data management system for JCOMM activities.

Recommendation 7.3b: JCOMM must work closely with the many other IOC programs in developing its implementation plans.

Capacity-building is an important activity. Within data management activities, it needs to cover all aspects from assembly of collected data, to processing and quality control, to archiving and providing access to the data. Information about the data collection, processing, etc., as referred to in other parts of this document is equally important. Suitable activities to increase the capabilities of WMO-IOC Members / Member States contributing to JCOMM to fully participate in data management and to use the managed data must be supported. However, JCOMM is not alone in wanting to address these issues. Both of the co-sponsors of JCOMM, the WMO and IOC, have capacity-building programs. Rather than construct something new, it is more appropriate for JCOMM to collaborate with these efforts. This can be done, for example, by ensuring suitable training materials on marine operations are present and by contributing instructors as appropriate. Joint activities between Member / Member States are another way to increase members' capabilities.

Recommendation 7.3c: JCOMM should collaborate with existing IOC (and WMO) capacity-building activities to ensure that the marine component is included.

7.4 With WMO

The WMO Information System (WIS) has an important role to play in providing data. The linkages of the meteorological side of JCOMM are closer to the developments of WIS than the oceanographic side. The DMPA is invited to provide input to the WMO Commission for Basic Systems (CBS) Expert Teams involved in the implementation of the WIS through the DMCG Chair, and the Chairs of the relevant Expert Teams[8]. The terminology used to describe the components of WIS is different than what is used by oceanographers, but the functions described are readily understood. In fact, DMPA had already taken some steps that compliment the work of WIS through its support of the IODE ODP in terms of metadata exchange, metadata catalogues, data and product delivery. At the meeting, connections were made between the work of IODE ODP and WIS and this work continues.

Recommendation 7.4a: DMPA and WIS should cooperate to ensure that all components of JCOMM data systems are available to WIS.

There are a number of other programs within WMO for which there are strong impacts on data management activities of JCOMM. These include the various committees, such the CBS Inter-Programme Expert Team on Data Representation, Maintenance and Monitoring (IPET-DRMM), which regulates how data are presented on the GTS, to the WMO Secretariat who maintains publications such as Pub 47. JCOMM must be engaged in these groups to represent its activities and to influence, as appropriate, activities in WMO.

Recommendation 7.4b: Deprecated

7.5 With International Council for Science (ICSU) World Data System (WDS)

There is a hierarchy of archives that exist in the world into which JCOMM data management activities fit. On the broadest international level is the World Data System (WDS) which is an interdisciplinary body of the International Council for Science (ICSU) WDS builds on the 50 year legacy of the World Data Centres and Federation of Astronomical and Geophysical data analysis Services, see http://www.icsu-wds.org/. , see.

The ICSU WDS promotes universal and equitable access to, and long-term stewardship of, quality-assured scientific data and data services, products, and information covering a broad range of disciplines. WDS aims at facilitating the scientific research endeavours under the ICSU umbrella by coordinating trusted scientific data services for the provision, use and preservation of relevant datasets. ICSU WDS is building worldwide "communities of excellence" for scientific data services by certifying member organizations that holders and providers of data or data products—from wide-ranging fields using internationally recognized standards. WDS Members are then the building blocks of a searchable common infrastructure with which to form a data system that is both interoperable and distributed. Membership of WDS increases exposure to international users and collaborators, and demonstrates a commitment to open data sharing, data and service quality, and preservation. WDS provides four different categories for membership: regular members, network members, partner members and associate members. Network members are organizations representing groups of data centres and/or data services that serve as coordinating agents for nodes that have common characteristics disciplines. In 2013 IODE was certified as a Network Member of WDS and through this membership all IODE NODCs will be linked to the WDS. The former World Data Centres for Oceanography at Silver Springs (USA), Obninsk (Russian Federation) and Tianjin (China) are regular members of the WDS.

Just as JCOMM must have close ties to IODE for oceanographic data, it must also have similar ties to WDS members managing data of interest. Indeed, the issues of standards, archives and access, all apply to consideration of interactions with WDS members as well. In this context, JCOMM should take the opportunity to build stronger relationships.

---

[8] See the list of Expert Team at http://www.wmo.int/pages/prog/www/CBS/CBS-WorkProgramme/OPAG_ISS_ETs.html

Recommendation 7.5a: DMPA, in collaboration with IODE, build stronger links between the observing and archive systems to ensure all data is available through the WDS.

WDS members are an international focal point for all of the data. Besides ensuring the stewardship and dissemination of the data, they are also in a position to create or collaborate on the production of climatologies. These products, as mentioned earlier, can be an important output from JCOMM.

Members / Member States can both contribute by timely provision of data to the WDS and benefit from recent and appropriate and timely updates to the global data set and climatologies.

Recommendation 7.5b: WMO-IOC Members / Member States participating in JCOMM support the timely assembly of data in the WDS and encourage timely updates and distribution of the global data sets and climatologies.

7.6 With Other Programmes

There are a host of other programmes and projects carried out in national and international fora that lie outside of the organizations discussed already. Many of the data management problems they address are the same ones experienced by JCOMM. Data collected under the JCOMM umbrella contribute to these programmes and WMO-IOC Members / Member States participating in JCOMM also are members of these programmes. There is a great deal of inter-programme communication through the individuals that contribute to JCOMM and these other programmes. Data management plans and implementation cannot and should not ignore these activities. This experience is valuable and the solutions are worthy to note.

Recommendation 7.6: JCOMM must develop a level of interoperability in data management with other major international and significant national programmes.

8. COMMUNICATIONS

Implementation of all of the recommendations posed will greatly enhance the capabilities of JCOMM to meet its objectives. However, it is clear that JCOMM is not an organization operating in isolation from other activities and programmes. Implementing the recommendations is only completing part of the work. It is important that after collaboration with JCOMM partners on many of their issues, JCOMM communicates these results globally. It is important to tell others what is being done, explain why, and show results. It is through this process that others will understand what JCOMM is doing and express interest in joining or pointing out similar endeavours.

One way to provide information is to use the Internet and WWW technology; JCOMM already has a website. There are also multiple sites (e.g., www.jcommops.org) associated with specific programs or Expert Teams. Within these sites, it will be necessary to have additional pages that provide information about standards adopted by JCOMM, about how to connect to data and information and the work program and results from the DMPA. None of this additional information is currently available (or written) but it is important and must be undertaken.

Recommendation 8a: DMPA undertake to design and populate web pages that explain its activities.

It is important for representatives of JCOMM to attend meetings of other organizations where interests intersect. At these meetings, JCOMM must make the case for what they are doing and why, and encourage even greater collaboration and cooperation.

Recommendation 8b: DMPA will provide its representatives (and encourage the necessary national and international support) to attend meetings of other organizations and committees whose interests intersect.

9. CONCLUSION

This plan presents a review of the various components of data management that must be considered as part of JCOMM. It makes a number of recommendations. Some of these are, in fact, underway either as formal projects in JCOMM, as an activity undertaken by one or more members, or as activities undertaken by other organizations with which JCOMM is linked. Most of the work requires coordination of activities across WMO-IOC Member / Member States participating in JCOMM. Developing this degree of cooperation will be a challenge. The national organizations of each Member / Member State have national priorities and objectives that must be met. Progress will be made by aligning these national requirements with activities at an international scale.

Of course, this is merely a plan and does not lay out the implementation steps. That is something that needs more work as this is where analysis of existing activities, forming working groups, pilot projects and experimenting with ideas will be explored. As a follow-on, once the DMPlan is accepted, an implementation plan should be drawn up that takes the accepted recommendations and lays out a work schedule and target timelines to realize the objectives of the recommendations.

ANNEX- ACRONYMS AND OTHER ABBREVIATIONS AND SYMBOLS

ADCP	Acoustic Doppler Current Profiler
Argo	CLIVAR-GODAE profiling float pilot project (not an acronym)
ASCII	American Standard Code for Information Interchange
BUFR	Binary Universal Form for the Representation of meteorological data (FM 94–XI Ext. BUFR)
CBS	WMO Commission for Basic Systems
CCl	WMO Commission for Climatology
CEOS	Committee on Earth Observation Satellites
CF	Climate and Forecast convention
CMM	Commission for Marine Meteorology (superseded by JCOMM)
CMOC	WMO-IOC Centres for Marine-Meteorological and Oceanographic Climate Data
CREX	Character form for the REpresentation and eXchange of Data (FM 95–XII CREX)
CTD	Conductivity, Temperature, and Depth probe
CWP	Community White Paper
DAC	Data Acquisition Centre
DBCP	Data Buoy Co-operation Panel
DMPA	JCOMM Data Management Programme Area
DMPlan	JCOMM DMPA Data Management Plan
DOI	Digital Object Identifier
E2E	End-to-End Data Management
E2EDM	End-to-End Data Management Pilot Project
ECV	Essential Climate Variables
ET	Expert Team
ET-DRC	Former CBS Expert Team on Data Representation and Codes (now IPET-DRMM)
ETDMP	JCOMM-IODE Expert Team on Data Management Practices
ET-SI	JCOMM Expert Team on Sea-Ice
ET-WCH	Expert Team on Waves and Coastal Hazards Forecasting Systems
FGDC	Federal Geographic Data Committee (USA)
FM-13	Report of surface observation from a sea station (FM 13–XII Ext. SHIP)
FTP	File Transfer Protocol
GARP	Global Atmospheric Research Programme
GCC	MCSS Global Collecting Centre
GCMD	Global Change Master Directory (USA)
GCOS	WMO-IOC-UNEP-ICSU Global Climate Observing System
GCOS-92	Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC
GCOS-IP	GCOS Implementation Plan (i.e., as detailed in GCOS-92, and its 2010 update, GCOS-138)
GDAC	Global Data Assembly Centre
GFCS	Global Framework for Climate Services
GHRSST	Group for High-Resolution SST
GLOSS	JCOMM Global Sea-level Observing System
GODAE	Global Ocean Data Assimilation Experiment
GOOS	IOC-WMO-UNEP-ICSU Global Ocean Observing System
GOSUD	Global Ocean Surface Underway Data Pilot Project
GTS	Global Telecommunication System
GTSPP	Global Temperature and Salinity Profile Programme
HABS	Harmful Algae Blooms
HDF	Hierarchical Data Format
HQ-GDMSC	High Quality Global Data Management System for Climate
HTTP	Hypertext Transfer Protocol
ICOADS	International Comprehensive Ocean-Atmosphere Data Set
ICSU	International Council for Science
IGOSS	WMO-IOC Integrated Global Ocean Services System (superseded by JCOMM)

IOC	Intergovernmental Oceanographic Commission of UNESCO
IODE	International Oceanographic Data and Information Exchange (IOC)
IMM	International Maritime Meteorological format
IMMA	International Maritime Meteorological Archive
IMMT	International Maritime Meteorological Tape
INDO	Indian Ocean
IPET-DRMM	CBS Inter-Programme Expert Team on Data Representation, Maintenance and Monitoring
ISO	International Organization for Standardization
ISO-23950	Information and documentation - Information retrieval (Z39.50) - Application service definition and protocol specification
JASIN	joint air-Sea Interaction Experiment
JCOMM	Joint WMO-IOC Technical Commission for Oceanography and Marine Meteorology
JCOMMOPS	JCOMM *in situ* Observations Programme Support Centre
MARPOLMON	Marine Pollution Monitoring Programme
MCDS	Marine Climate Data System
MCSS	Marine Climatological Summaries Scheme
META-T	Water Temperature Metadata Pilot Project
MT0	BUFR Master Table number 0 (Meteorological Data)
MT10	BUFR Master Table number 10 (Oceanographic Data)
NetCDF	Network Common Data Form
NODC	National Oceanographic Data Centre (IODE)
OceanSITES	OCEAN Sustained Interdisciplinary Timeseries Environment observation System
ODAS	Ocean Data Acquisition System
ODASMS	ODAS Metadata Service
ODP	IODE Ocean Data Portal
ODSBP	Ocean Data Standards and Best Practices
OGC	Open Geospatial Consortium
OOPC	Ocean Observations Panel for Climate
OPA	JCOMM Observations Programme Area
OPeNDAP	Open-source Project for a Network Data Access Protocol
OSMC	Observing System Monitoring Centre
PA	Programme Area (of JCOMM)
Pub 47	WMO Publication No. 47 (International List of Selected, Supplementary and Auxiliary Ships)
QARTOD	Quality Assurance of Real-Time Oceanographic Data
QC	Quality Control
RNODC	Responsible National Oceanographic Data Centre (IODE)
RNODC / DB	Responsible National Oceanographic Data Centre for Drifting Buoys (IODE)
SensorML	Sensor Model Language
SOC	Specialized Oceanographic Data Centre (of former IGOSS, then JCOMM, then deprecated to MCDS GDAC)
SOO	Ship of Opportunity
SOOP	SOT SOO Programme
SOOPIP	SOOP Implementation Pnel
SOT	JCOMM Ship Observations Team
SFSPA	JCOMM Services and Forecasting Systems Programme Area
SST	Sea Surface Temperature
TAC	Traditional Alphanumeric Code form
TDC	Table Driven Code form
TT-SAT	JCOMM cross-cutting Task Team on Satellite Data Requirements
TT-TDC	DMPA Task Team on Table Driven Codes
UN	United Nations
UNEP	United Nations Environment Programme
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNFCCC	United Nations Framework Convention on Climate Change
VOS	Voluntary Observing Ship

VOSClim	Voluntary Observing Ship Climate Project
WCRP	WMO-IOC-ICSU World Climate Research Programme
WCP	World Climate Programme
WDS	ICSU World Data System
WESTPAC	IOC Sub-committee for the Western Pacific
WIGOS	WMO Integrated Global Observing System
WIS	WMO Information System
WMO	World Meteorological Organization
WOCE	World Ocean Circulation Experiment
WWW	World Weather Watch
XBT	Expendable Bathythermograph
XCTD	Expendable Conductivity, Temperature, and Depth probe
XML	Extensible Markup Language