WORLD METEOROLOGICAL ORGANIZATION

INTERGOVERNMENTAL OCEANOGRAPHIC COMMISSION (OF UNESCO)

JOINT WMO/IOC TECHNICAL COMMISSION FOR OCEANOGRAPHY AND MARINE METEOROLOGY (JCOMM)
SHIP OBSERVATIONS TEAM

FOURTH SESSION

GENEVA, SWITZERLAND, 16 TO 21 APRIL 2007

SOT-IV/Doc. I-6.1
(15.II.2007)

ITEM I-6.1

Original: ENGLISH

PROGRAMME OPERATIONS AND DEVELOPMENT

Data Standards

(Submitted by the Secretariats)

Summary and purpose of document

This document is advising on recent recommendations by the JCOMM Data Management Coordination Group regarding observational data standards. These are being expressed in the context of the JCOMM Data Management Strategy.

ACTION PROPOSED

The Ship Observations Team is invited to:

(a) to comment on these developments;

(b) to relay any specific recommendation it may have regarding the JCOMM Data Management Strategy to the DMCG;

Appendix: None.

DISCUSSION

1. Introduction

1.1 At its second session, Geneva, 10-12 October 2006, the JCOMM Data Management Coordination Group (DMCG) reviewed and updated a draft of the JCOMM Data Management strategy. The strategy has then been discussed through a wider audience. The Strategy will be finalized after discussions at the Nineteenth Session of the IOC Committee on International Oceanographic Data and Information Exchange (IODE-XIX), Trieste, Italy, 12-16 March 2007. The draft strategy includes guidelines regarding the maintenance and development of observational data standards for the JCOMM community. Any developments with regard to observational data standards should eventually be consistent with the Strategy, which is encouraging JCOMM to 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. Much of the text below includes excerpts from the draft strategy directly related to observational data standards. Specific recommendations from the draft strategy are indicated in brackets.

1.2 The strategy recognizes that 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.

1.3 According to the strategy, 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. 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.

1.4 The strategy therefore makes the following recommendations:

- JCOMM to develop a process to accredit standards to be recommended for use across all activities (Rec. 7.1d).
- DMPA to develop a plan for coordination of the accreditation process and carrying out of evaluations (Rec. 7.1e).
- JCOMM to establish a highly visible and accessible repository where information about JCOMM standards can be found (Rec. 7.1f).
- 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 (Rec. 7.1g).

2. Real-time and delayed mode data exchange

2.1 The following segments have to be considered for real-time data exchange:

- Data transmission from the sensors to the acquisition platforms
- Data collection from the observing platforms
- Data distribution to data users and data exchange

2.2 The Strategy indicates that 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. It recommends that JCOMM should encourage instrument manufacturers to standardize the formats of the data and information coming from instruments used at sea (Rec. 4.1).

2.3 As far as real time and delayed mode data exchange, a number of data exchange systems are being used (see below). Some data formats can be used for both real-time and delayed mode or archival purposes.

2.3.1 The WMO Information System (WIS) and its Global Telecommunication System (GTS) are now promoting the transition from Traditional Alphanumeric Codes (TAC) to Table Driven Codes (TDC). TDC are used for both real-time and archival purposes at operational centres.

The DM strategy makes the following recommendations:

- DMPA to lead the development of the detailed plan to change GTS data reporting from TACs to TDCs (Rec. 4.2a).
- The DMPA in association with the appropriate WMO committee should evaluate BUFR Master Table 10 (MT10) for its relevance to present needs (Rec. 4.2b).
- 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 (Rec. 4.2c).

Detailed information about TDC aspects is provided in document I-6.2.2.

2.3.2 Data distribution systems developed for Argo, OceanSITES, GOSUD, and GTSPP

These systems are basically using formats such as NetCDF. They are used for both real-time and delayed mode purposes as well as for archiving purposes. The strategy makes the following recommendations:

- JCOMM to support the widespread use of netCDF as a data exchange format (Rec. 4.3.1a).
- JCOMM to encourage usage of Climate and Forecast (CF) convention for variable naming in netCDF and stay informed of CF updates to meet JCOMM contributors' needs (Rec. 4.3.1b).
- JCOMM to stay informed on netCDF maintenance and developments (Rec. 4.3.1c).

2.3.3 XML and other formats

2.3.3.1 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.

2.3.3.2 The flexibility of xml is also one of its weaknesses. Both the sender and receiver of the message must understand the meaning of the xml tags. 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.

2.3.3.3 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.

2.3.3.4 The strategy makes the following recommendations:

- DMPA to monitor the development of xml and encourage appropriate use for the exchange of

data and metadata (Rec. 4.3.2a).
- DMPA to encourage the development of vocabularies used in xml that are as close as possible to those used in other formats (Rec. 4.3.2b).
- 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 (Rec. 4.3.3a).

4. Marine climatology and archiving

4.1 Besides the data systems listed above that may also be used for delayed mode data exchange or for archiving, the formats used in the context of the Marine Climatological Summaries Scheme (MCSS) such as the International Maritime Meteorological Tape (IMMT) format need to be considered. For example, Version 3 of the IMMT format was adopted by JCOMM-II.

4.2 Also, a standardized International Marine Meteorological Archive (IMMA) format required for the exchange of historical ship data digitized from national logbooks is being developed by the Expert Team on Marine Climatology (ETMC). The format is operational for the International Comprehensive Ocean-Atmosphere Data Set (ICOADS), for the Climatological Database for the World's Ocean 1750-1854 (CLIWOC), and helping to meet requirements of the VOSClim project. The format is similar to IMMT in using a simple ASCII representation, but more flexible and extensible through the definition of "attachments" tailored to the requirements of different marine data types (e.g., historical or contemporary).

4.3 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. The DM strategy therefore recommends that JCOMM work with partners to encourage the evolution of exchange format to more robust forms (Rec. 4.3.3b).

4.4 Also, 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.

4.5 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.

4.6 The DM Strategy therefore makes the following recommendations in this regard:

- 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 (Rec. 7.5a).
- JCOMM members support the timely assembly of data in WDCs and encourage timely updates and distribution of the global data sets and climatologies (Rec. 7.5b).

5. Metadata

5.1 There are different types of metadata to consider depending upon their applications. These range from metadata about the observing instruments or platforms, to discovery metadata and metadata profiles. Current initiatives are focusing for example on:

- Ontologies (the science of describing the kinds of entities in the world and how they are related) and vocabularies such as the Marine Metadata Interoperability Project.
- Operational and scientific application of the observational data (e.g. META-T, ETDMP) and are mainly interested on characteristics of instruments, data quality, etc.
- Describing document origins and contents (e.g. the Dublin Core Metadata Initiative)
- Interoperability of data management systems (e.g. WIS, SeaDataNET) and are mainly interested in metadata profiles.

5.2 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. 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.

5.3 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.

5.4 The DM strategy is proposing the following recommendations:

- The DMPA to examine existing metadata initiatives to develop a categorization that aligns with the purpose of the metadata (Rec. 5.6a).
- The DMPA to use the metadata categorization to develop a plan on which metadata initiatives align with its work and become engaged in these activities (Rec. 5.6b).
- JCOMM to encourage all agencies keeping information about instruments, platforms, etc., to place this information on-line and keep it up-to-date (Rec. 5.6d).
- JCOMM to develop a strategy for managing the international suite of these metadata sources so that they are easily found and used (Rec. 5.6e).
- JCOMM to define its requirements for discovery metadata and embody these in a formal metadata structure (Rec. 5.6c).
- JCOMM to pursue the creation of standards for data discovery metadata and encourage these to be used to support interoperable catalogue services and registries (Rec. 6a).
- JCOMM to explore how commercial search engines can be used as another way to search catalogues so that users can use Internet tools to locate data (Rec. 6b).
- JCOMM must consider interoperability issues with satellite data providers so that satellite and in-situ data are easily compared (Rec. 7.1b).
- JCOMM must develop a level of interoperability in data management with other major international and significant national programmes (Rec. 7.6).

_____

Appendix: None