Limited distribution **IOC/IODE-XIX/48**
21 January 2007
Original: English

INTERGOVERNMENTAL OCEANOGRAPHIC COMMISSION
(of UNESCO)

Nineteenth Session of the IOC Committee on International Oceanographic Data and Information Exchange (IODE-XIX)
Trieste, Italy, 12-16 March 2007

JCOMM Data Management Strategy

By Robert Keeley, Coordinator JCOMM Data Management Programme Area

This document is a draft version prepared by the author on 11 December 2007.

Table of Contents

JCOMM Data Management Plan

1. The Vision and Objectives of JCOMM

The stated vision of JCOMM is that of an organization which coordinates, regulates and facilitates, 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://ioc.unesco.org/jcomm/ at IOC and http://www.wmo.ch/web/aom/marprog/marprog.html at WMO.

The long-term objectives of JCOMM are:

(i) To enhance the provision of marine meteorological and oceanographic services in support of the safety of navigation and safe operations at sea; contribute to risk management for ocean-based economic, commercial and industrial activities; contribute to the prevention and control of marine pollution, sustainable development of the marine environment, coastal area management and recreational activities, and in support of the safety of coastal habitation and activities; 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;

(ii) 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, as part of the Global Ocean Observing System (GOOS) and the World Weather Watch (WWW), and in support of the World Climate Programme (WCP), the World Climate Research Programme (WCRP), the Global Climate Observing System (GCOS), and other major WMO and IOC programmes;

(iii) To coordinate and regulate the maintenance and expansion of a comprehensive database of marine meteorological, oceanographic and sea ice data, in support of marine services, operational meteorology and oceanography and the WCP;

(iv) To manage the evolution of an effective and efficient programme through the selective incorporation of advances in meteorological and oceanographic science and technology; and to work to ensure that all countries have the capacity to benefit from and contribute to these advances, and to the work of JCOMM in general.

2. Purpose and Scope of this Plan

JCOMM was formed from two agencies, one in the Intergovernmental Oceanographic Commission, IOC, and one in the World Meteorological Organization, WMO. The IOC contribution was the Intergovernmental Global Ocean Services System, IGOSS. It 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, the WMO contributed the Commission on Marine Meteorology, CMM. Their 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 they built and maintained archives of marine data. This plan 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 2006) has three Programme Areas, PAs, one for Operations (OPA), one for Services (SPA) and one for Data Management (DMPA). When JCOMM was formed a decision had to be made about how to organize the cross cutting activity that is data management. The groups in the OPA and SPA 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. 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 on a par with 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. It is the challenge to the DMPA to work within the requirements of the activities within the OPA and SPA 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 strategy, 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 part of 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. JCOMM must position itself to handle this broader range of variables.

New observing technologies are being developed at a fast pace. In addition to 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 observations with 2-way communication with the sensors. Open ocean moorings with both meteorological and oceanographic measurements available immediately are coming into being. All of these data will be challenging for JCOMM to 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 in the oceanographic community. Now, ocean and atmospheric models are starting to be fully coupled together. 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. They permit us to fill in data gaps and to predict conditions where data are sparse. When these results are reformulated as products they become particularly useful in decision making, disaster mitigation and a host of other uses.

Models can be operated in research or in operational 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 only.

Results from operational models are the basis for many products that translate into decision making tools. These results are valuable to save and can be used in the same way as are historical observations. That is, new products can be derived from model archives and incorporated into decision making.

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 JCOMM Data Management Strategy 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. No common terminology has developed to talk about these different kinds of metadata, so this document will provide examples of the kind of information considered whenever the term is used.

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; the management of the real-time data was left to IGOSS. IODE is a close partner in managing the oceanographic data and is a co-sponsor of some 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 http://www.wmo.ch/web/www/WISweb/home.html). 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, being co-sponsored by WMO, is a contributor to WIS.

Both oceanographic and meteorological data contribute to the holdings within the World Data Center, WDC, system (http://plato.wdcb.rssi.ru/wdc/wdcmain.html). It is expected that the WDC system will ultimately archive all of the data collected by JCOMM. The JCOMM data management activities, therefore, need to provide data and information to those WDCs and work with them to build a complete, global data system.

Finally, the creation of the Global Climate Observing System and all of its components places expectations on what JCOMM will provide in support. JCOMM programs are mentioned in

more than 20 of the actions in the GCOS plan (see http://www.wmo.ch/web/gcos/Implementation_Plan_(GCOS).pdf). Of these, some relate directly to the data systems. This plan must provide the direction that will ensure these expectations are met.

Details of 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 programmes are well 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 co-sponsoring agencies of WMO and IOC. This Strategy provides the broad outlines and recommendations by which the DMPA will support attaining the vision of JCOMM. It is not intended that this plan provide the details of how the recommendations will be met. This is the subject of an implementation plan that must be built from the agreed strategy. It should be expected that as technology advances, and as the implementation develops, there will be changes in emphasis or new capabilities not anticipated by this Strategy.

3. Organization of the Plan

An examination of the long-term objectives of JCOMM and other expectations of JCOMM, such as in the GCOS plan, 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.
- information be maintained about the observing practices so that older data may be compared to more recent data without instrument biases.
- there be a degree of standardization in such areas as data formats, content, naming conventions, etc.
- data version control be addressed.
- data management activities and experiences are made available equally to all JCOMM members.

Many contributors to JCOMM have data management capabilities already in place. This means that a JCOMM Data Management Strategy must recognize the distributed nature of the international data system. To meet JCOMM objectives will take an increased level of cooperation and coordination by all members of JCOMM and a number of international organizations. Interoperability will be a central issue. Therefore, a heavy emphasis will be 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.

To organize the discussion, this document will divide the tasks into main areas and make recommendations in each. The major data management themes that will be 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 and the links between JCOMM and other organizations.
- Communications – including the dissemination of information about JCOMM data management, training materials, performance measures, reports, etc.

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 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 limited 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 grown up to manage these data streams.

While there is some hope that limitations on communications bandwidths 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.

Both WMO and IOC have sought for decades to standardize communications of data, largely in the context of reporting data over the Global Telecommunication System (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.

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 GTS

The Global Telecommunications System, GTS (see http://www.wmo.ch/web/www/TEM/gts.html), is one of the means of data exchange used by JCOMM. For WMO it is the transmission mechanism of choice for operational, time critical data exchange. This is still true in the development of the WMO Information System, WIS. Being co-sponsored by WMO, JCOMM will need to adopt this same view.

On the GTS, the Traditional Alphanumeric Code forms (TACs) have strongly regulated formats and contents (for example FM-13 SHIP). This has been of advantage in that there is a commonly understood protocol for naming variables, for units of reporting and additional

information to be sent with the observations. The capabilities and variety of the TACs are more developed for meteorological data than for oceanographic. However, it is increasingly the case that either new variables or new information about how existing variables are observed needs to be exchanged. The TACs are an older technology, with roots extending back to the International Synoptic Land Code agreed to in Rome in 1913. It is a technology that WIS is phasing out.

Replacing the TACs are Table-Driven Codes (TDCs) such as the Binary Universal Form for the Representation of meteorological data, BUFR (see http://www.wmo.ch/web/www/WMOCodes.html). TDCs offer a flexible form that allows data to be structured within the exchange format in ways that are linked to the kind of data observed. As an example, a BUFR message is built using a set of carefully defined tables of variables and attributes, generally using only SI units, with rules about how the information is structured into a well formed message. The BUFR tables show their heritage in that they are much more capable of describing meteorological variables than oceanographic ones. One of the complications of oceanographic data, which does not appear in meteorological data, is that oceanography spans a variety of disciplines. Whereas it is possible to draw the analogy between physical oceanographic variables and meteorological ones, oceanography also includes chemistry, geophysics and biology. The present structure of BUFR tables is not particularly well suited to handling these other disciplines, nor the metadata that are crucial for correct interpretation of the observed values.

There is also a character form of TDCs, in this case closely linked to BUFR, called CREX, (Character form for the Representation and EXchange of data). The chief difference between CREX and BUFR is that a CREX message is in ASCII characters and so is directly readable by people and hence removes the software dependency of BUFR.

WMO has expressed the requirement to move all of its GTS, and possibly other observational data exchanges to use TDCs. JCOMM must be a partner in this process. 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

- devising and validating appropriate BUFR templates that include and extend the information now sent in TACs.
- encouraging ocean and meteorological centres to develop capacity to both read and write TDCs.
- 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 toTDCs.

A number of years before JCOMM was formed, there was some activity to build another set of BUFR tables whose structure is better tuned to the requirements for all kinds of ocean (and some atmospheric) data. BUFR was designed with the capacity to handle a number of different tables, so this was simply exploiting an existing capability. This new set of tables was designated Master Table 10 (MT10) and met the requirements at the time for an alternate set of tables that BUFR could support. No ocean centres were or are capable of encoding or decoding BUFR data and so the use of these tables was not adopted. However, with encouragement from JCOMM to get both ocean and meteorological data reported in BUFR, MT10 should be revisited and evaluated against current needs.

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

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, including software reliability, human readability, 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, 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 process that requires "handshakes" between every partner in the exchange.

4.3.1 netCDF

There are preferred practices that are starting to emerge for data exchange using the Internet. The use of 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), OceanSITES (see http://www.oceansites.org/) and GOSUD (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 when even 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 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 support the widespread use of netCDF as a data exchange format.

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

Working with the groups that maintain and extend netCDF would be a useful activity for JCOMM. Another development that should be noted in this context is the planned convergence of netCDF with the Hierarchical Data Format (HDF see http://www.hdfgroup.org/). Considering long-term archive requirements, it should be noted that formats such as netCDF and HDF, like BUFR, are extremely complex, and highly dependent on software.

Recommendation 4.3.1c: JCOMM stay informed on netCDF maintenance and developments.

4.3.2 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, its 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 a 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.

Recommendation 4.3.2a: DMPA monitor the development of xml and encourage appropriate use for the exchange of data and metadata.

Recommendation 4.3.2b: DMPA encourage the development of vocabularies used in xml that are as close as possible to those used in other formats.

4.3.3 Other Formats and Data Structures

Of course, netCDF is not the only format used for exchanging data on the internet. 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.

There is a present 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 is 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 or communications channels that the full resolution data can not be sent as soon as observed. 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.

Many of the data structures used in delayed mode exchanges of data are either relatively inflexible and consequently difficult to change from a technical standpoint, or the change mechanisms are so cumbersome that required changes take inordinate amounts of time to accomplish. Both of these impede exchange and tend to encourage the creation of "new and better" formats. JCOMM and its other data management partners need to tackle this problem head on to encourage the evolution to more capable exchange formats that are flexible and yet relatively simple to alter.

Recommendation 4.3.3b: JCOMM work with partners to encourage the evolution of exchange formats to more robust forms.

5. Data Processing

Members of JCOMM maintain and support their own national archives as required. The strategies and resources (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 so 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

In the simplest terms, the raw observations coming from an instrument may be considered to be one 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 verification of 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 (GOOS publication no. 42, annex 4) 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.

There is 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.

But 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 reasons, the full resolution data stream coming from oceanographic sensors at sea is 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 archives and sit there until the higher resolution forms arrive in delayed mode. So there is now in existence both a low resolution form and a high resolution form. 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 more difficult process.

There has been a pilot project operating within the Ship Observation Team (SOT) program 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 to be of 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 ocean data 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 delayed mode form arrives. 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 come once more 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 a strategy.

5.2 Data Quality

Assessing the quality of data is a complicated process that uses knowledge of oceanography and meteorology, knowledge of the area and time in which the data are collected. The degree of assessment 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 of errors that pass through the procedures is relatively low, and 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 to call on.

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.

Still, 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 use flags to indicate the quality of the data. These typically are placed on each observation.

There are a number of places where standardization of practice would benefit users. Presently 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 reconciling if a test carried out by one group is functionally the same as a test carried out by another. 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 the pass or fail of a test by an observation or perhaps group of observations. 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 the scheme is that there is no interpretation of the "goodness" of the data, simply a statement of success or failure of a test. Users are therefore able to decide for themselves which tests they wish to take into consideration 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. A second 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.

Many programmes currently use a data quality flagging scheme that provides an assessment of the quality of observations. However, there is no universal way that quality is indicated. Within the ocean community, there has been a tendency to standardize on the flagging scheme inherited from IGOSS. This scheme has some differences with the scheme used in the meteorological community as embodied in the BUFR tables. Finally, another group working in the U.S. is currently devising another interpretation. The problem is that although single digits are used by all, a particular digit value has different meanings for each scheme. Although a user can sort this out with appropriate documentation from each data provider, it is an inconsistency that they should not have to deal with.

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, one can imagine a complicated series of quality flags that may well be confusing to a user. The other extreme is for the receiver to use the attached flags as guidance, but if additional 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 work with all appropriate bodies to come to agreement on a single scheme to indicate quality of data.

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

JCOMM is interested in data that arrive both in real-time and in delayed mode. In a broader context, the data collected by JCOMM are transferred to various archive centres and projects around the world. In this way many copies of the same original data exist in many different centres. But because of processing that takes place at each of these centres, the data and information 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. The result is that a user will get different copies of the data depending on who has provided the copy, or perhaps duplicate or near duplicate copies of the original.

The same thing can, of course, occur in an individual archive where the same data arrive from two different sources. If these arrive at separate times, or are processed by different people, the fact that they are duplicates may escape notice. Then, when the data are provided to a user, these duplications may have undesirable impacts on the analyses.

The detection of exact duplicates is relatively straightforward and can be taken care of through 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 is inserted at the surface derived by extrapolation. 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 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 archives and passed from one to another, 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. In fact schemes based on this idea are being employed in some data systems already.

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 in that as an archive adopted the practice of employing unique identifiers, 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. There are, for example, 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.

But 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.

A few years ago, this idea was presented in the context of a discussion of Marine XML. The idea was that information could be assembled into packages of standard content that were called “bricks” (see http://ioc.unesco.org/iocweb/iocpub/iocpdf/IODE17_19_sgxml.pdf). For example, standard content for the origin of data would be information about who sent the data, their address, the name of the project under which the data were collected, the name of the platform and so on. The essential idea was that each brick dealt with discrete and separable information. It is possible to construct bricks to contain observations and in this, for example, could be found data quality flags. A suite of bricks were envisioned that contained the varied information content of both measurements and information that make up a collection of data. By assembling the bricks in different ways, a variety of data types could be built into a collection. The most powerful aspect of this, though, is that adopting such a strategy would standardize the content and thereby remove the ambiguities that are present in many data exchanges.

The metadata standards developed as part of ISO (International Standards Organization) or FGDC (Federal Government Data Committee) have at their heart similar ideas. However, they do not yet go as far into the detail and content of observations, nor do they yet provide the flexibility. Still, because there are similarities, it would be worthwhile to take this work into consideration when deciding what information to include and how the information should be structured.

Recommendation 5.4: JCOMM explore the ideas embedded in xml "bricks" as a standard way to organize and preserve information and data.

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 to another, 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. 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 a 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

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 and is strongly linked to both the WMO Information System developments (see http://www.wmo.int/web/www/WISweb/home.html) 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.

There is the well known ISO organization which develops a broad collection of standards including for metadata. (see http://www.iso.org/iso/en/ISOOnline.frontpage).

Each of the 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. But 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). They divide 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. It would be far better to take the approach of developing categories of metadata, and to define the content to suit the purpose. We 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 on 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 (Federal Government Data Committee, see http://www.fgdc.gov/) standard used within the U.S., in the ISO19115 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: JCOMM define its requirements for discovery metadata and embody these in a formal metadata structure.

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 meteorology, there is Publication 47: International List of Selected, Supplementary and Auxiliary Ships (see http://www.wmo.ch/web/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 and platforms (see http://jcomm.coi.gov.cn/ for information about this system called ODAS). 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 such things 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 member states 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 and reflects the most recent information.

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.

Recommendation 5.6e: JCOMM to develop a strategy for managing the international suite of these metadata sources so that they are easily found and used.

5.7 Model Data

Computer modeling is an activity carried out in many JCOMM countries, 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 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. These latter characteristics define the key attributes that determine if model results have value to archive. JCOMM should consider the results of such operational 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 how long it should be archived for. In addition to this, the data assimilation schemes, observational inputs used, 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 SOCs and RNODCs

SOCs (Specialized Oceanographic Centres) are a legacy from IGOSS. They were formed by member agencies meeting a particular need within IGOSS and volunteering to carry out this activity. SOCs were of different kinds, such as focused on managing data or on capacity building activities. These included SOCs for BATHY and TESAC data, for sea level, and surface drifters.

RNODCs (Responsible National Oceanographic Data Centres) were a creation of IODE. These included RNODCs for the Southern Ocean, for surface drifters, for MARPOLMON, for IGOSS, for WESTPAC, for JASIN, for ADCP and for INDO. In the last review of IODE, it was determined that the RNODC system was working only in a few cases. Consequently, they were abolished with IODE intending to find another mechanism to support the functions. In fact, there is still a strong reason for the activities performed by some of the former RNODCs. This issue is not resolved at this point in time.

Where it is of interest to JCOMM, the activities of former SOCs and RNODCs should be formally recognized either as a component of the appropriate PAs, or adopted as an extension of other activities of interest. Where there are overlaps of activities, actions should be taken to decrease the level of overlap and to enhance the cooperation.

Recommendation 5.8: JCOMM and IODE seek efficiencies in the operations of former SOCs and RNODCs.

6. Access

Finding data of interest in the world of distributed archives and data sources is not easy and it certainly is 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. It is therefore 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 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, through the use of the FGDC metadata standard and the ISO model. There is much work going on now 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 ISO23950 allows queries to cross from one catalogue to another. This achieves interoperability, without the requirement of centralizing the catalogue. These 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 explored and tested.

Recommendation 6a: JCOMM pursue the creation of standards for data discovery metadata and encourage these to be used to support interoperable catalogue services and registries.

Recommendation 6b: JCOMM explore how commercial search engines can be used as another way to search catalogues so that users can use internet tools to locate data.

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 posed by a user. 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 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 Open Geospatial Consortium (OGC see http://www.opengeospatial.org/) standards such as in web map services, may be of use. This strategy would identify a set of data browse services that every archive centre could 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, described later.

Recommendation 6c: JCOMM explore the implementation issues of existing or proposed methods for supporting browse functions.

6.3 Data Delivery

Once the source of data has been identified and the data of interest have been verified to be in the archives, the next step is to provide access to the data. Ideally, this would take place through an interactive query that selects only the data of interest and passes the results directly to the person making the query. But it is unrealistic to expect that this can be done in every case. The reality is that some data providers do not have the computer capabilities to support this service. In other cases, the volume of data requested may be so large that a server trying to provide the data would fail. A further consideration is the national policies on providing data. Each member state needs to analyze their capabilities and national policies and determine for itself what can be supported.

Recommendation 6d: Each member state of JCOMM needs to examine its ability 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 is discussed later.

Not all data users will be satisfied by the functionality provided by WIS. This means that JCOMM will also need to participate in other implementations 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 6e: 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.

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. Such climatologies are very valuable in testing whether newly arrived data appear to have unusual values and so may, in fact, be in error. At the moment, whatever climatologies are used, they may not be the same everywhere and this can cause confusion.

Such climatologies should be built with the active collaboration of appropriate members of the scientific community. Such collaboration ensures that there are sound principals behind the choices made in 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 model 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, HABS, sudden deepening of storms, etc.

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 SPA. 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 6f: DMPA consider ways to collaborate or build products that have wide spread applicability to members.

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. This metadata is 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 6g: JCOMM 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 member states of JCOMM. Both WMO (see http://www.wmo.ch/web/spla/Res40Cg-XII.doc) and IOC (see http://www.iode.org/contents.php?id=200) have data policies that have been constructed with careful consideration of the view of member states. Different countries will have more explicit policies that will apply to exchanging data, or delivering data to clients as envisaged in this Strategy. Clearly JCOMM needs to operate within these intergovernmental and national policies.

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. JCOMM wishes to promote as open access as possible and achieving this while guarding assets complicates the technology solutions. Never-the-less, JCOMM must recognize this as a necessity and encourage members to take all appropriate precautions.

7. Coordination and Linkages

As an international organization co-sponsored by WMO and IOC, JCOMM has many connections among its internal programmes as well as to organizations and groups outside. 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 at sea observation programs. It includes programs using surface drifters (Data Buoy Cooperation Panel, DBCP), Volunteer Observing Ships (VOS), Ships Of Opportunity (SOOP), the international tide gauge network (GLOSS), 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 DBCP, recognizing the limitations of their initial systems are now building new ones to hold information about their platforms, (i.e. ODAS).

The SPA also has observing programs incorporated into their work. For examples, the ET concerned with waves and the ET on ice 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 SPA 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 SPA activities. The interaction between the data managers of the different groups has been only through informal discussions with the result that there is only a small degree of commonality.

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

It was recognized early 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 appropriate 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.

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. There is much work to be done in the domain of standards. However, there have already been significant activities by other groups so that all of what is needed may not have to be developed by JCOMM. Indeed, 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 choice, JCOMM may need to devise its own standards, though this should not be done without careful consideration.

Recommendation 7.1c: JCOMM should first adopt an existing standard or best practice, as a second option adapt an existing one, or failing that create its own.

JCOMM will need a process to adopt, adapt or create its standard practices. There is no such process at the moment, though there are examples of similar activities such as within the WMO domain in such committees as the ETDRC (Expert Team on Data Representation and Codes) and elsewhere. Because JCOMM should only, as a last resort, create its own standards, it does not require the same process as in ISO or OGC. Instead, JCOMM requires a process that can recognize where standards are required, identify candidates to be considered, evaluate candidate practices and then recommend their use across JCOMM. The accreditation process for standards will require both a group to coordinate this activity and assistance by JCOMM members to take part in the evaluation process.

Recommendation 7.1d: JCOMM develop a process to accredit standards to be recommended for use across all activities.

Recommendation 7.1e: DMPA develop a plan for coordination of the accreditation process and carrying out of evaluations.

As a standard is adopted, this information must get out to JCOMM members and they will need to take steps to implement it. There will, therefore, be a role for communications and a repository for the documentation of the standards used by JCOMM. This could well be served by JCOMMOPS, or some other suitable and widely visible agency.

Recommendation 7.1f: JCOMM establish a highly visible and accessible repository where information about JCOMM standards can be found.

Members will have varying abilities to respond to adopting recommended standards. It is unlikely that a standard will be implemented across all JCOMM members simultaneously. Indeed, if this is a requirement for a standard to be effective, 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: As part of the accreditation process, consideration must be given to how to implement the standard across JCOMM members as rapidly as possible. Due consideration should be given to how capacity building resources may be used.

Coordination must also take place with the other programmes in IOC, WMO, 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, DMPA needs to adopt a reporting process whereby members 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 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.

Within the OPA there has developed a simple quarterly reporting system (see http://www.oco.noaa.gov/index.jsp?show_page=page_status_reports.jsp&nav=observing) whose goal is to provide a concise view of where the observing system stands in meeting GOOS objectives. The target audience for this report is senior members of governments who have the ability to influence budgets.

The results shown in these quarterly reports are still limited to only 5 variables whereas there are a number of other variables that should also be represented. These reports represent one measure of how well the data systems are functioning. DMPA needs to provide the necessary support to fill out the missing portions of this reporting system.

Recommendation 7.1j: DMPA in collaboration with OPA, and SPA encourage the completion of quarterly reporting of other important variables following the model used by OPA.

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. There are a number of measures that can be thought of as possible candidates to be used by all data systems. This could include measures such as:

- the percentage of data reporting in real-time with detected problems.
- the percentage of total data received that report in real-time.
- the time delays between receiving real-time and delayed mode versions of the same measurements.
- the mean time to report a real-time observation (report time – observation time)

Other measures can be imagined. The objective of defining such a list 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 in OPA and SPA.

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

7.2 With IODE Activities

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. 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 the work needed to manage the data is diverse.

Recommendation 7.2a: IODE and JCOMM formalize the relationship between the organizations. It is suggested that the chair of IODE be named a member of the DMPA-CG and the chair of the DMPA be named an Officer of IODE.

Where there is a high degree of overlap of interests in types of data, it is important to consider streamlined operations. In this spirit, IODE and JCOMM share the ETDMP (Expert Team on Data Management Practices) 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 the opposite is also true. 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 data streams and archives.

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, OOPC, GHRSST, Argo, etc. JCOMM is implicated in many of these.

JCOMM figures prominently in the GCOS Plan, being mentioned in 24 actions. These span the range of JCOMM activities in all Program Areas. Those with direct mention of data management functions include:

- O6: Improve meta-data acquisition and management for a selected, expanding subset of VOS (VOSClim) together with improved measurement systems.
- O11: Ensure real-time exchange and archiving of data. Ensure historical sea level records are recovered and exchanged.
- O24: Promote development of integrated analysis products and reanalysis using historical data.
- O33: Develop and implement comprehensive data management procedures.
- O34: Undertake a project to develop an international standard for ocean meta-data**.**
- O35: Undertake a project to apply the innovations emerging from the Future WMO Information System initiative, 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**.**
- O36: Plan and implement a system of regional, specialized and global data and analysis centres.
- O39: Develop plans for and coordinate work on data assembly and analyses**.**

This strategy document addresses meeting the first part of O33. Implementation of the recommendations of this strategy will meet the second part of O33 and will address all of the other actions.

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

The GODAE project (see http://www.bom.gov.au/bmrc/ocean/GODAE/) is nearing its end. It has developed a number of products, intercomparisons, common output strategies and so on. The work that has been done is important and directly relevant to implementing recommendations that are part of this strategy. Similarly, the GHRSST (see http://www.ghrsst-pp.org/) project has much to offer in demonstrating, among other products, how in-situ and satellite observations can be used together. Argo (see http://wo.jcommops.org/cgi-bin/WebObjects/Argo) shows 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 implementation for JCOMM.

Recommendation 7.3b: JCOMM must work closely with the many other IOC programs in developing its implementation plans so that the experience gained is used in implementing recommendations of this strategy.

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 JCOMM members to fully participate in data management and to use the managed data must be supported. But 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 makes more sense for JCOMM to collaborate with these efforts. This can be, for example, by ensuring suitable training materials on marine operations are present and by contributing instructors as appropriate. Joint activities between member states is another way to increase members capabilities.

Recommendation 7.3c: JCOMM should collaborate with existing WMO and IOC 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 (see http://www.wmo.ch/web/www/WISweb/home.html). The linkages of the meteorological side of JCOMM are closer to the developments of WIS than the oceanographic side. However, at the invitation of WIS, the DMPA chair and a member of the CG were invited to a meeting to discuss the development of WIS. 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 E2EDM (End to End Data Management) pilot project. At the meeting, connections were made between the work of E2EDM 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 ETDRC, who regulate how data are presented on the GTS, to the 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: DMPA ensure appropriate experts are fully engaged in appropriate WMO activities.

7.5 With ICSU WDCs

There is a hierarchy of archives that exist in the world and into which JCOMM data management activities fit. On the broadest international level there are the World Data Centers (WDCs, http://www.ngdc.noaa.gov/wdc/wdcmain.html). These were set up many years ago by ICSU (International Council for Science, see http://www.icsu.org/index.php) for various disciplines, including meteorology and oceanography, and they continue to operate. Their mandate has been to act as the global archive for data of one kind or another such that a client anywhere in the world could come to a WDC and find any data that might have been collected. WDCs rely on data exchange agreements with national data centres.

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

Recommendation 7.5a: DMPA initiate a discussion with WDCs to build stronger links between the observing and archive systems and how WDCs operate. This should be done with appropriate other partners.

Because of the position of WDCs in the international data system, they are the focal point for all of the data. Besides ensuring the safe keeping 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 can both contribute by timely provision of data to WDCs and benefit from recent and appropriate and timely updates to the global data set and climatologies.

Recommendation 7.5b: JCOMM members support the timely assembly of data in WDCs 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 JCOMM members 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. The experience is valuable and the solutions are worthy of 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 enhance greatly the capabilities of JCOMM to meet its objectives. However, it is clear that JCOMM is not an organization operating in isolation from others. Implementing the recommendations is only completing part of the work. It is important that after collaboration with our partners on many of their issues, JCOMM communicate these results to the wider world. It is impotant to tell others what is being done and why, and when results can be shown. It is through this process that others will understand what JCOMM is doing and express interest in joining or pointing out similar endeavours being made by them.

One way to provide information is to use the Internet and WWW technology. JCOMM already has a web site. There are also sites, e.g. http://wo.jcommops.org/cgi-bin/WebObjects/JCOMMOPS and http://icoads.noaa.gov/etmc/, 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 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 cooperation.

Recommendation 8b: DMPA will provide members 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 a formal project 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 JCOMM Member States. Developing this degree of cooperation will be a challenge. The national organizations of each member 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 since this is where analyses of existing activities, forming working groups, pilot projects and experimenting with ideas will be explored. As a follow on, once this plan 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 objective of the recommendation.

[end]