STUDY ON INTEGRATION OF DATA MANAGEMENT ACTIVITIES BETWEEN WMO PROGRAMMES

by

David E. McGuirk

December 2003

CONTENTS

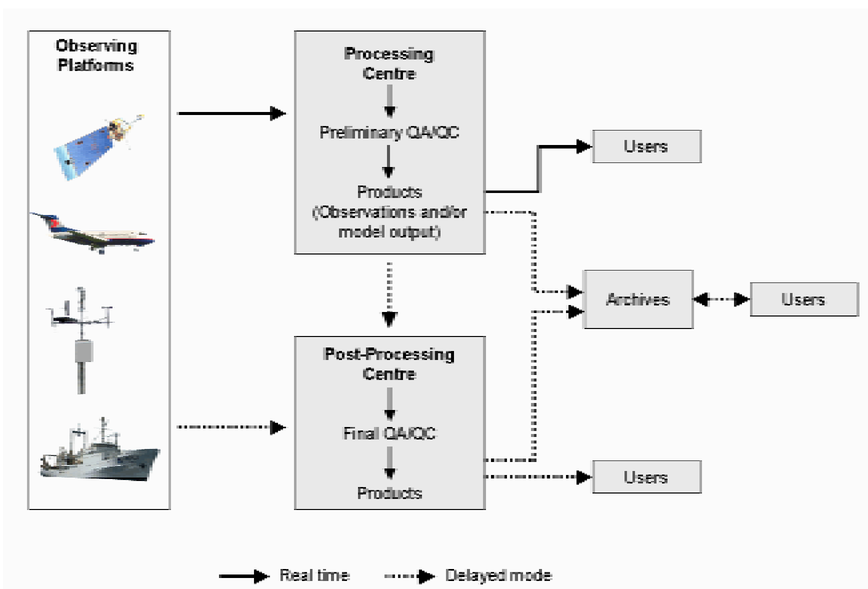
More data, more data
Right now and not later.
Our storms are distressing,
Our problems are pressing.
We can brook no delay
For theorists to play.
Let us repair
To the principle sublime
Measure everything, everywhere
All of the time.

*Aaron Fleishe*r, 1957
Theme song of the Sixth
Weather Radar Conference

It has been recognised for many years that data and its effective management are essential for the success of scientific programmes. Data is the foundation on which our knowledge of the environment is built and data management is essential to ensure effective and efficient use of this resource. This is particularly true for international programmes, where data must be exchanged and understood regardless of language.

Most WMO Programmes have instituted various measures and activities to support their data management requirements. Although the aggregated requirements for each of the Programmes are unique, there are similarities and overlaps in many of their specific requirements. It is therefore likely that benefits could be gained from further coordination and integration of related data management activities among the Programmes.

The current state of the ensemble of WMO information systems is extremely complex. However, a review of the overall data management strategies for several Programmes reveals that most have several fundamental capabilities in common. Nearly all Programmes collect data, transmit these data to one or more processing centres, perform quality control, generate products, transmit these products to users, and archive the data and products for future use. Several Programmes further divide the data flow into real-time and delayed mode data streams as shown in the figure below.



Real-time and operational data flow is often essentially one-way, that is, products are routinely transmitted to collection centres or broadcast to users without any explicit action from the recipient. A variety of telecommunication services are used for real-time transmission including leased fixed circuits (both terrestrial and satellite) and the Internet. Delayed mode transmission utilizes the Internet, post and, less often, private circuits.

Access to data held in archives is usually provided on a two-way or request/reply basis. That is, users contact the archive, request products and these products are then transmitted to the user. Although the Internet is rarely used for real-time data transmission it is commonly used for access to archives.

Several Programmes utilize the standard protocols and data representation forms of the WMO Basic Systems for transmission of their real-time and operational delayed-mode data. These include the WWW, Marine Meteorology and Oceanographic Activities, the Aeronautical Meteorology programme, and the World Climate Programme. However, other Programmes utilize processing procedures that are entirely unique. Furthermore, even those Programmes that utilize the Basic Systems for operational needs have usually developed their own procedures to handle archives, research data, and ad hoc requests. For the most part these have been developed with little or no coordination between Programmes.

This lack of standardisation has led to a number of problems. Notably:

- There is little connectivity between applications developed to serve the needs of different Programmes
- There are a large number of different applications whose development has not been coordinated making integration of data sets technically challenging
- Multidisciplinary application of meteorological, hydrological and oceanographic data is hampered by lack of agreed standards needed to effectively identify, acquire and use all of the relevant data

The multiplicity of systems operated for different Programmes has resulted in incompatibilities, duplication of effort and higher overall costs for Members. Continuing to develop systems in this uncoordinated manner will exacerbate these problems and will further isolate the WMO Programmes from each other and from the wider environmental community. It will increase the difficulty in sharing information between programmes, which is growing increasingly important for inter-disciplinary research.

This problem has been recognised within WMO for many years and a number of efforts have been made to address it. Beginning in the mid 1990s a series of inter-programme data management coordination meetings were held. Representatives of all of the scientific and technical programmes of WMO, as well as representatives of other international programmes with significant data management components were represented. These meetings provided a useful forum to exchange views and information on the activities of the various programmes and led to the effort to develop a comprehensive WMO information system, which could encompass and address the needs of all WMO Programmes.

WMO Congress, at its fourteenth session, agreed that an overarching approach to information management within WMO was required: a single coordinated global infrastructure, the Future WMO Information System (FWIS). FWIS would be used for the collection and sharing of information for all WMO and related international programmes. The FWIS vision provided a common roadmap to guide the evolution of the information system functions performed by current WMO Programmes into an integrated system that efficiently meets all of the requirements of Members for the relevant international environmental information.

Regarding data management standards, principles, practices and guidelines currently in place, for the most part, Programmes provide overall guidance on preferred practices for data management but do not specify binding standards. For Programmes that comprise sub-programmes, each of the sub-programmes has usually instituted its own data management practices and procedures with little guidance or direction from the parent Programme. A notable exception to this trend is the WWW, which has a long history of specifying comprehensive standards within the WMO Technical Regulations. The breadth and detail of data exchange standards and procedures specified for the WWW have certainly contributed to the success of the WWW over the past 40 years. Nonetheless, these regulations are perceived to be complex and inflexible by some of the other Programmes, which have consequently managed their own data management activities with more flexible and informal arrangements.

It is extremely difficult, if not impossible, to specify specific requirements for the diverse set of information needed to meet the requirements of all of the Programmes. Only a few Programmes have defined and documented most of their data management requirements. For the majority of Programmes, some requirements have been specified, usually for real-time data, while other requirements are only vaguely known.

By considering the requirements that have been articulated by one or more of the WMO Programmes or projects, the following emerging requirements can be identified.

- A widely available and electronic (on-line) catalogue of all meteorological and related data for exchange to support WMO Programmes is required.
- It should be possible to rapidly integrate real-time and non-real-time (archive) data sets to better interpret weather events in a climatological context.
- There is a need to identify the potential of observation sites established by one Programme to meet the requirements of other Programmes.
- Need to harmonize data formats, transmission standards, archiving and distribution mechanisms to better support inter-disciplinary use of data and products.
- Require a standard method for station numbering beyond the existing WMO numbers, which is not adequate to define all GAW, climate, hydrological or agromet stations.
- Need standard practices for the collection, electronic archival and exchange of metadata, both high-level and detailed, especially for stations and instruments.

The analysis of ongoing and planned activities of the WMO and related programmes indicates that there has been good cooperation and coordination between most Programmes. However, there is also considerable room for improvement. In general, shortcomings of existing data management activities within the WMO Programmes can be traced to one or more of three basic causes:

- Insufficient recognition of the importance of data management in programme planning
- Insufficient knowledge, resources or commitment to management of data
- Inadequate coordination or communication

For the most part, real-time collection and dissemination of operational data are well coordinated. Much of these data are carried over the GTS. However, even for these data many problems have been identified, principally loss of data and a lack of real-time monitoring information. Shortcomings in real-time data flow are usually due to a shortage of resources or current capabilities and are not due to a breakdown in coordination.

Collection and transmission of delayed mode data, use of data for by multiple programmes, and ad hoc access to data and products are problematic. These functions have not been as well coordinated, there has been more duplication of effort, and incompatible standards have been developed or used.

It is evident that many of the problems in existing data management systems and capabilities are primarily due to an historic lack of resources and are thus unlikely to be remedied by any recommendations made in this report. However, some deficiencies are primarily the result of poor communication or insufficient coordination and there is reason to hope that these problems could be remedied. Specifically:

- There are a large number of different applications developed to serve the needs of different Programmes whose development has not been coordinated and there is little connectivity between applications resulting in inadequate/inflexible data exchange standards and incompatible data formats
- Inflexible or incompatible data transmission standards, procedures and protocols have been adopted

- Diverse and incomplete quality control of observational data and insufficient documentation of procedures for these data to be effectively used by other Programmes
- No standard practices for the collection, electronic archival and exchange of metadata, both high-level and detailed, especially for stations and instruments
- Lack of agreed standards needed to effectively identify, acquire and use all of the relevant data, particularly from archives, hampers multidisciplinary application of meteorological, hydrological and oceanographic data
- Poor coordination between observing systems, making it difficult to identify the potential of observation sites established by one Programme to meet the requirements of other Programmes.

RECOMMENDATIONS

Within WMO, each Programme has traditionally worked independently to develop and implement its own plans and activities. Programmes have worked together towards a common goal only where possible benefits were clear and immediate. An example of success in this area can be found in the cooperation between the WWW and Programmes that use the Basic Systems to generate and transport their data and information.

There are many reasons for the traditional independence of the Commissions and Programmes.

First and foremost, coordination between departments and between Technical Commissions has historically not been a high priority within WMO. It is clear from the Technical Regulations that it is assumed most activities are to be carried out within a Commission. The WMO budget is defined and administered by Programmes and the departmental structure of the Secretariat mirrors that arrangement. Creation of inter-commission groups thus introduces significant complications in management, budgeting and reporting responsibilities.

Second, many experts and staff see coordination as a cost with little benefit. That is, the effort required to coordinate activities with others can significantly slow progress and may require compromises and introduce complexity that would not be needed if the activity were narrowly focused on the needs of a single Programme. Since major activities within WMO are usually initiated in response to a pressing need, time spent on coordination is often seen as introducing unacceptable delay. This attitude also applies to standards, which are developed only when there is a compelling need and, again, usually within a single Commission.

Third, bureaucratic rivalries between Commissions and departments are an unfortunate reality. This does not imply that directors or staff make deliberate decisions not to cooperate. Rivalry instead manifests itself in the feeling that “we know best” or “we can do it better”. Thus, each Programme sees its expertise as superior and its requirements unique. In such an atmosphere, it is quite rare for one Programme to approach another Programme or department to ask for assistance.

Over the past several years it has become increasingly evident that many scientific and technical problems are inherently interdisciplinary in nature. It has been recognized for more than a decade that data and information management is an area where better coordination and cooperation could lead to increased efficiency and improved services. New crosscutting WMO Programmes (Space and Natural Disaster Prevention and Mitigation) presage increased requirements for well-coordinated information management. Several attempts have been made to address this issue. However, the organizational and bureaucratic obstacles have made this difficult.

Although progress has been made over the past several years, current mechanisms for coordinating data management activities within and between departments in the Secretariat and between the Technical Commissions could be further improved. The principal problems within the Secretariat are the result of poor communication, budget issues and the low priority accorded to coordination. Between Technical Commissions the problems generally result from poor communication, complications resulting from reporting responsibilities, inappropriate expertise at meetings, competing or incompatible requirements, and the perception that joint projects are slow and unwieldy. The following recommendations are suggested to address these problems.

Within the Secretariat:

a. Establish a data management coordination team within the Secretariat. Members of this team, comprising representatives of all of the scientific and technical departments, would be responsible for ensuring the activities within their own department are coordinated with similar activities in other departments. The team should also develop and oversee plans for inter-programme activities to accomplish common goals. It would be worthwhile for the team to meet once every few months to discuss ongoing activities and accomplishments. However, recognizing that conflicting schedules would make it difficult to hold regular meetings, once the team were established it might be sufficient to exchange information via an e-mail mailing list.

b. Budget support for inter-commission work must be made explicit. A formal mechanism for funding cross-programme activities in data management should be defined. The preferred option is for Executive Council to establish, approve and commit funds to a crosscutting programme in data management. There are several models that could be used to manage this programme. It could be managed by dedicated staff within an existing department, an independent office could be created under the Deputy Secretary General, as was done with the WMO Space Programme, or it could be matrix-managed by staff from several departments, perhaps comprising the coordination team recommended above. If it is not possible to create a formal crosscutting programme then each department should be directed to establish a line within its budget to support this work.

c. Improve feedback from experts to the Secretariat. Experts attending inter-commission task teams as representatives of their Programme or Commission often do not provide sufficient feedback of the results to the responsible staff in the Secretariat (or their Commission). These experts should be required to submit reports on their participation, which should include the results of the meeting, with emphasis on the requirements and concerns of their particular Programme and if or how they have been addressed.

d. Improve circulation of meeting reports. By improving the dissemination of the most pertinent results of meetings, communication of important activities could be improved at little cost. All meeting reports should include a concise (less than one page) summary of their most important results. Concrete results, such as proposed standards, are particularly important. These summaries should be distributed to the directors and chiefs of all of the scientific and technical departments. All reports should be accessible via the Web and links to these reports should be prominent and obvious from the front page of each department's Web site. It would probably be worthwhile to standardize the links to this information on each department's home page.

Within and between Technical Commissions

e. Many issues affecting more than one Commission can be resolved using the current mechanism of "weakly inter-commission" teams. That is, those teams that are established within a Commission but with invited representatives of other relevant Commissions. From this point onward within this summary these will be referred to as "inter-programme teams". To ensure the most effective experts are nominated, the attendance of the invited representatives should be funded by the Commission he/she represents. Since it is difficult to anticipate all opportunities for such inter-programme teams, every Programme should include some funds within its budget to support its participation in such efforts.

f. Some critical issues can best be addressed by “independent inter-commission” teams. In order to be most effective, these teams need to truly independent of any particular Commission. Membership of these teams should be based on the expertise needed with equal weight given to the requirements of all participating groups. Otherwise there is a perception that the views of one Commission (usually the designated lead) are accorded higher priority than the views of the others. If a crosscutting programme in data management were established (as suggested in recommendation b above), this programme would be responsible for the funding and management of such independent teams related to data management. If an umbrella programme is not established, then each inter-commission team should be directly funded by and report to Executive Council. Suggested mechanisms to manage such teams are discussed under recommendation b above. Since the participation in these teams will be supported by a dedicated budget, the team leader or responsible officer in the Secretariat should have a significant role in confirming team members.

g. Inter-agency, inter-organization coordination or data management activities can provide important benefits. Ocean data management, where JCOMM facilitates coordination with IOC, is a good example. Other Programmes within WMO that could benefit from further coordination with outside organizations include hydrology, agricultural meteorology and the World Climate and Weather Research Programmes. Officers responsible for these activities should reach out to other relevant organizations to investigate possible collaborative arrangements pertaining to data or information management.

h. Currently, FWIS is perceived as catering primarily to the requirements of CBS, while at the same time, moving too slowly to address pressing requirements of other Programmes. An independent inter-commission team, with representation from all of the players on an equal basis, should refocus further development and implementation of FWIS. The team should review the various components of FWIS and concentrate further efforts on a few specific and achievable objectives. Some users have suggested that highest priority should be given to developing an effective means to help users to find, acquire and use pertinent data from all WMO Programmes.

i. Continue to hold informal meetings of the Presidents of Technical Commissions. These meetings provide an effective forum for the presidents to exchange views and to keep abreast of significant developments and activities within the other Commissions. Reports of these meetings should be posted on the Web so that experts not in attendance might benefit from this same exchange of information.

j. Experts representing a Commission should be made explicitly aware of their responsibilities. Experts attending inter-programme or independent inter-commission task teams as representatives of their Programme or Commission should be reminded of their responsibilities in this capacity. Before attending the meeting they should be briefed of the Commission’s position on pertinent topics and should be given the authority to speak on the behalf of his/her president when necessary. Furthermore, effectiveness is enhanced when there is continuity in the experts chosen to attend follow-on meetings of the same group. These experts should submit reports on their participation to their president noting the requirements and concerns of their Commission and if or how they have been addressed by the meeting.

1. INTRODUCTION

It has been recognised for many years that effective management of data is essential for the success of scientific programmes. Data is the foundation on which our knowledge of the environment is built and data management is essential to ensure effective and efficient use of this resource. This is particularly true for international programmes, where data must be exchanged and understood regardless of language.

Most WMO Programmes have instituted various measures and activities to support their data management requirements. Although the aggregated requirements for each of the Programmes are unique, there are similarities and overlaps in many of their specific requirements. It is therefore likely that benefits could be gained from further coordination and integration of related data management activities among the Programmes.

Current WMO information systems have been developed to meet a diverse set of requirements. The principal system is the Global Telecommunication System (GTS) along with the related data processing and management functions that have been developed to serve the World Weather Watch (WWW). The GTS has a number of significant strengths: it is an operational private network that mainly provides for the exchange of real-time high-priority data, it is mature, well tested and operated according to well-defined procedures and shared responsibilities.

Other information systems have been developed to meet the needs of other Programmes and Commissions and each have their own advantages. Given the diversity of these systems it is difficult to provide a concise summary. However, most share a common strength: they have been developed by individual programmes to meet their own specific requirements. Thus, the systems are generally focused in their approach and do not entail compromises or inefficiencies that can sometimes result from development of generalised systems.

The current state of the ensemble of WMO information systems is extremely complex. Consider the data flow for a single programme JCOMM/IODE as illustrated in Figure 1. The figure highlights the complexity of the information management of a typical Programme. As a consequence it should be obvious that a complete and detailed data flow diagram for all WMO Programmes would be indecipherable.

Despite the complexity of the details, a review of the overall data management strategies for several Programmes reveals that most have several fundamental capabilities in common. The essential components of this fundamental data flow are illustrated in Figure 2. Nearly all Programmes collect data, transmit these data to one or more processing centres, perform quality control, generate products, transmit these products to users, and archive the data and products for future use. Several Programmes further divide the data flow into real-time and delayed mode data streams as shown.

Real-time and operational data flow is often essentially one-way, that is, products are routinely transmitted to collection centres or broadcast to users without any explicit action from the recipient. A variety of telecommunication services are used for real-time transmission including leased fixed circuits (both terrestrial and satellite) and the Internet. Delayed mode transmission utilizes the Internet, post and, less often, private circuits.

Access to data held in archives is usually provided on a two-way or request/reply basis. That is, users contact the archive, request products and these products are then transmitted to the user. Although the Internet is rarely used for real-time data transmission it is commonly used for access to archives.

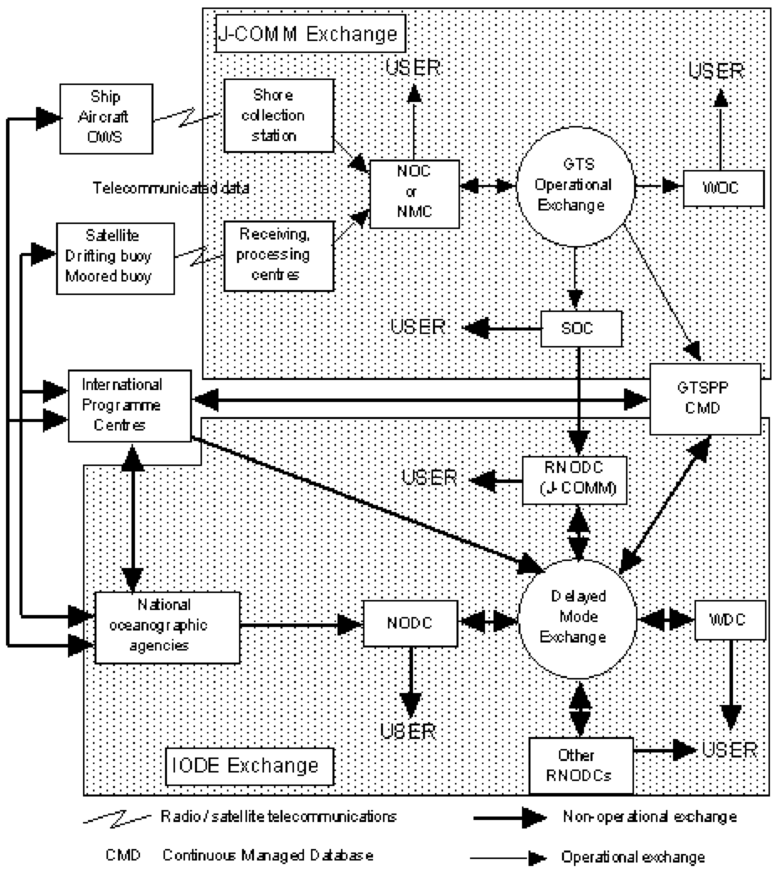


Figure 1. JCOMM/IODE Data Flow

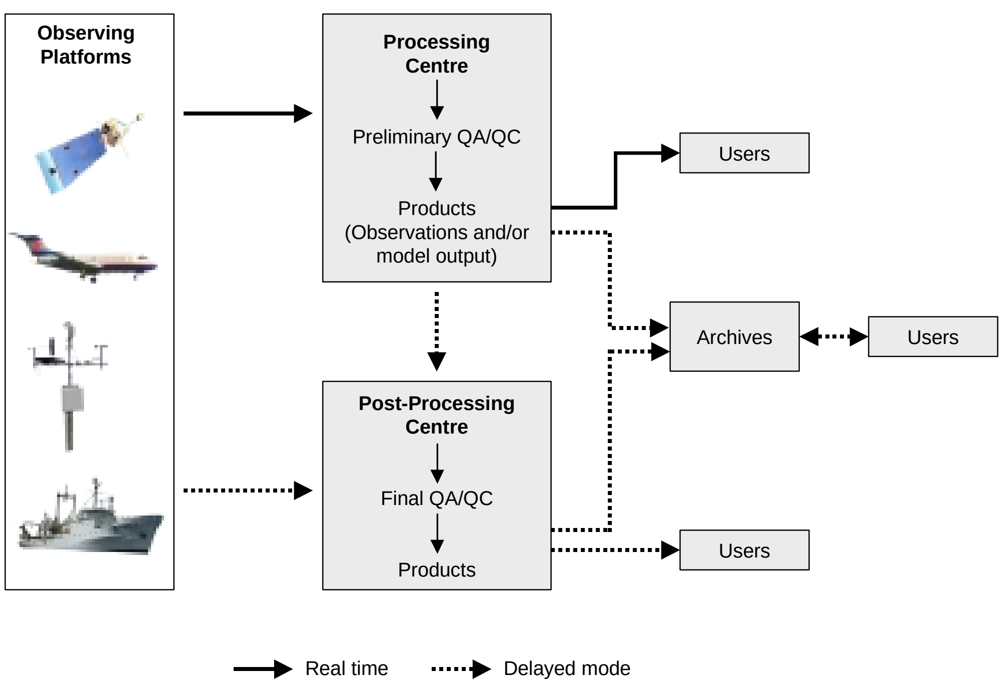


Figure 2. Simplified programme data flow

From the high-level perspective of Figure 2, it can be seen that many WMO programmes share some data management requirements in common. Several Programmes utilize the standard protocols and data representation forms of the WMO Basic Systems for transmission of their real-time and operational delayed-mode data. These include the WWW, Marine Meteorology and Oceanographic Activities, the Aeronautical Meteorology programme, and the World Climate Programme. However, other Programmes utilize processing procedures that are entirely unique. Furthermore, even those programmes that utilize the Basic Systems for operational needs have usually developed their own procedures to handle archives, research data, and ad hoc requests. For the most part these have been developed with little or no coordination between Programmes.

This lack of standardisation has led to a number of problems. Notably:

- There is little connectivity between applications developed to serve the needs of different Programmes
- There are a large number of different applications whose development has not been coordinated making integration of data sets technically challenging
- Multidisciplinary application of meteorological, hydrological and oceanographic data is hampered by lack of agreed standards needed to effectively identify, acquire and use all of the relevant data

The multiplicity of systems operated for different Programmes has resulted in incompatibilities, duplication of effort and higher overall costs for Members. Continuing to develop systems in this uncoordinated manner will exacerbate these problems and will further isolate the WMO Programmes from each other and from the wider environmental community. It will increase the difficulty in sharing information between programmes, which is growing increasingly important for inter-disciplinary research.

This problem has been recognised within WMO for many years and a number of efforts have been made to address it. Beginning in the mid 1990s a series of inter-programme data management coordination meetings were held. Representatives of all of the scientific and technical programmes of WMO, as well as representatives of other international programmes with significant data management components, such as the Committee on Earth Observation Satellites (CEOS), the International Geosphere Biosphere Programme Data and Information System (IGBP-DIS) and the International Council of Scientific Unions (ICSU) World Data Centres, were represented. These meetings provided a useful forum to exchange views and information on the activities of the various programmes and led to the effort to develop a comprehensive WMO information system, which could encompass and address the needs of all WMO Programmes.

WMO Congress, at its fourteenth session, agreed that an overarching approach to information management within WMO was required: a single coordinated global infrastructure, the Future WMO Information System (FWIS). FWIS would be used for the collection and sharing of information for all WMO and related international programmes. The FWIS vision provided a common roadmap to guide the evolution of the information system functions performed by current WMO Programmes into an integrated system that efficiently meets all of the requirements of Members for the relevant international environmental information. Congress emphasized that the implementation of FWIS should build upon the most successful components of existing WMO information systems in an evolutionary process. FWIS is further discussed in section 4.9.

2. EXISTING DATA MANAGEMENT STANDARDS, PRACTICES AND GUIDELINES

For this study, representatives of each of the scientific and technical programmes of WMO were interviewed to ascertain the data management standards, principles, practices and guidelines currently in place for their respective Programmes. For the most part the Programmes provide overall guidance on preferred practices for data management but do not specify binding standards. For Programmes that comprise sub-programmes, commonly each of the sub-programmes has instituted their own data management practices and procedures with little guidance or direction from the parent Programme.

A notable exception to this trend is the WWW, which has a long history of specifying comprehensive standards within the WMO Technical Regulations. The breadth and detail of data exchange standards and procedures specified for the WWW have certainly contributed to the success of the WWW over the past 40 years. Nonetheless, these regulations are perceived to be complex and inflexible by some of the other Programmes, which have consequently managed their own data management activities with more flexible and informal arrangements.

A concise summary of the guidelines, standards and regulations instituted by each WMO programme is provided in the text boxes below, followed by a brief discussion of each.

2.1. **Agricultural Meteorology programme** (Commission For Agricultural Meteorology)

> While there are published guidelines on collection and management of agrometeorological data within the Guide to Agricultural Meteorological Practices, the current Guide is more than twenty years old. At present, there are no defined standards for exchange of digital agrometeorological data or metadata.

The Guide to Agricultural Meteorological Practices (WMO-134) provides general guidance on best practices for operation of an agricultural meteorological service. Most of the Guide is more than 20 years old and is seriously out of date. An update is currently being prepared and is expected to be available within the next few years. Within the existing Guide there are references to WMO technical regulations and other guidance material. However, these references are limited and in many cases use obsolete references that are no longer valid. Most importantly for this study, the Guide to Agricultural Meteorological Practices contains no material on digital exchange formats or mechanisms for either data or metadata.

Training material for the “Roving Seminar on Data Management for Applications to Agriculture” provides guidance on data collection and dissemination, including basic information on types, uses and limitations of data of interest to Agriculture. There are extensive discussions on data quality assurance, analytical procedures and techniques, including the use of Geographic Information Systems (GIS). The document provides an introduction to the Internet, which explains how it can be used to retrieve information and gives examples of how to make information available to the public. The document also outlines strategies for data collection and dissemination at the national and regional level. The material does not describe specific data exchange or metadata standards.

2.2. **Aeronautical Meteorology programme** (Commission for Aeronautical Meteorology)

> Exchange standards for aeronautical meteorological data are well defined and there has been close cooperation with the WWW and CBS in defining data representation forms and transmission procedures. CBS data processing centres play a well-coordinated role in the World Area Forecast System.

The standards and guidelines for data relevant to aeronautical meteorology are specific and well defined in the WMO Technical Regulations Volume II - Meteorological Services for International Air Navigation. These standards are closely coordinated with and equivalent to CBS standards. Guidance on data management principles and recommended practices are described in the Guide on Meteorological Observation and Information Distribution Systems at Aerodromes (WMO-731) and the Aircraft Meteorological Data Relay Reference Manual (WMO-958).

2.3. **Atmospheric Research and Environment Programme** (Commission for Atmospheric Science)

> Within the GAW separate groups are responsible for establishing data management guidelines for each of six basic parameter types. Since each applies to a different set of parameters it has not been considered necessary to coordinate these procedures among the groups. Compliance with guidelines is voluntary.

Among the activities of AREP the Global Atmosphere Watch (GAW) has the most well defined data management component. The backbone of the GAW is a network of Global, Regional and Contributing monitoring stations. About 80 countries host GAW stations through either their National Meteorological or Hydrological Service (NMHS) or through collaboration with other national scientific organizations. The 22 Global GAW stations are situated near upper air stations of the WWW Global Observing System (GOS) located in remote locations. A database of network station information is maintained in the GAW Station Information System within the Secretariat. This database is not closely coordinated with the list of stations maintained in WMO Pub. 9, Volume A.

Scientific Advisory Groups (SAGs) have been established to organize and coordinate the monitoring for the six basic parameter types. Each SAG is responsible for preparing measurement guidelines, defining data quality objectives, and developing and overseeing implementation of standard operating procedures. Since each SAG is responsible for a different set of parameters, it has not been considered necessary to coordinate these procedures among the SAGs.

GAW World Calibration Centres and Quality Assurance/Scientific Advisory Centres help to ensure the data submitted to the five GAW World Data Centres (WDCs) are of high quality. The WDCs are operated and maintained by their individual host institutions and each focuses on selected atmospheric chemical constituents. As a consequence, each WDC has its own standards, capabilities and requirements. Observational data are transmitted to the WDCs through various mechanisms; primarily the Internet and special purpose regional networks. Each WDC has its own preferred format(s) for receiving data but these preferences are considered guidelines and compliance is voluntary. For all measurement groups, GAW is developing a common terminology on data quality objectives that is consistent with the Bureau of International Weights and Measures and applicable to all GAW measurements.

GAW data management guidelines and standard procedures are explained in several GAW reports and technical documents. Some representative documents include:

a. Report of the WMO-NOAA Expert Meeting on GAW Data Acquisition and Archiving (Asheville, NC, USA, 4-8 November 1995) (TD-755)
b. Global Atmosphere Watch Measurements Guide (TD-1073)
c. Strategy for the Implementation of the Global Atmosphere Watch Programme (2001-2007) (TD-1077)
d. Updated Guidelines for Atmospheric Trace Gas Data Management (April 2003) (WMO TD-1149)

For the most part, these documents provide an overview of best practices rather than specific standards. For example TD-1073 states where data are archived but does not specify how or in what form it is transmitted there. An excerpt from TD-1073 is given below. Emphasis on selected text for this report is indicated in bold italics.

> "Measurement programs maintain records of sample collection details (e.g., collection location, date), raw analysis data (e.g., chromatograms, voltages), processed analysis data (e.g., mixing ratios, isotopic ratios), instrument diagnostic data (e.g., temperature, pressure), and standard gas calibration histories. In many instances, these data are maintained for each trace gas constituent measured. Further, individual laboratories may operate both continuous and discrete (flask and cylinder) measurement strategies from both fixed (land surface, towers, and ice

cores) and moving (ship and aircraft) platforms. All of this information must be managed so that measurement data can be readily viewed, re-processed, selected, analyzed, and disseminated. ***A poor data management strategy can render even the very best measurements almost useless.***

Thus, it is imperative that measurement programs employ a well-designed data management strategy. The carbon cycle measurement community, and in particular, the CO^2 community, has spent considerable effort developing instrument and technical guidelines for scientists entering into the atmospheric trace gas measurement field [*WM*O, 1998]. Guidelines which emphasize measurement techniques, calibration methods, and common pitfalls, provide fledgling measurement programs with a "recipe" for obtaining high-precision measurement results more quickly. A similar effort is required for the management and maintenance of data produced by these programs. ***A first attempt to establish guidelines for atmospheric trace gas database management is presented. The data management strategy described is a compendium of concepts in use among many of the laboratories making trace gas measurements.*** The National Oceanic and Atmospheric Administration (NOAA) Climate Monitoring and Diagnostic Laboratory (CMDL) Carbon Cycle Greenhouse Gases (CCGG) Group has committed considerable resources towards the management of discrete and continuous data from the cooperative air sampling network, the CMDL baseline observatories, and the tall towers and light aircraft programs. Many of the examples presented are from CCGG programs that address many of the same data management considerations as other carbon cycle measurement laboratories. The majority of guidelines set forth here have been in use by CCGG for many years; however, the CCGG strategy has considerable room for improvement and continues to evolve as the measurement programs expand and new technologies become available. ***This discussion represents the direction in which the CMDL Carbon Cycle Greenhouse Gases Group is moving towards a robust database management system.***"

2.4. Global Climate Observing System

GCOS relies upon facilities provided by other WMO Programmes, principally the WWW and GAW, for its data and information. It relies upon CBS to define transmission protocols and data representation forms. General principles for management of climatological data have been adopted and are described in the Joint Data and Information Management Plan.

GCOS, GTOS and GOOS data management procedures and activities are defined in the Joint Data and Information Management Plan (GCOS-60). This plan is intended to complement the program plans for each of the three observing systems. It, therefore, provides general guidance, principles and best practices and does not define specific procedures, formats, standards or protocols. Data management plans for each of the three observing systems are expected to provide detailed information on the system-specific requirements and procedures. Additional, somewhat generic, information is provided in the GCOS Data and Information Management Plan (GCOS-13), which is now more than eight years old.

2.5. Hydrology and Water Resources Programme (Commission for Hydrology)

Several documents provide guidance on management of hydrological data. Since there is no routine international exchange of hydrological data there are currently no standard protocols or data representation forms defined.

Several documents provide guidance on management of hydrological data. These include the Guide to Hydrological Practices (WMO-168), Hydrological Data Transmission (WMO-559), HOMS Reference Manual, and Guidelines for Computerized Data Processing in Operational Hydrology and Land Water Management (WMO-634). These documents describe general best practices for collection and management of data and provide information and recommendations on the need to define standards. However, they do not define detailed standards themselves. An excerpt from the Guidelines for Computerized Data Processing in Operational Hydrology and Land Water Management illustrates the nature of the material.

> "The adoption of standard formats should not impose any constraints on the user's own data-processing activities, rather it should minimize the resources involved in the movement of hydrological data between data-processing centres. …. In the absence of exchange standards, each agency must develop separate software routines to handle the data formats used by each other agency. This leads to a proliferation of software development tasks. Further, it is often found that receiving agencies cannot even read the data because of some non-standard/unspecified feature of the recording media or the data set."

Currently, there is limited hydrological data and/or information transferred between WMO Members. There are no international exchange standards. In many cases there are not even standards for exchange between regions within a country. It is of course important to share data among countries within a river basin that spans national boundaries. Where these data are shared the cooperating parties have generally agreed on the arrangements among themselves. Although the HYDRA and HYFOR codes are defined in the Manual on Codes, they do not appear to be used for routine exchange.

The main hydrology related information systems in operation within WMO are largely hard copy based in the form of the guidance given in the Guide to Hydrological Practices and within Technical and Operational Hydrology Reports. Computer based information systems include INFOHYDRO (an Excel Spreadsheet) which provides information on the agencies that collect hydrological data and their networks and the Hydrological Operational Multipurpose System (HOMS) which provides access to technology, software, guidelines and manuals (some over the Internet). Runoff data are provided to the Global Runoff Data Centre from participating Members but there are no standards defining the format of these data.

The GTS has been proposed for use in the collection of hydrological data within the WHYCOS projects. However, currently alternative methods for product and data delivery are being used and are considered to be less expensive and more accessible than the GTS.

2.6. **Marine Meteorology and Oceanographic Activities** (Joint Commission for Oceanography and Marine Meteorology)

Exchange standards for marine meteorological data are well defined and there has been close cooperation with the WWW and CBS in defining data representation forms and transmission procedures. Some unique exchange formats are now being developed. There has been limited coordination of data management guidelines, with each sub-programme developing its own practices and procedures. Recent efforts to define metadata standards have acknowledged the WMO Core Metadata Standard but have not yet embraced it as a starting point.

Over its long history, the marine meteorology and oceanographic community has developed a wealth of material describing data management standards, practices and guidelines. The first international agreement concerning a formalized recording of weather observations from the seas in ships' logbooks was made way back in 1853. Traditional logbooks were only episodically digitized and exchanged through bilateral agreements. The digitalization and exchange of digitized meteorological journals were not formalized until 1960.

The standards and guidelines for management of these data are defined in a number of publications including the Guide to Marine Meteorological Services (WMO-471) and the Manual on Marine Meteorological Services (WMO-558). Exchange standards for marine meteorological data are well defined and, since there has been close cooperation with the WWW and CBS, most are published in the Manual on Codes (WMO-306). A notable exception concerns the International Maritime Meteorological Tape (IMMT). Most of the codes in the

IMMT format with the exception of those added for the VOSClim project are defined in the Manual on Codes since they follow the FM 13-X Ship code. Because CBS did not agree to expand the FM 13-X Ship code for the VOSClim project, the additional observed elements have been added to the IMMT-2 format version modified for VOSClim. Further divergence is possible since there is currently an effort to define a standard for exchange of marine data using XML.

The Joint Commission for Oceanography and Marine Meteorology (JCOMM) was established relatively recently and has only begun a process in which oceanography and marine meteorology would transition from the existing largely unconnected set of monitoring, data management and service activities to a fully coordinated and integrated system. It is hoped that the transition process would selectively broaden and modify the tasks of existing mechanisms towards an agreed new structure.

2.7. **World Climate Programme** (Commission for Climatology)

The Guide to Climatological Practices provides extensive guidance on best data management practices. The Guide has been coordinated with the WMO Technical Regulations and guidance material developed by other WMO Programmes. Standard data representation forms for exchange of some data (CLIMAT, CLIMAT TEMP) over the GTS are specified in the Manual on Codes. Guidelines on maintenance of metadata are provided but no standards for exchange of these data are defined.

The Guide to Climatological Practices (WMO-100) provides guidance on activities and best practices necessary to provide climatological services by Members. An updated version of the Guide has been available for review since 2001. It is nearing completion and should be published in 2004. The Guide contains a chapter dedicated to climate data management. This chapter contains general information on what data management encompasses, defines principles of effective data management and discusses methods and recommended practices for data collection, handling and processing. There is a comprehensive section on preferred procedures for quality control of climatological data.

The Guide has been well coordinated with the WMO Technical Regulations and guidance material developed by other WMO Programmes. For example, section 2.6.4 on Quality Control states

> "Detailed guidance on on-site quality control of observations and reports is given in the Manual on the Global Observing System, Part VI and Attachment VI.1 (WMO No.544) and in the Manual on the Global Data-processing System,Vol.1,Attachment II.I, Table 1 (WMO No. 485). The following procedures should be followed where there is an observer, or other competent personnel, on-site."

The section on data exchange (3.3.1) refers to both the GTS and Internet.

While the Guide provides a wealth of useful information, for the most part it describes general best practices rather than specifics. For example, it notes the need for documentation to describe delivery formats but does not provide a recommendation or specification for any standard exchange formats.

As a further example, consider the following excerpt from section 2.6.9 Station Documentation (Metadata).

> "The use of climatological data will be greatly facilitated if the climatological or other responsible section maintains complete documentation of all stations in the country. The information should be kept up to date and should be easily obtainable, in station catalogues, data inventories, and climatic data files, upon application to the climatological division. Beginning with the official name and any index number, station documentation should include such geographical information as latitude, longitude, elevation, the name and mailing address of the station or associated agency, and a description of the local topography, soil types and surrounding land use. Complete information regarding the observing programme, including lists of variables measured and hours of observation,

> should be recorded. Documentation should contain a complete history of each station including the date of beginning of records, changes of site, changes in exposure of each set of instrumentation including height above ground, dates of closure or interruption of records, names of observers and their periods of service. In view of the increasing importance of detecting climate change and variability, records should be kept of dates of changes in instrument types, of instrument replacements and calibrations, and of the results of calibrations and re-calibrations. Finally, station documentation should include a record of all inspection visits, especially comments about the site, exposure, quality of observations and station operations."

This material provides extremely useful and important information. However, in considering the requirements for inter-programme cooperation and integration, additional information, such as standards for exchange of metadata are needed.

As described in the Guide (3.3.5.1) "The Climate Data Information Referral Service, INFOCLIMA, is a service for the collection and dissemination of information on the existence and availability of worldwide climate data. The information comprises descriptions of available data sets that are held at data centres and/or published. Information is obtained from Member countries and also from contributions by individual data centres and international organisations. The referral system is international in scope. INFOCLIMA does not handle actual climate data, but only the metadata."

INFOCLIMA has attempted to provide a comprehensive reference to climate data sets. It currently includes information on more than 1250 data sets held by more than 300 centres in 125 countries. INFOCLIMA has proven to be difficult to keep up to date and it is now thought that a standard or mechanism to support cross-catalogue searches might be more effective.

While a formal format for provision of data set descriptions for inclusion in INFOCLIMA is not defined, examples are provided, such as the following:

#1143 (B001) : GLOBAL SYNOPTIC DATABASE CROSS-REFERENCE
: cryosphere
FORM : single set
MEDIA : digital
PERIOD : 1966 - current
AREA : global
ELEMENTS : precipitation ,air temperature ,air temperature (max/min)
wind, air pressure, humidity, weather, cloud (sky cover, type, height)
visibility, snow (depth, coverage), state of ground
ARRANGEMENT : per station or square
chronology/time series
ORIGIN OF DATA : in situ obs - manual, registered
CENTRE 629B (02/10/92): Deutscher Wetterdienst, Seewetteramt Hamburg - Germany

2.8. World Climate Research programme

> Data management within WCRP represents a microcosm of WMO as a whole. WCRP encompasses several sub-programmes and these in turn often comprise several projects. Each project is responsible for developing its own data management plans and procedures and there has been little coordination among these disparate activities. For some projects, data management is considered a high priority and comprehensive planning has been done. For others, it is of secondary importance and has consequently received less attention. Although there are some exceptions, for the most part each project has adopted its own standards and the WMO Basic System protocols and data representation forms are rarely used.

The World Climate Research Programme (WCRP) encompasses several sub-programmes including the Arctic Climate System Study (ACSYS), Climate Variability and Predictability (CLIVAR), the Global Energy and Water Cycle Experiment (GEWEX), the Stratospheric

Processes And their Role in Climate (SPARC) study, the World Ocean Circulation Experiment (WOCE), and WCRP modelling activities. Each of these programmes or experiments comprises several projects or activities.

For the most part, each individual project is responsible for developing its own data management plans and procedures. For some, comprehensive planning for data management has been done and standards and procedures have been agreed. For others no data management plans have been developed and coordination within the project has been managed informally.

Consequently, it is difficult to characterise the existing data management standards, practices and guidelines for the WCRP as a whole. Some projects make use of CBS data representation forms or other WMO standards where they are applicable. For example, the ACSYS dataset catalogue uses the new WMO Core Metadata profile as the basis for the structure of its database.

Some projects use formats developed in the research community such as the Network Common Data Form (NetCDF) or the Hierarchical Data Format (HDF). Others have developed specialized formats tailored to their own requirements.

Although it is apparent that data and products produced by the various programmes and projects within WCRP could be of considerable value to other projects, both within and beyond WCRP, no explicit mechanism for sharing this information has been defined. No standard dataset catalogue or equivalent metadata standard has been developed or adopted.

2.9. **World Weather Watch** (Commission for Basic Systems)

> Of all WMO Programmes the WWW has without doubt defined and instituted the most comprehensive and specific standards, practices and guidelines for transmission and management of its data and information. These are described in numerous guides and manuals and codified within the WMO Technical Regulations. Nonetheless, standards to support some emerging requirements, such as inter-disciplinary research, are only in the early stages of development or implementation.

The WWW has a long history of specifying comprehensive specific standards to control the transmission and management of data and information within its Programme and these standards are codified within the WMO Technical Regulations. Over the years these standards have become so extensive that most have been extracted from the Technical Regulation document and published instead within the Manual on the GTS (WMO-386), the Manual on Codes (WMO-305) and the Manual on the Global Data Processing System (GDPS) (WMO-485), which are themselves regulatory.

Extensive guidance on best data management practices and procedures for the WWW has been developed and published in several guides and technical reports. These include the Guide on the GDPS (WMO-305), Guide on Binary Codes (TD-611), Guide on the Automation of Data-processing Centres (WMO-636) and the Guide on World Weather Watch Data Management (WMO-788).

Twelfth WMO Congress specified that the Basic Systems should endeavour to serve the needs of all WMO Programmes. As a result, CBS has included representatives of the other Technical Commissions in several of its Open Programme Area Groups, expert teams and task teams. It has tried to reflect the requirements of other Programmes in its work and where necessary has broadened and expanded its data exchange standards and procedures to meet these expanded and emerging requirements. As an example, to support more effective access to data and products from different Programmes CBS has recently led an effort to develop a WMO metadata standard that could be used for development of a comprehensive catalogue (or catalogues) of WMO data sets. This standard is only in early stages of implementation and has not been widely publicized.

2.10. WMO Space Programme¨

> CGMS is responsible for defining data transmission formats and protocols for data and products from meteorological satellites. There has been close cooperation between the satellite operators and the WWW and standard code forms are used for relevant satellite products. However, satellite imagery and high-resolution products are too large to be transmitted over the GTS and are disseminated through alternate mechanisms. The responsible operators individually define formats and protocols for products from experimental satellites.

The Co-ordination Group for Meteorological Satellites (CGMS) provides a forum for the exchange of technical information on geostationary and polar orbiting meteorological satellite systems, such as reporting on current meteorological satellite status and future plans, telecommunications matters, operations, inter-calibration of sensors, processing algorithms, products and their validation, data transmission formats and future data transmission standards. The definitions of products, formats, schedules etc. for satellite broadcast products are published in Information on Meteorological and other Environmental Satellites (WMO-411). Guidance and information on methodology for processing or utilising satellite data and products is provided in Satellite Data Archiving, SAT-14, (WMO/TD-909) and Application and Presentation Layer Specifications for the LRIT/LRPT/HRIT/HRPT Data Format, SAT-19.

CEOS coordinates the CEOS International Directory Network (CEOS IDN), which is an international effort to assist researchers in locating information on available data sets. The CEOS IDN is sponsored by the CEOS Working Group on Information Systems and Services. The CEOS IDN provides free, on-line access to information on worldwide scientific data in the Earth sciences: geoscience, hydrospheric, biospheric, satellite remote sensing, and atmospheric sciences. The CEOS IDN describes data held by university departments, government agencies, and other organizations. The directory contents are shared among the nodes using the Directory Interchange Format (DIF), which is further described in section 2.11.

Historically there has been close cooperation between the satellite operators and the WWW and standard code forms are used for relevant satellite products. However, satellite imagery and high-resolution products have usually been too large to transmit over the GTS and are transmitted between Members via leased lines or the Internet. Alternative dissemination methods (ADM) for satellite data and products are now being implemented. Additional information on ADM is provided in section 4.10.

2.11. Other programmes

Research organizations, universities and government agencies other than NMHSs that are engaged in environmental issues have defined several standard data formats for meteorological and related data. A few of these standards are widely used and worth noting.

The Network Common Data Form (NetCDF), developed by the UCAR Unidata Program defines an interface for array-oriented data access and includes a library of software that provides an implementation of the interface. The NetCDF library also defines a machine-independent format for representing scientific data. Together, the interface, library, and format support the creation, access, and sharing of scientific data. The software is available free of charge. NetCDF is commonly used within universities and research organizations for exchange of meteorological data and products, particularly in North America.

The Hierarchical Data Format (HDF) was developed by the U.S. National Center for Supercomputing Applications (NCSA). HDF provides software support and file formats for scientific data management, particularly for array or matrix datasets. The HDF software includes I/O libraries and tools for analysing, visualizing, and transforming scientific data. The software is supported by NCSA and is freely available. Software is available to convert between NetCDF and HDF files.

There are a number of standards that pertain to metadata, particularly high-level metadata that define the contents and context of entire datasets. The International Standards Organization (ISO) has defined the Standard for Geographic Metadata for general-purpose metadata, in the field of geographic information. This standard, ISO-19115, while exceedingly broad, comprehensive and thus complex, provides a solid foundation for defining a standard for exchange of metadata relevant to any datasets that contain data referenced to earth coordinates, i.e. geospatial data. ISO-19115 was approved as an international standard only in 2002 and it is not yet clear how widely it will be accepted or adopted.

The FGDC Content standard developed by the Federal Geographic Data Committee (USA) has been adopted for digital geospatial metadata since the mid 1990s. Use of this standard was mandated for government agencies within the USA for describing their geospatial datasets and has, therefore, been widely used. Several agencies have developed filters to export the contents of the dataset catalogues in this format.

The Directory Interchange Format (DIF) has been adopted by CEOS as its standard for exchange of metadata within its international network of interoperable dataset directories. DIF was developed by the NASA Global Change Master Directory in the late 1980s and has evolved over the years to support fields required by other standards, such as the FGDC standard. Support for ISO-19115 has recently been added.

The Open GIS Consortium (OGC) is an international industry consortium of more than 250 companies, government agencies and universities participating in a consensus process to develop publicly available specifications for exchange and processing of geospatial data. Open interfaces and protocols defined by OpenGIS specifications support interoperable solutions that “geo-enable” Web and mainstream information services. They are intended to empower technology developers to make complex spatial information and services accessible and useful within all kinds of applications. OpenGIS Implementation Specifications, available over the Internet, detail the agreed upon interfaces that OGC develops through its consensus process. These are software engineering specifications that allow software developers to build products that implement one or more of the OpenGIS specifications. The software should then be able to communicate with other software that implements the specifications, allowing different layers for a single map to be collected from diverse providers over the Internet and displayed as a single product by a standard Web browser.

3. ESTABLISHED DATA MANAGEMENT REQUIREMENTS

To determine the information systems requirements of the WMO scientific and technical Programmes in light of current and near-term technological capabilities, representatives of many of these Programmes were consulted. Despite efforts to specify and quantify these requirements in detail, all agreed that it was extremely difficult, if not impossible, to specify volume and timeliness requirements for the diverse set of information needed to meet the requirements of all of the Programmes.

Only a few Programmes have defined and documented most of their data management requirements. For the majority of Programmes, some requirements have been specified, usually for real-time data, while other requirements are only vaguely known. For these less specific requirements, communication links are established that provide maximum capacity within available financial resources and data and product flow are prioritised to ensure the most pressing requirements are met. Other information is transmitted up to the capacity of the “pipe” available. In several cases postal services are used to deliver non time-critical data.

For the GAW, requirements for collection and transmission of data are well established. Volumes are very low and, in general, data are not needed in real-time so timeliness is not critical. However, some air chemistry parameters are becoming more useful for real-time purposes. This could provide an opportunity for enhanced cooperation with the WWW Programme, with which there has traditionally been little or no connection.

Within the WCP the requirement to collect CLIMAT and CLIMAT TEMP reports has been well defined for many years. Nonetheless, the current procedures and protocols within the WWW for transmitting these data from observing stations to processing and archival centres are clearly inadequate as an unacceptable percentage of reports fail to reach their intended destination. This problem has proven to be so intractable that efforts are now underway to generate equivalent reports from SYNOP and TEMP reports, even though they would provide an imperfect substitute.

Within the Marine, Aeronautical and WWW Programmes timeliness and volume requirements for their real-time data and products are well defined. For the most part, collection and delivery of these routine data and products are adequately addressed by existing arrangements, including the GTS the Internet and various bilateral agreements. Innovative arrangements are currently being investigated to meet growing requirements to transmit full resolution model and satellite products, both routinely and on an ad hoc basis.

By considering the requirements that have been articulated by one or more of the WMO Programmes or projects, the following emerging requirements can be identified.

- A widely available and electronic (on-line) catalogue of all meteorological and related data for exchange to support WMO Programmes is required.
- It should be possible to rapidly integrate real-time and non-real-time (archive) data sets to better interpret weather events in a climatological context.
- There is a need to identify the potential of observation sites established by one Programme to meet the requirements of other Programmes.
- Need to harmonize data formats, transmission standards, archiving and distribution mechanisms to better support inter-disciplinary use of data and products.
- Require a standard method for station numbering beyond the existing WMO numbers, which is not adequate to define all GAW, climate, hydrological or agromet stations.
- Need standard practices for the collection, electronic archival and exchange of metadata, both high-level and detailed, especially for stations and instruments.

4. CURRENT AND PLANNED DATA MANAGEMENT ACTIVITIES OR INITIATIVES

4.1. Agricultural Meteorology programme

The Agricultural Meteorology programme has a number of ongoing activities that pertain to data management. It conducts roving seminars on management of agro-meteorological data to disseminate new techniques and methods. It has prepared material on best practices and in 2000 developed a CD-ROM containing software for agroclimatic data management.

CAgM has appointed a representative to the Inter-programme Task Team on FWIS. This expert has been an active participant and has made significant contributions to the development of the FWIS concept.

The CAgM has established an Expert Team on Database Management, Validation and Application of Models, and Research Methods at the Eco-Regional Level. Among other responsibilities the team is expected to:

(a) recommend efficient ways and means of database management, including computer technology, standardized analytical techniques and integrated information management systems;
(b) determine and specify the needs of agriculture, rangelands, forestry and fisheries in future agrometeorological information management systems;

4.2. Aeronautical Meteorology programme

The Aeronautical programme depends upon the WWW for most of its data management activities. However, it does direct a few projects with significant data management components.

While the Aeronautical programme works to foster close contacts and cooperation with the aviation community to explore opportunities to provide enhanced services, it is also investigating innovative delivery mechanisms for forecasts and data. As an example, in the implementation of the World Area Forecast System (WAFS), the Aeronautical programme is responsible for ensuring the delivery of WAFS products to Members. These products are now being delivered operationally through a dedicated satellite broadcast. They are also available via the Internet. There has been some utilization of spare capacity within this satellite broadcast to deliver WWW data and products.

Automated meteorological observations from aircraft contribute to the GOS, with the number of reports increasing dramatically over the past several years. These data are collected via a private telecommunication system, which is already used by the aviation industry to transmit information to and from commercial aircraft. Global coordination and data management have been supported to ensure optimum cost-effectiveness.

4.3. Atmospheric Research and Environment Programme

The GAW comprises three missions:

1. Systematic monitoring of atmospheric chemical composition and related physical parameters on a global to regional scale
2. Analysis and Assessment in support of environmental conventions and future policy development
3. Development of a predictive capability for future atmospheric states

The monitoring component undertakes the most significant data management activities. Figure 3 shows the major activities and facilities comprising this part of the GAW programme. Support for these is provided, in large part, by WMO member countries that participate in the GAW. A network of measurement stations is the backbone of GAW monitoring. It consists of GAW Global and Regional stations with additional observations made at Contributing stations. Approximately 80 countries host GAW Global and Regional stations through either their NHMS or through collaboration with other national scientific organizations.

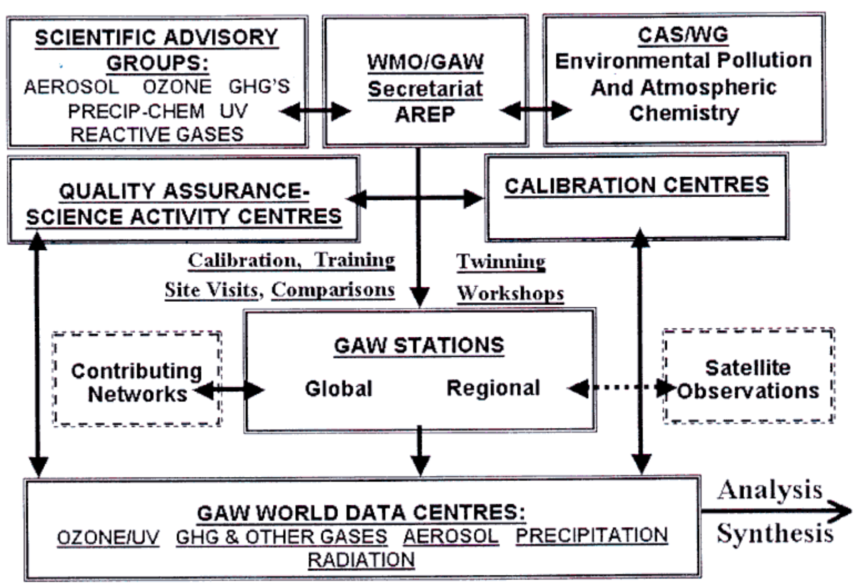


Figure 3. GAW monitoring components

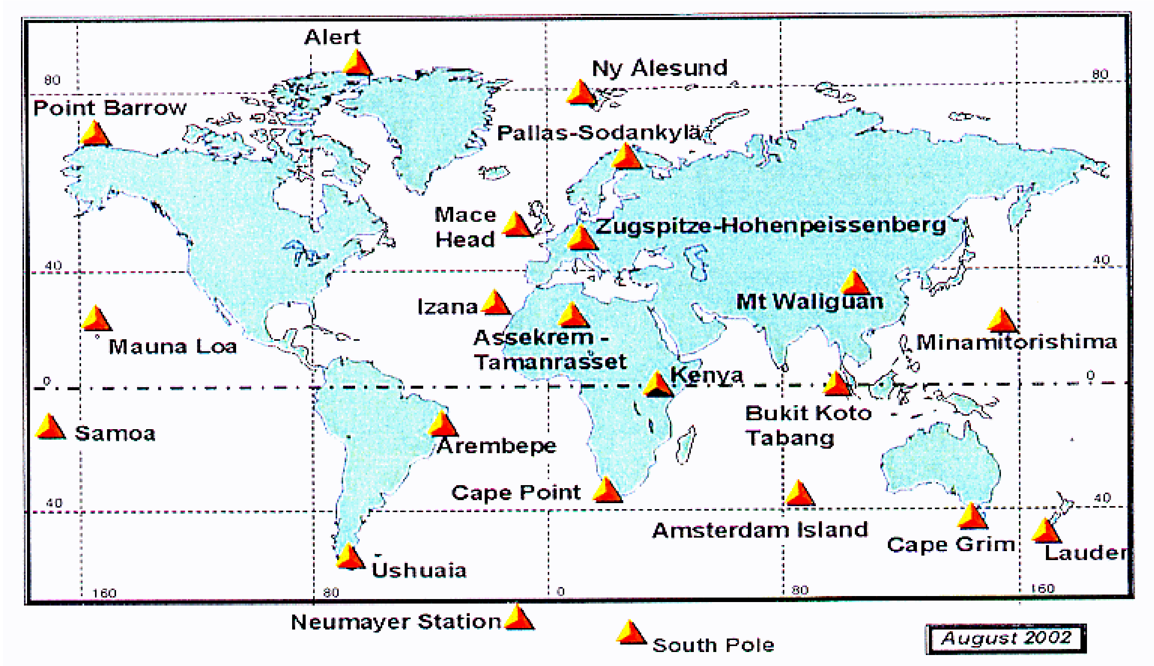


Figure 4. GAW Global atmospheric chemistry observatories

The GAW World Data Centres (WDCs) perform a major role in the management of GAW data. The WDCs are operated and maintained by their individual host institutions and focus on selected atmospheric chemical constituents. They collect, document and archive atmospheric measurements, the associated metadata from the measurement stations, and quality assurance information. They make the data freely available to the scientific community, usually via the Internet. In some cases, WDCs also provide additional products including data analyses, maps of data distributions, and data summaries:

1. The World Ozone and Ultraviolet Radiation Data Centre WOUDC (Toronto, Canada) is operated by the Meteorological Service of Canada (MSC), began as the World Ozone Data Centre in 1960 and in June 1993 began receiving data on ultraviolet radiation. At present, there are over 400 registered stations represented in the archive.
2. The World Data Centre for Greenhouse Gases WDCGG (Tokyo, Japan) was established at the Japan Meteorological Agency (JMA) in October 1990. It collects and distributes data on the mixing ratios of greenhouse (CO2, CH4, CFCs, N2O, O3 etc.), other related reactive gases (CO, Nox, SO2, VOC, etc.) and associated meteorological parameters. As of February 2000, 182 stations in 42 countries had submitted observational data for 13 species of greenhouse and related gases.
3. The World Data Centre for Precipitation Chemistry WDCPC (Albany, USA) is operated by the Atmospheric Sciences Research Center of the State University of New York at Albany. Data includes precipitation acidity or alkalinity, the major cations ammonium, calcium, potassium, magnesium, sodium and the major anions: sulphate, nitrate, chloride.
4. The World Data Centre for Aerosols WDCA (Ispra, Italy) is operated by the European Union's Joint Research Centre and was set up to archive aerosol related observations made under GAW.
5. The World Radiation Data Centre WRDC (St. Petersburg, Russian Federation) was established in 1964 at the Main Geophysical Observatory of the Russian Federal Service for Hydrometeorology and Environmental Monitoring. It is the central repository of global, diffuse and direct solar radiation, downward atmospheric radiation, net total and terrestrial surface radiation (upward), spectral radiation components (instantaneous fluxes), and sunshine duration, on hourly, daily or monthly basis.

TABLE 1. A summary of the facilities responsible for components of the GAW calibration and quality assurance system as of May 2003. The facilities have assumed global responsibilities, unless indicated (Am: Americas; E/Af: Europe and Africa; As/O: Asia and the South-West Pacific). It should be emphasized that several major network operators contribute substantially to quality assurance of global atmospheric composition measurements in addition to those facilities listed below. (! Mark represents missing components that the WMO/GAW secretariat with members and partners are close to filling)

SPECIES	QA/SAC	WORLD CALIBRATION CENTRE	REFERENCE STANDARD HOST	WORLD DATA CENTRE
CO_2	JMA *(AsIO)*	CMDL	CMDL	JMA
CH_4,	EMPA(Am, E/Af), JMA A/O	EMPA(Am, E/Af), JMA A/O	!	JMA
N_2O	UBA	IMK-IFU	!	JMA
CFCs	!			JMA
Total Ozone	JMA (As/O)	CMDL[1], MSC[2], MGO[3]	CMDL[1], MSC[2]	MSC
Ozone Sondes	FZ-Julich	FZ-Julich	FZ-Julich	MSC
Surface Ozone	EMPA	EMPA	NIST	JMA
Precipitation Chemistry	ASRC	ASRC	ISWS	ASRC
CO	EMPA	EMPA	CMDL	JMA
VOC	UBA	IMK-IFU	NCAR	.IMA
SO_2				JMA
NO_5				JMA
Aerosol Phys. Characteristics	UBA	WCCAP		JRC
Optical Depth		WORCC	WORCC	JRC
UV Radiation	ASRC-SUNY (Am)	SRRB (Am)		MSC
Solar Radiation		PMOD/WRC	PMOD/WRC	MGO
85Kr, 222Rn		EML		JMA
7Be, 2lOPb		EML		EML

- ASRC- Atmospheric Science Research Centre Albany, NY (US NOAA GCOS);
- CMDL, Climate and Monitoring Diagnostics Laboratory, NOAA, USA;
- EML, Environmental Measurements Laboratory, OHS, USA.
- EMPA, Zurich, Swiss GAW;
- FZ-iuelich, Forachungs Zentrum Juelich, Germany;
- IMK-IFU, Institute fur Umwelt Forschung, Garmisch-Partenkirchen, Germany (UBA supported);
- ISWS, Illinois State Water Survey, USA
- JRC European Joint Research Centre, Ispra, Italy;
- JMA Japan Meteorological Agency GAW,
- MSC Meteorological Service, Research Directorate, Canada;
- MGO Main Geophysical Observatory, St. Petersburg, Russia;
- NIST, US National Institute for Standards and Technology;
- NCAR National Center for Atmospheric Research USA;
- SRRB Surface Radiation Research Branch of NOAA ARL, USA;
- UBA Federal Environmental Agency, Germany;
- WCCAP: World Calibration Centre for Physical Aerosol Properties, Leibniz Institute for Tropospheric Research, Leipzig, Germany
- WORCC: World Optical Depth Research and Calibration Centre and WRC World Radiation Centre,,Swiss GAW, Physikalisch-Meteorologisches Observatorium, Davos, Switzerland.

I. Dobson instrument
2. Brewer Instrument
3. Russian filter instrument

4.4. Global Climate Observing System

GCOS relies upon other Programmes to conduct most of its operational activities, including data management. However, some unique tasks have been identified and these are being undertaken by specialized centres, which also participate as centres within other WMO Programmes.

GCOS Monitoring Centres

GCOS, CBS and CCl collaborated to define two basic networks of stations to comprise the GCOS Surface Network (GSN) and the GCOS Upper Air Network (GUAN). GCOS has recognized the need to monitor the performance of these networks and established the GCOS Monitoring Centres. Germany operates the GSN Monitoring Centre with special regard to precipitation while Japan monitors temperature data. ECMWF and the Hadley Centre monitor GUAN data. These centres:

- Monitor the availability, timeliness and completeness of the incoming data and messages received via GTS or other communication medium with the objective of improving the performance of the network being monitored;
- Perform fundamental quality control and assurance procedures on the incoming data and metadata to ensure the basic quality and completeness of the data set;
- Make basic quality-controlled data available to NMHSs, WDCs and others for their use in a variety of climate applications and products.

GCOS Analysis Centres

The GCOS Analysis Centre provides higher-level quality control of both the daily and monthly GCOS network data. NCDC performs this role for the GSN and NCDC and the Hadley Centre covers the GUAN. For the daily data, this includes updating and quality controlling the data, applying bias corrections, calculating monthly statistics from daily data and providing daily and derived monthly data, metadata and products to users. For the monthly data, this includes analyzing the data; improving bias adjustments and the monthly station database; creating global and regional monthly statistics; and developing and providing gridded products with reduced biases. The centre also reports on historical data and metadata reception.

GCOS Archive

WDC-A Asheville (NCDC) serves as the GCOS Archive for the GSN and GUAN.

- It will archive both the monthly and the daily data (in delayed-mode), as well as historical data, including metadata, for each station.
- It will make all GCOS data and products available to all potential users on a free and unrestricted basis.
- Data in the WDC may come either from data available at WDCs (e.g., from the Global Historical Climatology Network); from quality-controlled data available at the Monitoring Centres; from data submitted, upon request, by national centres (e.g., NMHSs) and available digitally and updated on a routine basis, or from any other source openly available to the archive for unrestricted further distribution.

CBS Lead Centres for GCOS Data

CBS Lead Centres for GCOS Data have been established, on a trial basis. They will facilitate the exchange of monitoring results with station operators in order that remedial action could be taken in a timely manner. These Lead Centres:

- Evaluate the monitoring results of the GCOS Monitoring and Analysis Centres;
- Coordinate activities with other GCOS Centres and/or other centres as appropriate;
- Liase with nominated Points of Contact for GCOS data to improve data availability and quality;
- Monitor and report to CBS and GCOS on action taken and progress achieved;
- Maintain the list of Points of Contact in co-operation with WMO Secretariat.

GOSIC

A single entry point for users enhances the distributed nature of GCOS data and information systems. The Global Observing Systems Information Centre (GOSIC) serves as a central source for information about the GCOS data system. It offers a search capability for GCOS data centres, to facilitate access to a worldwide set of observations and derived products. GOSIC does not hold data. Rather, it maintains metadata about the data sets that are available in the three programs and points to the data centres for the data and information. This metadata is managed by the NASA Global Change Master Directory (which is a node of the CEOS IDN) and is maintained in the DIF format.

4.5. Hydrology and Water Resources Programme

<u>WHYCOS</u>

The World Hydrological Cycle Observing System (WHYCOS) has been developed to assist Members in maintaining and improving their hydrological information systems. Modelled on the WWW, and using the same information and telecommunications technology, WHYCOS plans to provide a vehicle not only for disseminating high quality information, but also for promoting international collaboration. It should also provide a means for the international community to more accurately monitor water resources at the global level, and to understand the global hydrological cycle. WHYCOS has two components:

- a support component, which strengthens cooperative links among participating countries;
- an operational component, which achieves “on the ground” implementation at regional and international river basin levels.

WHYCOS is based on a global network of reference stations, which transmit hydrological and meteorological data in near real-time, via satellites, to National Hydrological Services (NHSs) and regional centres. These data enable the provision of constantly updated national and regionally distributed databases, of consistently high quality. WHYCOS aims to support, in all parts of the world, the establishment and enhancement of information systems that can supply reliable water-related data to resource planners, decision makers, scientists and the general public.

WHYCOS does not replace existing hydrological observing programmes, but supplements them. Composed of regional Hydrological Cycle Observing Systems (HYCOSs) implemented by cooperating nations, WHYCOS aims to complement national efforts to provide the information required for wise water resource management. In its initial phase WHYCOS has focused on establishing components in international river basins. The first steps in implementing WHYCOS have been made through regional HYCOSs in the Mediterranean, Southern Africa, and West-Central Africa. These have been made possible by collaboration between WMO, the World Bank, the European Union, and the Government of France.

The individual regional HYCOS components primarily address the priorities expressed by the participating countries and thus serve the needs of National Hydrological Services within a specific region. Therefore, the established hydrological information systems utilize different technologies, using either the GTS or the Internet. In the future HWR plans to develop integrated and standardised information concepts to serve all HYCOS components and to link with and participate in the Future WMO Information System.

<u>HOMS</u>

The Hydrological Operational Multipurpose System (HOMS) has been established to transfer technology in operational hydrology. This technology is usually in the form of descriptions of hydrological instruments, technical manuals or computer programs, which have been made available for inclusion in HOMS by the National Hydrological Services from the techniques that they use in their own operations. This is an important aspect of the HOMS philosophy in that it ensures that the technology transferred is not only ready for use but also works reliably.

HOMS is organized as a cooperative effort of WMO Members. Participating countries designate a HOMS National Reference Centre (HNRC), usually in the National Hydrological Service. This centre provides national components for use in HOMS, handles national requests for HOMS components to be supplied by other HNRCs, advises users on HOMS, and generally coordinates and publicizes HOMS activities in the country.

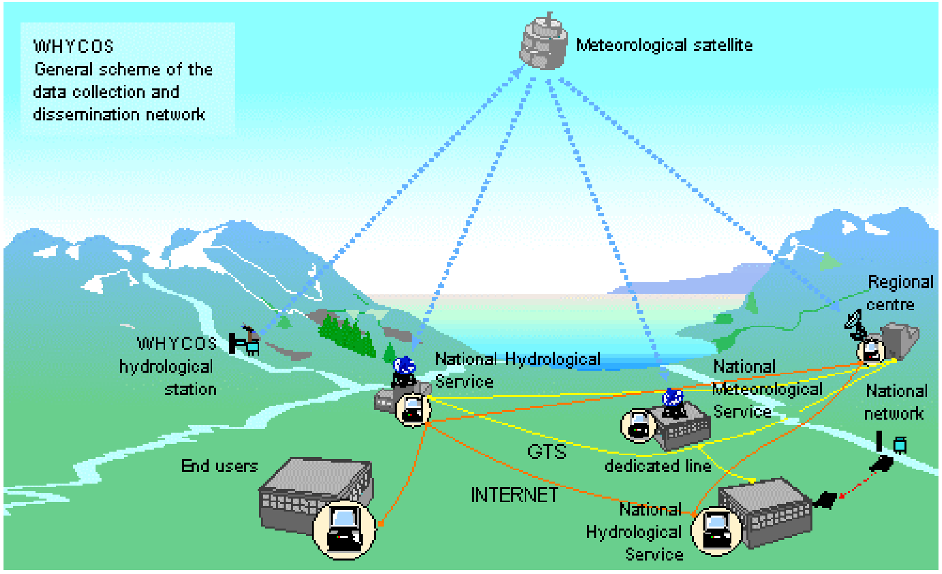


Figure 5. WHYCOS data flow

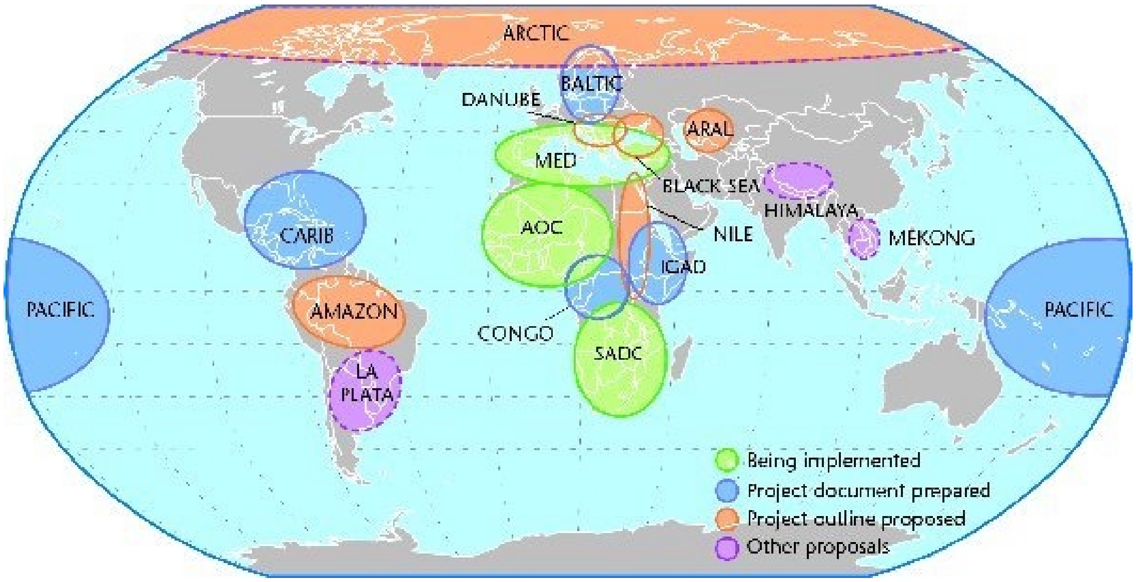


Figure 6. Status of Regional Hydrological Cycle Observing Systems

Being implemented	Project document prepared	Outline prepared	Being considered
MED-HYCOS SADC-HYCOS AOC-HYCOS	IGAD-HYCOS Congo-HYCOS CARIB-HYCOS Baltic-HYCOS Aral-HYCOS Pacific-HYCOS	Black Sea-HYCOS Danube-HYCOS Amazon-HYCOS Nile-HYCOS Arctic-HYCOS	La Plata-HYCOS Himalayan-HYCOS Mekong-HYCOS

GTN-H

The concept of a Global Terrestrial Network for Hydrology (GTN-H), established in 2001, is the result of the joint efforts of the HWR Department, GCOS and the Global Terrestrial Observing System (GTOS). The GTN-H is a global hydrological "network of networks" for climate that is building on existing networks and data centres and producing value-added products through enhanced communications and shared development. The goal of the GTN-H is to meet the needs of the international science community for hydrological data and information to address global and regional climate, water resources and environmental issues, including improved climate and weather prediction; detection and quantification of climate change; assessment of impacts of climate change; assessment of freshwater sustainability; and understanding the global water cycle.

The core functions of the GTN-H are to address the following:

- Provision of timely access to global hydrological data and metadata for users;
- Generation of relevant products and related documentation, satisfying timeliness and quality requirements of users;
- Promotion of standardization in observations and the use of 'best practices';
- Promotion and facilitation of free and unrestricted exchange of data and products within existing frameworks, e.g., WMO resolutions 40 (Cg-XII) and 25 (Cg-XIII);
- Soliciting of user feedback and measures to ensure responsiveness to changing needs;
- Monitoring and evaluation of GTN-H performance;
- Identification of key observational requirements of GTN-H, including requirements for satellite observations; and
- Provision for capacity building.

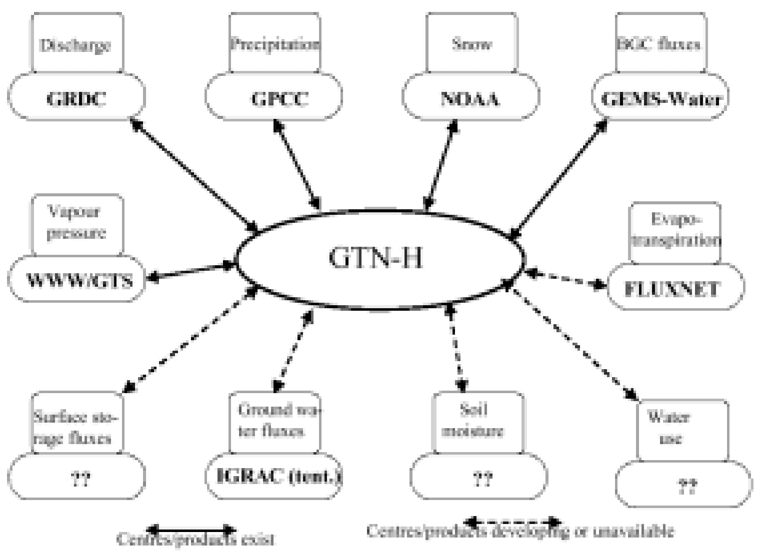


Figure 7. Initial configuration of the GTN-H

GTH-H has considered approaches for data collection, management, quality control and access and has determined that technical challenges related to these issues can be met by employing current and emerging technology and standards, best practices and available infrastructure. In particular:

- Metadata is a very important component, not only to describe the data, but to provide a useful contribution to the data quality assurance procedures;
- Communications systems and protocols and data standardized data formats are essential for transmitting and sharing data (e.g., GTS, File Transfer Protocol (FTP), Extensible Markup Language (XML));

- Web services are becoming increasingly available to facilitate access to data from distributed and disparate sources.

Several specific demonstration projects have been undertaken as part of the GTN-H development. These include:

Development of GTN-H Website: To develop a public website to help GTN-H users to discover and access GTN-H data and information products, and to provide linkages to GTN-H partners.

Inventory of Existing Data Products, Databases, and Organizations: Compile a general inventory of existing data products and databases, data sources, organizations, and other information of relevance to the GTN-H and its users. The inventory will comprise a "Discovery-level" meta-database (using the Dublin Core format) that describes GTN-H products, databases, and organizations. This will be integrated into the GTN-H website.

Demonstration of Metadata: Propose standardized detailed metadata formats for selected GTN-H data types and demonstrate their use in enabling the user to discover and access data and related information. Develop a mapping application that presents the user with a selected number of points representing a range of hydrological variables; for each point, the metadata record can be displayed, including a link to the website of the data provider. Using the FGDC metadata standard, propose a standardized metadata format for selected data types:

Map Product on Real-time Hydrological Conditions: Develop a pilot web application that demonstrates the retrieval, integration and presentation of real-time hydrometric data for selected large rivers from several countries.

4.6. Marine Meteorology and Oceanographic Activities

There are many observing instruments and cooperative arrangements contributing oceanographic and marine meteorological data to the GOS and Global Ocean Observing System (GOOS). Data flow is therefore quite complex, with a number of alternative communication systems used to transmit data from the various observing systems.

Argo is a global array of 3,000 free-drifting profiling floats that measure the temperature and salinity of the upper 2000 m of the ocean. This allows continuous monitoring of the climate state of the ocean, with all data being relayed and made publicly available within hours after collection. Deployments began in 2000. The under water floats, bob up to the surface every 10 days, transmit their data in real time via Service Argos. After elementary quality control, the data are processed into GTS messages in BATHY code and distributed. Argo data can also be retrieved on an ad hoc basis from the Global Data Assembly Centres: the Coriolis Project (France) and the Global Ocean Data Assimilation Experiment Server (USA).

The Automated Shipboard Aerological Program (ASAP) is built around upper-air observation stations that are operated on board cargo ships. Vessels enlisted in the ASAP program typically make two launches per day, although some perform up to four. On board equipment processes the sounding data into TEMP or SHIP TEMP messages. They are then relayed by satellite (Inmarsat), and injected into the GTS for international dissemination.

Within the Ship of Opportunity Programme expendable bathythermographs are deployed from recruited ships. The data are collected on board and transmitted by satellite (GOES; Meteosat, Inmarsat, Argos) to the responsible NMHS. After quality control they are formatted into BATHY, TESAC messages and switched onto the GTS. Delayed mode data are retrieved after the vessel returns to port, where operating agencies apply detailed quality control and transmit the data to a Regional National Oceanographic Data Centre, which then relay the data to a World Data Centre every 12 months.

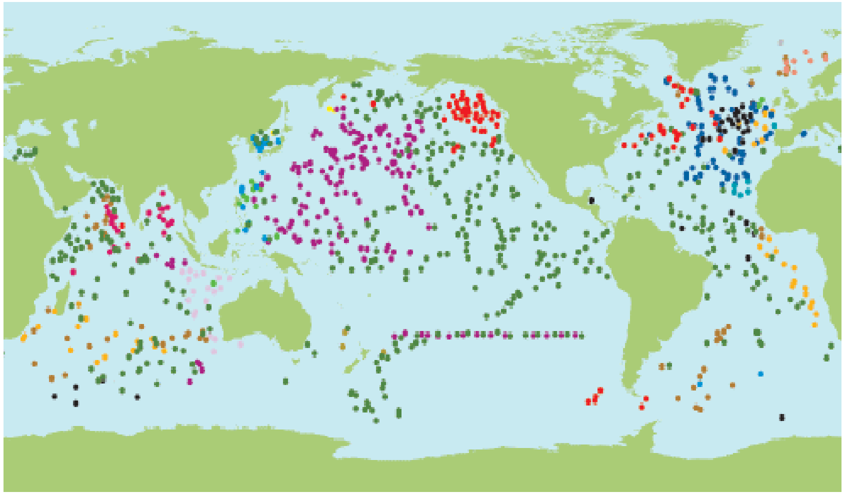


Figure 8. Argo floats as of October 20, 2003

The Data Buoy Coordination Panel is responsible for the international coordination of ocean data buoy programmes. The data are transmitted by satellite via Service Argos, which quality control the data, prepare BATHY or TESAC messages and relay the bulletins to MeteoFrance and NOAA/NWS. MeteoFrance and NOAA/NWS inject the bulletins onto the GTS for distribution.

VOS and VOSCLIM

Voluntary Observing Ships (VOS) and Climate Subset Projects (VOSClim) collect surface marine observations for the GOS. To provide timely and complete information, and to ensure that no reports or information from participating ships are lost, data are submitted in both real time and delayed mode. Real time reports from participating ships are inserted onto the GTS in SHIP code. It is the responsibility of Data Assembly Centres to identify and extract these reports on the basis of a published and continuously updated list of call signs of participating ships.

In addition, all observations are recorded for delayed mode submission in either paper or electronic logbooks. Expanded logbooks were specially designed for the VOSClim project, and include additional required information. The recruiting countries digitise these observations in the revised IMMT format (IMMT-2), apply the agreed minimum QC procedures, and forward the digital data sets to the Global Collecting Centres (GCCs) for the WMO Marine Climatological Summaries Scheme (MCSS). The GCCs apply their normal QC and related procedures, and forward the data to the Data Assembly Centre in IMMT format (on an appropriate medium or via the Internet) with a minimum delay. Electronic logbooks will necessitate some extension to existing procedures. When recruiting countries collects these logbooks, they will carefully screen them for duplicates before full data sets are compiled. The data flow within VOSClim is depicted in Figure 9.

The VOSClim project also requires real time observational data monitoring, and comparison with model fields. To this end, The Met Office has established a Real Time Monitoring Centre, which already undertakes such monitoring of ship observations on a routine basis. Observations from participating ships will be identified by this centre, and associated with co-located model field values. These data sets will also be transferred on a regular basis to the Data Assembly Centre. The monitoring centre will:

1. Extract GTS reports of project ships (by call sign) and decode.
2. Associate project observed variables (pressure, air temperature, humidity, SST, wind speed and direction) for each project ship with co-located model field values (4 times daily).

3. Compile data sets of observations and associated model field values and transfer to the Data Assembly Centre.
4. Provide ship monitoring statistics for all VOSClim ships to the Data Assembly Centre (monthly).

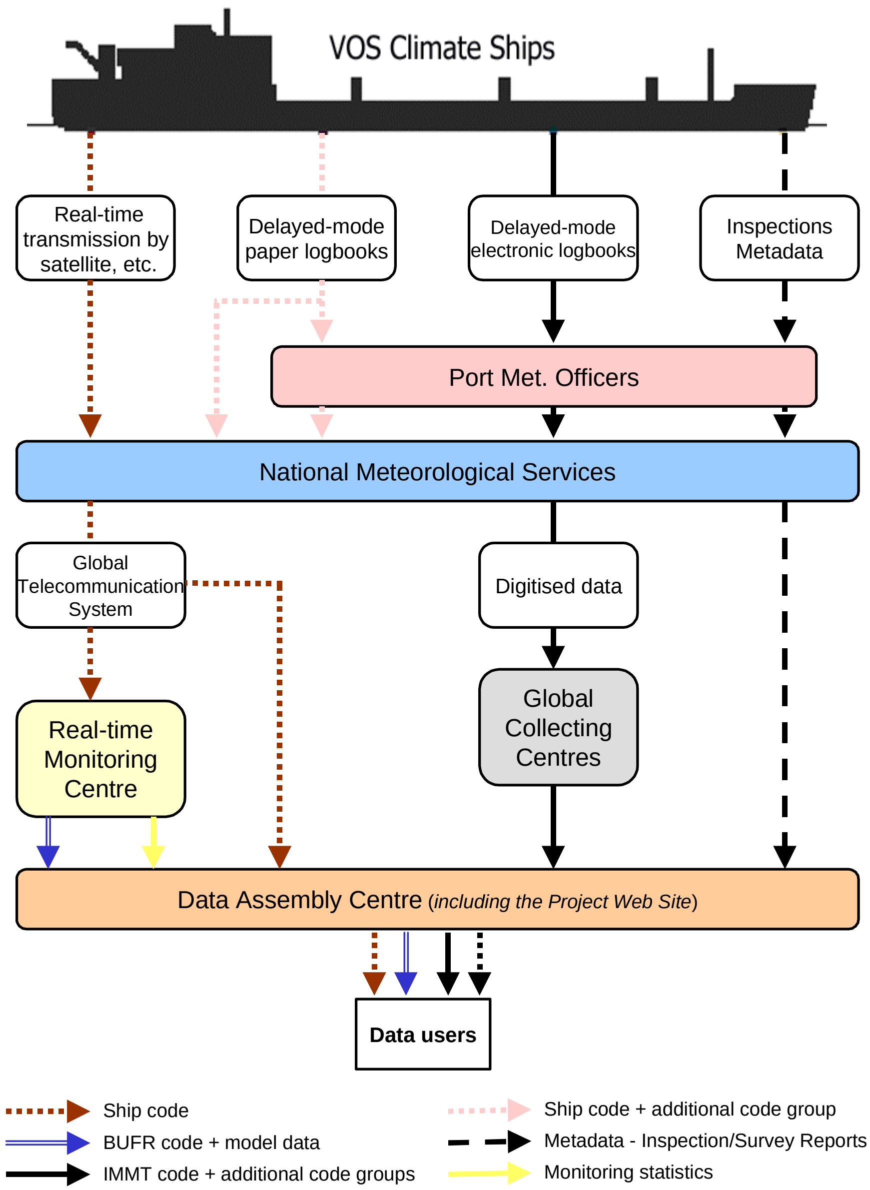


Figure 9. Data flow and data management with JCOMM VOSClim

Details of metadata, including digital imagery, instrument exposure and date of any changes, and ratings of the quality of instrument exposures, will be stored in a master index of ships. This will be developed as a supplement to, but separate from the main ship catalogue (WMO-No. 47). The catalogue will be continuously updated and made available through the Data Assembly Centre, as well as distributed in hard copy form to all Port Meteorological Officers through the national contact points. This catalogue will contain details of the instrument locations for each ship in an agreed format and will also contain details of the results of regular ship inspections.

The data collected during the project will be collected, quality controlled and archived by the project Data Assembly Centre, NCDC. The Centre will create and maintain a database so that the information on instrument types, exposure and observing practice can be automatically associated with each observation. The database should be freely accessible to registered users. The Data Assembly Centre will:

1. Extract and decode GTS reports of project ships (by call sign).
2. Receive the real time reports from the Real Time Monitoring Centre.
3. Collect delayed mode reports of project ships from participants.
4. Merge real time and delayed mode reports, eliminate duplicates and compile a complete project data set.
5. Collect metadata and survey reports for project ships and compile a complete data set.
6. Make project data sets available to users on request.
7. Maintain a project web site, information exchange mechanism and electronic newsletter.

GTSPP

The Global Temperature-Salinity Profile Programme (GTSPP) is a joint programme of the Intergovernmental Oceanographic Commission (IOC) Committee on International Oceanographic Data and Information Exchange (IODE) and the JCOMM. IODE is a technical committee of IOC and WMO. The data flows of national and international programmes within which GTSPP must find its place are illustrated in Figure 10. The boxes in the figure represent generic centres. A given international JCOMM or IODE centre may fit within several boxes in carrying out its national and international responsibilities.

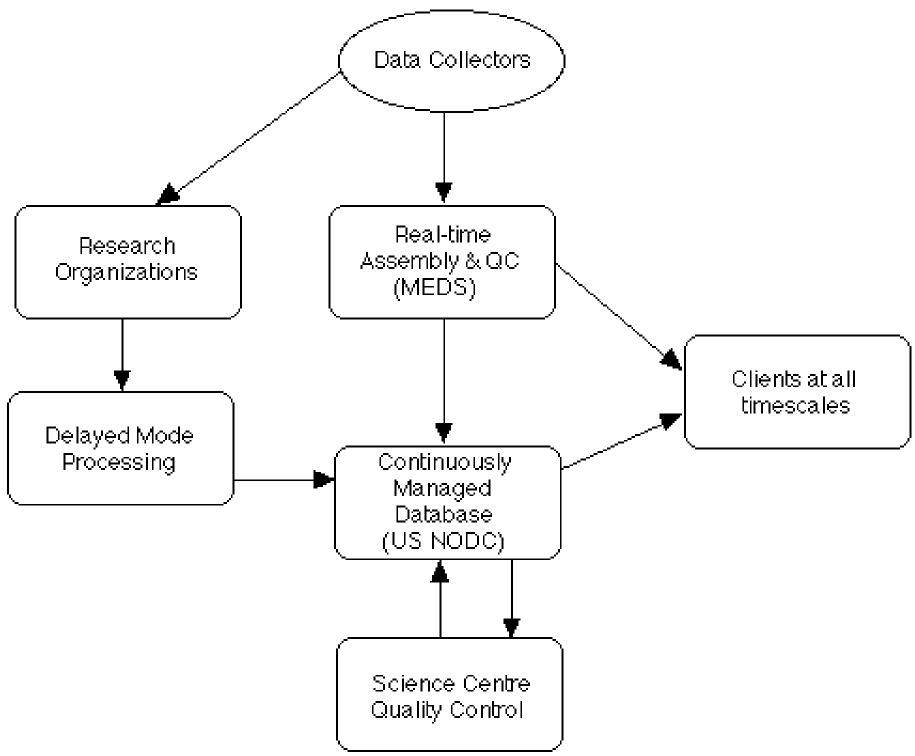


Figure 10. **GTSPP data flow**

There are seven essential elements of the GTSPP:

1. Near real time data acquisition
2. Communications infrastructure
3. Continuously managed database
4. Data flow monitoring
5. Delayed mode data acquisition
6. Quality control procedures
7. GTSPP data and information products

The ability to acquire data and to ensure that it is provided to the GTSPP Continuously Managed Database in a timely manner is dependent on the communications infrastructure. Operation of GTSPP involves a range of communication mechanisms, which include:

- satellite ship to shore (international),
- HF radio ship to shore,
- high speed networks (primarily the Internet),
- magnetic or optical media,
- GTS

In the future it is intended to move to faster and more automated means of communications for data acquisition and dissemination, and for the dissemination of data products and information.

The Internet and the World Wide Web have become increasingly important in the dissemination of data, information products, and information concerning the programme. Some of the products have near real time uses, the applicability of which can be greatly enhanced by their timely distribution. Publication of such products on the Web has become the option of choice to provide the data quickly to clients that have access to the Internet. For users without access to the Internet, other options for telecommunicated products must be implemented. For example, some products have been circulated to some Members by fax.

The concept of the continuously managed database (CMD) has been a significant advance in making data available to the user community sooner and in a form that is more convenient for them. Prior to the CMD concept, users had to contact three or more sources to obtain the most complete data set at any given time. The real time data from one source, historical data from another, and the delayed mode data, if available at all, would probably have to be obtained from one or more oceanographic laboratories. The data would be in several formats, and there would be duplicates of observations in several of the files obtained. Metadata would be scarce and varied and the use of flags and indicators would be different between the files from different sources.

The CMD has been designed to avoid these problems. Data should be available in one format with duplicates removed. The QC flags and the metadata indicators are standardized. The database should include the latest data available; as delayed mode data will replaced the lower resolution data collected from the GTS. The continuously managed database has been a very successful advance in data management within the GTSPP.

In the development of the GTSPP there were several problems with the manner in which QC systems were implemented in data centres. To address these issues standard quality control procedures were agreed and published in the GTSPP QC manual. Metadata and QC flags were standardized for real-time, delayed mode, and historical versions of the data. A table of information carried with each observation identifies where the data had been and what each organization had done to it.

The GTS is used to circulate reports around the world in real time. Two of the goals of the GTSPP are to improve the performance of IOC and WMO data exchange systems and to circulate monitoring reports on data flow and data availability. To accomplish this goal for the GTS, a monitoring system has been put in place for the flow of oceanographic reports on the GTS, in particular BATHY and TESAC reports.

The result of this monitoring indicated that various centres were receiving of the order of 65% to 95% of the data from month to month. Not only has this monitoring allowed the centres to improve their capture of real time data, but it has ensured that the data centres capture 95% or more of the available data instead of losing up to 30% per month.

4.7. World Climate Programme

The WCP depends upon the WWW for observation and transmission of CLIMAT and CLIMAT TEMP reports. There are, however, many other ongoing data management activities. Principal among these are efforts to define metadata for climate applications, data catalogues, and development and implementation of climate database management systems.

Metadata

The Commission for Climatology (CCl) Expert Team on Metadata for Climate Applications is tasked to:

(a) update, develop and record the requirements for metadata, particularly in support of the international exchange of climate data, noting the needs of climate change detection;
(b) recommend procedures for the recording and reporting of metadata by Members;
(c) develop any coding requirements for the transmission of metadata;
(d) develop the basis of an implementation plan to facilitate the agreement of procedures for the inter-national exchange of metadata and/or their deposition in major data centres;
(e) maintain close links with relevant groups such as CIMO, CBS, JCOMM and GCOS on related issues;

Dataset Catalogues

The CCl Expert Team on Dataset Catalogues has been tasked to, among other things:

(a) document guidance on dataset registration including national climate data catalogues;
(b) ensure that the *Manual on Codes* (WMO-No. 306) Volume I.1, Part A is current and complete for all climate observations;

There is clearly overlap between tasks (c) and (d) of the metadata team, task (a) of the catalogue team and the development of the WMO Core Metadata Standard by the Inter-programme Task Team on Integrated Data Management led by CBS. A representative of CCl has actively participated in the development of the WMO Core Metadata standard by this team. This issue will be examined further in section 6.

Climate Data Management System

WCP is also pursing the Climate Data Management System (CDMS) project. Nearly twenty years ago, the Climate Computing (CLICOM) project developed a PC-based data management system for use in developing country climate services. The CDMS project is now underway to evaluate and complete the documentation of new CDMSs. These systems will gradually replace the CLICOM system. A team of experts evaluated and documented each system in 2002 and a report for each system along with a comparative criteria table were produced. Among the extensive list of criteria are questions pertaining to the ability to import and/or export data in standard GTS message formats (TEMP, PILOT, SYNOP, METAR, CLIMAT, CLIMAT TEMP). There are also several criteria relating to collection, management and export of station-level metadata.

4.8. World Climate Research Programme

WCRP encompasses several sub-programmes and these in turn often comprise several projects. Each project is responsible for carrying out its own data management activities. Some of these activities are quite extensive.

ACYSYS

The principal objective of Arctic Climate System Study (ACSYS) data and information management is to provide timely information on ACSYS-related data sets and ancillary information, and to facilitate the information flow within the ACSYS community. The information flow also pertains to related programmes and projects of other organizations such as CLIVAR and GEWEX as well as new programmes. The ACSYS Data Management and Information Panel (DMIP) serves as a link between various ACSYS projects and as information source for data information of related projects within the spatial domain of ACSYS.

Both the hierarchy of data and information collected for ACSYS since 1994 and the historical network data collected in connection with and independent of ACSYS, are recognised. The DMIP activities cover raw data, corrected/calibrated data, and assimilated data (including results of re-analysis of datasets and high-level data products such as gridded data products).

The ACSYS DMIP operates as a system of collaborating data and information providers; i.e., global or topical data centres. The day-to-day management activities are implemented through the ACSYS Data Information Service (ADIS). ADIS holds metadata about information sources and description of datasets, and it provides access to major ACSYS-related publications. The system is not intended to hold actual data. The principal medium for communication is the Internet. However, it is recognized that some locations have limited Internet access and may have to be serviced by other means (fax or postal services).

The metadata about ACSYS datasets held by ADIS includes information on the content of the datasets, its location, contact persons, etc. The metadata format used is based on the ISO 19115 Geographic Metadata International Standard and the WMO Core Metadata profile is the basis for the structure of the database.

Data quality is of paramount concern to all ACSYS activities. Hence, all ACSYS data sets should be accompanied by documentation, which allows an informed judgement on the usefulness of the data for specific research activities. Data quality is also subject to user requirements and is therefore, as such, not an absolute term. The quality, or usefulness, of a specific data set for research purposes is therefore to be assessed by the user. The adoption of data standards; i.e., maximum allowable error bars, sensitivity of sensors, time-and spatial resolution of data sets, are specific for each data set and may be agreed upon (where applicable) by scientists and data providers, if this data will be collected or re-analyzed or otherwise processed for ACSYS purposes.

ADIS is the technical tool of DMIP that implements communication and flow of information. It contains a meta-database consisting of existing and progressing datasets and their status, as well as other related information. The status of proposed and planned data collection activities, in addition to that of on-going ACSYS research projects are also included. Regional overview of collected data, visualization of space- and time-distribution of key data sets; etc., are used to show the state of the data-collection efforts. Links to other collaborating centres and institutions are given to assist the users.

Many ACSYS-DMIP activities are of direct interest to other WCRP projects such as GEWEX, and its associated projects such as BALTEX and GAME. CLIVAR is also expected to benefit from these activities. DMIP reports to G3OS JDIMP and maintains liaison on an informal basis with G3OS concerning data and information management.

GEWEX

The GEWEX Hydrometeorology Panel (GHP) Data Management Working Group (DMWG), is to "Assist the GEWEX Hydrometeorology Panel in the coordination and facilitation of data management activities and issues between the GEWEX Continental Scale Experiments, the Global Runoff Data Centre, and the International Satellite Land-Surface Climatology Project."

The major responsibilities of the DMWG are, among others, to:

1. act as point of contact for data management questions on behalf of the GHP or assist (whenever possible) in data requests. The DMWG will also coordinate data management activities/issues with other GEWEX panels/organizations where appropriate.
2. facilitate data policy issues, data exchange restrictions, and data access/procedures.
3. develop and draft a data management plan for those GHP enhanced observing periods. This Data Management Plan will address specific issues regarding identification of

“standard” and “validation” data sets (i.e. formats, resolution, frequency, parameters, etc.) and establish specific working protocol procedures among the GHP participating members.

4. develop a plan (including resources) to compile those “composite” data sets or CD ROMs approved by the GHP with data contributions from three or more GHP members.

CLIVAR

CLIVAR's data management policy is articulated in its Initial Implementation Plan which itself gives broad guidelines on CLIVAR data management. A CLIVAR Data Task Team was set up to:

1. define CLIVAR's requirements for a data and information system
2. assess the extent to which existing data management systems meet the CLIVAR requirements
3. ensure the rapid, responsive delivery of data and data products
4. ensure the secure but accessible archival of CLIVAR data
5. ensure the delivery of information on the location and availability of CLIVAR data
6. make recommendations on actions that need to be taken to ensure an adequate CLIVAR data and information system.

This group was disbanded following its first meeting. However the key recommendations of the panel have since been implemented by the International CLIVAR Project Office (ICPO). These were:

- to establish data/products liaison members for each of CLIVAR's panels and working groups with a remit to identify key data and data products relevant to CLIVAR
- given CLIVAR's key role in the role of the oceans in climate, to transition the existing WOCE Data Assembly Centres (see below) into CLIVAR Data Assembly Centres.

The ICPO is now working, through an identified member of staff with responsibilities for liaison over CLIVAR data issues, to develop the activities of the Data Assembly Centres for CLIVAR and proposals for a Data Assembly Centres Workshop are currently being developed. CLIVAR is also working with IOC on development of a joint hydrographic and carbon database linked to the WOCE repeat sections. Because of the broad nature of CLIVAR, the current strategy is for CLIVAR to work through a distributed network of data centres accessed via links on CLIVAR Data Information Pages of the CLIVAR web site. These pages have been established and are under active further development in consultation with the CLIVAR panel and working group liaison members. By its very nature, many of the datasets CLIVAR deals with are "CLIVAR-relevant" rather than produced by the CLIVAR project per se. A current exception is the data produced under the CLIVAR Variability of the American Monsoon (VAMOS) Panel field programmes which are managed and archived through a VAMOS Project Office run by UCAR's Joint Office for Science Support (JOSS).

At its last meeting, the CLIVAR Scientific Steering Group agreed that data management in CLIVAR would be driven by a new panel, the CLIVAR Global Synthesis and Observations Panel. The Scientific Steering Group also agreed on the need for a focussed CLIVAR data management workshop and for the development of a coordinated plan drawn together through the help of a data consultant. These issues are being taken forward currently through the ICPO.

SPARC

The SPARC data centre’s objective is to facilitate data exchanges between participating scientists. Established in June 1999, the centre provides access to a several datasets and their number is growing rapidly. Since the number of datasets is limited they are listed individually rather than in a formal dataset catalogue. Therefore, it has not been necessary for the centre to adopt a metadata standard.

WOCE

The WOCE data management structure was an internationally distributed system, which utilized the expertise of scientists to attain the highest possible data quality and documentation. Each measurement technique produced a different data "stream", and the WOCE data management system brought these streams together to form a single data resource. The data stream consisted of several elements with the flow being from Principal Investigator (PI) to Data Assembly Centre (DAC) to Special Analysis Centre (SAC) to users and then to an archive as illustrated in Figure 11.

The Data Assembly Centres (DACs) were managed by scientists to handle, assemble and quality control the data sets, and often generate data products. The Special Analysis Centres (SACs) performed data analysis and synthesis functions, including the generation of derived data sets. The WOCE Archive is distributed through these CD ROMs and across the centres of the World Data Centre System. The WOCE Data Information Unit (DIU) acted as a central online source of information on the status of WOCE, tracking all data collection, processing and archiving activities, and providing the primary interface between the WOCE data system and all users.

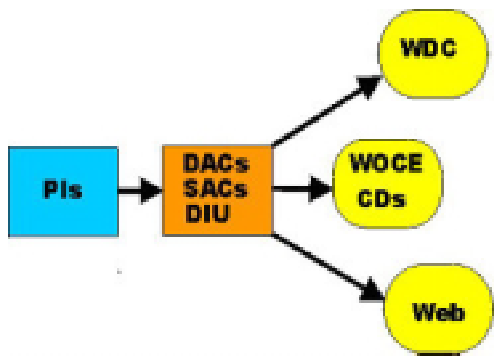


Figure 11. WOCE data flow

4.9. World Weather Watch

There are several current activities within the WWW with a major data management component. These include the projects on enhanced use of data-communication services, the improved Main Telecommunication Network, integrated WWW monitoring, metadata standards, and the Future WMO Information System.

The project on Enhanced Utilisation of Data Communication Systems aims to:

a. Develop recommended practices and technical guidance material for the implementation of data communication facilities (GTS and Internet) at WWW centres;
b. Review telecommunication and information system requirements of the WWW and other WMO Programmes that can be effectively met by the Internet;
c. Develop recommendations on coordinated use of the Internet to meet in the relevant short-term (1-3 years) requirements of all WMO.

IMTN

The goal of the Improved Main Telecommunication Network project (IMTN) is to develop and implement an improved MTN, including supported applications, data transport functions, responsibilities of MTN centres, interfaces and gateways between the MTN and RMTNs. It will be implemented through data-communication network services from a small number of providers. A first phase would mix network services and point-to-point circuits; a second phase would provide full MTN connectivity through the network services. The project will facilitate a progressive implementation, which could be adapted to the needs and resources of the Members concerned and could respond to changing requirements. The IMTN is expected to permit savings for most centres on recurrent costs in comparison with the current leased circuits, while enabling capacity upgrades. The IMTN consists of two parts:

(a) The implementation of a “cloud I” providing the interconnectivity between Regional Telecommunication Hub (RTH) / World Meteorological Centres (WMCs) Washington and Melbourne and RTHs Tokyo, Bracknell, Brasilia and Buenos Aires, including RTH/WMC Moscow in a further step;

(b) The implementation of a “cloud II” as an extension of the Region VI RMDCN, providing the interconnectivity between RTHs Bracknell, Toulouse, Offenbach, RTH/WMC Moscow and other adjacent RTHs, i.e. RTHs Nairobi, Dakar, Algiers, Cairo, Jeddah, New Delhi and Beijing. The inclusion of the Tokyo-Beijing and Tokyo-New Delhi circuits would also provide an effective interconnectivity between both “clouds”.

Metadata

The effort to develop a discovery-level metadata standard is closely linked with the Future WMO Information System (FWIS), since a standard approach to describing data is central to concepts behind the proposals for FWIS. “Metadata” is information that describes data. There are three stages in making use of data; a different level of detail is required for metadata at each stage:

(a) Discovery level – The first step in using data is to find out (“discover”) where the data of interest might be obtained;

(b) Request level – Once potential sources of data has been identified, the user can approach one of the data providers with a more specific request;

(c) Usage and management level – The third level of metadata is more detailed, describing not only the general aspects of the data but also specific details (such as the precision and units and details of quality control and processing procedures).

Over the past few years, CBS experts and representatives of other Technical Commissions have developed a “Core WMO Metadata Profile” within the context of the newly approved ISO standard for geographic metadata (ISO 19115). The core metadata are intended to be adequate for the “discovery” level.

WWW Monitoring

The project on integrated WWW monitoring comprises two parts: an operational trial of the proposed integrated monitoring and an extension of the special MTN monitoring. With respect to the relevant extension of the special MTN monitoring, all participating RTHs on the MTN had been invited to implement extensions, as from October 2002:

With expanding use of the BUFR code, in particular through the migration from traditional alphanumeric codes to table-driven code forms, it is increasingly important to monitor data presented in BUFR. Therefore, RTHs on the MTN have been invited to participate in the preparation of a pilot study and in preliminary tests for the monitoring of BUFR bulletins.

Several WMO Programmes have expressed a critical need for real-time monitoring of data flow over the GTS. However, the project on integrated WWW monitoring does not propose to meet this real-time requirement.

Future WMO Information System

In 1998 CBS reiterated its commitment to ensure that the Basic Systems meet the requirements of all WMO and related international programmes. The Inter-programme Task Team on the Future WMO Information System was established in 2000 to develop a strategic overview of WMO information systems and services. At that time the president of CBS invited the presidents of each of the other Technical Commissions to identify the information system requirements of their respective programmes and to nominate experts to participate in the work of the task team.

The WWW's primary requirement is for an information system focused on the collection of observational data, and the exchange of products based on these data, in real time. This exchange is underpinned by Resolution 40 (Cg XII), which relates to conditions applying to the free and unrestricted exchange of meteorological and related data and products, and is generally achieved via the GTS.

The requirements of the Technical Commissions other than CBS are met, in part, by the WWW's real-time datasets exchanged via the GTS, and also by exchange via the Internet. The Internet is used for the ad-hoc exchange of files and for the routine downloading of specific files held on Member's servers (including satellite data and data types not generally exchanged on the GTS).

Current practice for the exchange of data and products on the GTS is to consolidate these into bulletins prior to their passage through the GTS. The GTS operates using a "store and forward" system by which data and products are forwarded from one RTH to the next in accordance with pre-agreed switching tables. RTHs also pass data on to adjacent National Meteorological Centres (NMCs) as agreed between the RTH and NMC. Thus, the GTS requires that bulletins may need to be switched through a number of RTHs before all their users are able to access them. Special monitoring exercises on the GTS show that while a number of centres may aim to collect the global set of observations, there are losses around the system that prevent this situation being completely realised.

In addition to this routine transmission of data, WMO programmes are increasingly seeking the capability to make ad hoc requests for data and products, a capability that is being explored with non-standardised, Internet-based request-reply systems. The growing need for ad hoc transmission and the inflexibility of the current store and forward GTS is leading some to development of non-standard solutions by the various projects and Programmes.

One option to address these problems might be to enhance the GTS in such a way as to generalize the services to all Programmes. However, the GTS would still suffer from inherent deficiencies, some of which are listed below:

- Use of proprietary high-level protocols that are not supported by the marketplace.
- Volume restrictions preclude the transmission of high volume data sets (in the order of gigabytes or terabytes).
- Lack of support for a request/reply system providing ad-hoc access to the data and products available for international exchange.
- Inability to facilitate information insertion and distribution to programmes and public and other clients beyond the meteorological community.
- Inability to rapidly (i.e. routinely near-real-time) identify where data losses are occurring and undertake remedial action.
- Inability to easily accommodate requirements that include short periods of high volume traffic followed by lengthy periods of low or no traffic.

Therefore, an alternative approach has been proposed: a coordinated global infrastructure, the Future WMO Information System (FWIS). It is envisioned that FWIS would be used for the collection and sharing of information for all WMO and related international programmes. The relationship between functions performed by FWIS and similar functions performed by current WMO Programmes is illustrated in Figure 12 below. The FWIS vision has attempted to provide a common roadmap to guide the orderly evolution of these systems into an integrated system that efficiently meets all of the international environmental information requirements of Members.

FWIS should provide an integrated approach to meeting the requirements of:

- Routine collection and automated dissemination of observed data and products ("push").
- Timely delivery of data and products (appropriate to requirements)
- Ad-hoc requests for data and products ("pull")

FWIS should also support:

- Different user groups and access policies, such as WMO Resolutions 40/25
- Data as well as network security
- Integration of diverse datasets

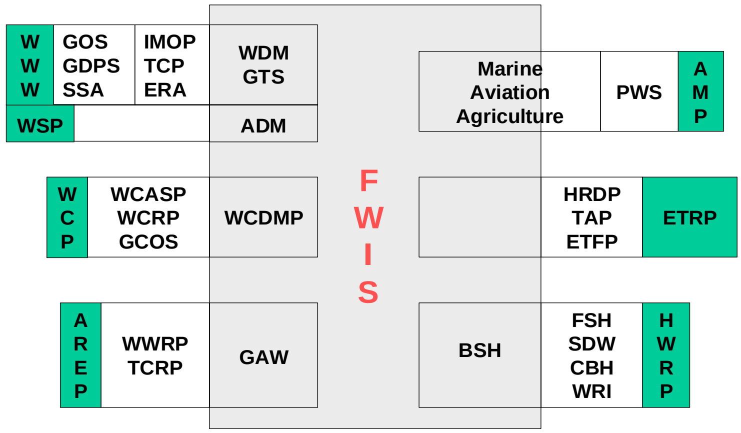


Figure 12. FWIS relationship to WMO Programmes

The implementation of FWIS would build upon the most successful components of existing WMO information systems. It would continue to rely upon the WMO communication system (initially the GTS) to provide highly reliable delivery of time-critical data and products. Currently, this requires a private network but this is likely to change as public communications services evolve.

For the near future, transmission of the current suite of global products will continue to be distributed to WMO Centres via the existing GTS infrastructure. However, implementation of request/reply systems and exchange of high volume datasets (e.g. radar data, satellite imagery, and high resolution model output) cannot be supported by the existing GTS. Realization of the FWIS vision would require that the existing GTS dedicated communication links and message switches be augmented by additional communications capabilities such as those provided by the Internet and other communication options.

4.10. WMO Space Programme

The European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) is implementing a multicast data distribution service called EUMETCast. It is based on a concept called 'Service Directories' into which the products to be multicast are transferred from the EUMETSAT ground segment. These service directories are part of the multicast system normally resident at the service provider's uplink site. The service directory structure is also used to allocate bandwidth characteristics to a particular set of products or directories.

It is proposed to extend EUMETCast to include broadcast of several data streams that are now transmitted by other means. A 1 Mbps channel would be used for the transmission of basic meteorological observation and forecast data for users in WMO RA-VI. Another channel would be used to carry DWDSAT products. DWDSAT is DWD's FAX-Europa (FAX-E) follow-on system. Météo-France has also proposed to provide their African data distribution service, RETIM-Afrique, via EUMETCast in a similar manner. On the user side, this would allow the same reception station to receive both the original C-band EUMETCast services, and the additional services with relatively minor software modifications.

Expert Team on Satellite System Utilization and Products (SSUP) noted the potential the Internet provided for distribution of meteorological data as a complement to direct broadcast systems (DBS). The team was well aware of current shortcomings in the Internet system, but was generally of the opinion that an effective Internet system for data exchange could be expected globally in the coming years.

It has been agreed that access to satellite data and products by WMO Members should be through a composite data access service comprised of both direct broadcast from satellite systems and alternative dissemination mechanisms (ADM). ADM would be the baseline while direct broadcast reception would serve as back-up as well as for those WMO Members unable to take advantage of ADM.

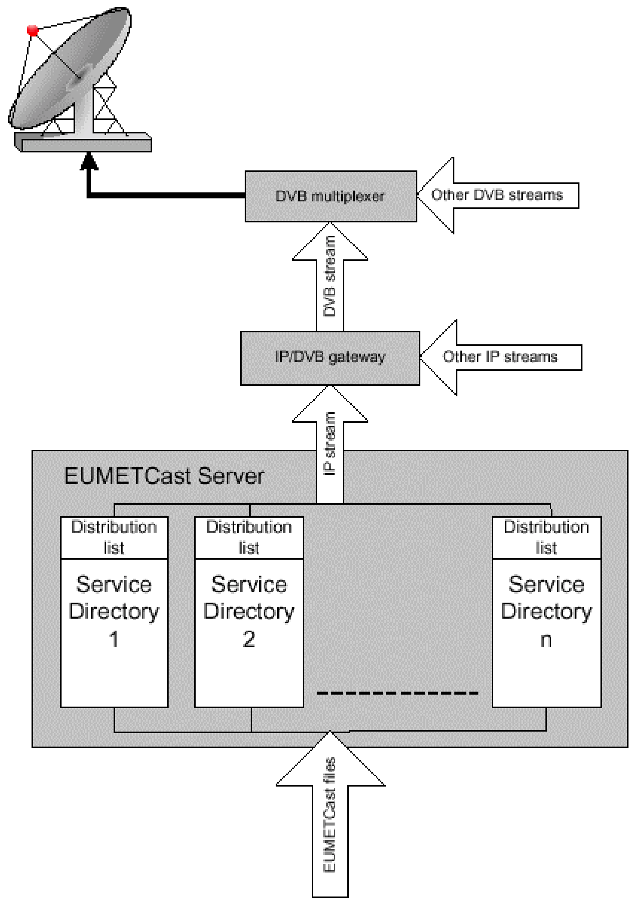


Figure 13. EUMETCast data flow

The team noted that ADM was likely to bring significant advantages including:

- High performance, allowing high data rates of several Mb/s, thus providing access to a wide range of data and products;
- Flexibility, allowing enhancement of dissemination during the lifetime of a satellite generation with additional products that were not included in the initial design;
- Capability to include data from spacecraft that were out of the visibility of the user;
- The availability of low cost user terminals would make data access affordable for a larger number of users;
- It could facilitate a smooth transition between different satellite generations for the user community;
- It would be possible to combine reception of satellite data with the reception of other meteorological data, which would save costs and facilitate operation, since the same or similar terminals could be used;

With regard to environmental satellites (i.e., R&D or operational satellites which were relevant to environmental observation, but not considered as operational meteorological satellites), it was recognized that they would be contributing to the space-based component of the GOS. However, data dissemination from these spacecraft did not imply the same level of coordination as operational meteorological satellites since they were beyond the scope and mandate of CGMS at the present time. Thus, there was an urgent need to take initiatives to establish near-real time dissemination data streams to the meteorological user community in a suitable way, i.e., without requiring that each WMO Member invest in a dedicated, expensive and complex receiving facility.

On the basis of but not necessarily corresponding strictly to WMO Regions, each satellite operator should cooperate with WMO Members to establish a new service to deliver the full suite of data and products from its satellite systems. This should include, but not be limited to, the comparable data content delivered by the standardized global dissemination service.

4.11. Activities of other (non WMO) organizations or groups

Unidata IDD

The Internet Data Distribution (IDD) system, developed by the UCAR Unidata Office, is used to distribute meteorological data via the internet to more than 150 universities in North America. They originated this service in 1995 and have continually improved the reliability and features of the software, which is distributed at no cost. The IDD utilizes TCP/IP protocols that are managed by the Local Data Manager (LDM) software. Transmission of data is reliable and currently handles traffic of more than 200 Gbytes daily with an average delivery time to top tier sites of less than 1 minute.

The IDD is also now used for communications of operational data and products within several NMHSs. In addition, the IDD is currently being used within the USA to transmit many gigabytes per day of Doppler radar data to NCDC for archival. In a survey of users in 2003, 97% of users reported being satisfied or very satisfied with the services provided by the IDD.

The system is similar to the GTS, in that both systems handle data from multiple sources, are configurable, and allow any site to inject data into transmission streams. A major advantage of the IDD is that it is freely available and utilizes communication protocols that are fully supported by industry standards. The software runs on many Unix operating systems as well on PCs running Linux.

The IDD ingests data products based on standard WMO message headers or other unique product headers. It is possible to send and receive any product encapsulated within a recognizable header.

Unidata monitors data transmission reliability using near real-time graphics that give a visual display of network latencies. Under normal network conditions these latencies are usually less than 5 minutes to any point in the network. However, if problems occur that degrade performance at a particular location, this is immediately highlighted by the monitoring software and the system can be reconfigured within minutes to provide alternative sources to the data.

Earth System Grid II

The Earth System Grid II (ESG) is a research project sponsored by the U.S. Department of Energy (DOE) Office of Science. The primary goal of ESG is to address the formidable challenges associated with enabling analysis of and knowledge development from global Earth System models. Through a combination of Grid technologies and emerging community technology, distributed federations of supercomputers and large-scale data and analysis servers will provide a seamless and powerful environment that enables the next generation of climate research.

Several organizations, primarily within the U.S. DOE, contribute to ESG as follows:

- Lawrence Berkeley National Laboratory: Climate storage facility and access
- Lawrence Livermore National Laboratory (LLNL): Model diagnostics and inter-comparison

- Argonne National Laboratory: Computational grids, and grid-based applications
- Oak Ridge National Laboratory: Climate data storage and computational resources
- University of Southern California, Information Sciences Institute: Computational grids, and grid-based applications
- NCAR: Climate change prediction and scenarios

The primary goal of ESG is to make climate data – particularly climate model data – an easily accessible community resource. ESG has adopted NetCDF as its data model and is also developing metadata standards for gridded climate data.

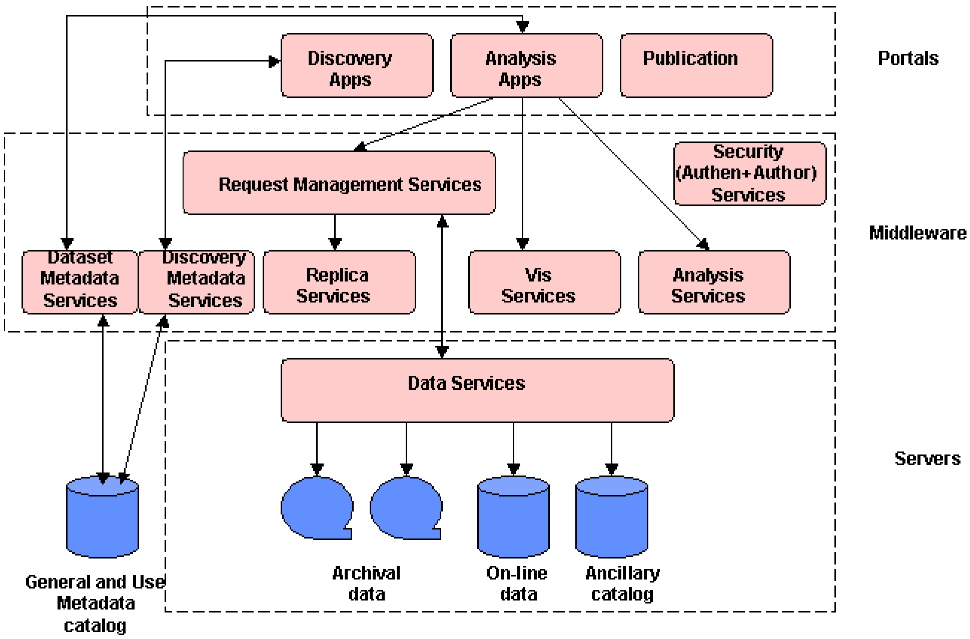


Figure 14. Earth System GRID Architecture (Draft)

<u>LLNL Climate Data Management System</u>

As part of its contribution to ESG the Lawrence Livermore National Laboratory has developed the Climate Data Management System. It is an object-oriented data management system, specialized for organizing multidimensional, gridded data used in climate analysis and simulation. <u>Climate Data Markup Language</u> (CDML) is the mark-up language used to represent metadata in the Climate Data Management System. CDML is an international Markup Language (XML) dialect geared toward the representation of gridded climate datasets. XML provides rigor to the metadata representation, ensuring that applications can access it correctly. XML also deals with internationalization issues, and holds forth the promise that utilities for browsing, editing, and other common tasks will be available in the future.

5. ADEQUACY, COMPLEMENTARITIES AND COMPATIBILITY OF THE DATA MANAGEMENT STANDARDS, PRACTICES, ACTIVITIES AND PLANS

The analysis of ongoing and planned activities of the WMO and related programmes indicates that there has been good cooperation and coordination between most Programmes. However, there is also considerable room for improvement. In general, shortcomings of existing data management activities within the WMO Programmes can be traced to one or more of three basic causes:

- Insufficient recognition of the importance of data management in programme planning
- Insufficient knowledge, resources or commitment to management of data
- Inadequate coordination or communication

For the most part, real-time collection and dissemination of operational data are well coordinated. Much of these data are carried over the GTS. However, even for these data many problems have been identified, principally loss of data and a lack of real-time monitoring information. Shortcomings in real-time data flow are usually due to a shortage of resources or current capabilities and are not due to a breakdown in coordination. However, the fact remains that critical requirements are not being met.

Collection and transmission of delayed mode data, use of data for by multiple programmes, and ad hoc access to data and products are problematic. These functions have not been as well coordinated, there has been more duplication of effort, and incompatible standards have been developed or used.

Historically, observing, transmission and information management systems have been designed and developed to meet the specific requirements of a single programme. The needs of other programmes for the same data have sometimes been recognized, but not always. Even when recognized, these requirements have most often been considered of low priority relative to the requirements of the funding programme. As a result, whenever budget pressures have forced compromises or tradeoffs, efforts to address the downstream or secondary requirements for the data have been the first to be cut. As a consequence, systems are often designed to meet a limited set of the highest priority requirements. Pressing deadlines and limited resources result in systems that meet critical operational requirement but are not flexible enough to easily take advantage of advances in technology or to respond to evolving requirements.

There are many cases where a relatively small adjustment by one programme could have produced a large benefit for another programme. However, in most cases this ancillary requirement was either not articulated or recognized or the additional investment was judged to be outside the scope of the responsible programme's mandate.

These lost opportunities are particularly evident in the relationship between the WWW and WCP. The WCP receives the bulk of its data from observations made for and by the WWW and transmitted over WWW facilities. However, the climatological requirements regarding these data have not always been recognized. Thus stations are often moved to a new location without any overlap in their operations, resulting in a detrimental impact on long-term data continuity. Additionally, the station operators have not consistently addressed the climatological requirement for detailed and specific metadata on the station exposure, instrumentation and operation.

A key component of data management includes timely quality control of the observations by the monitoring centres and notification to observing system operators and managers of both random and systematic errors, so that corrective action can occur. For the WCP an operational system is needed that can track, identify, and notify network managers and operators of observational irregularities, especially time-dependent biases, as close to near-teal time as possible. Such feedback systems are currently not routine at monitoring and analysis centres. Equally important is the follow-up required by operators and managers who are responsible for implementing timely corrective measures.

It is clear from detailed network reviews that the flow of real-time data to the user community and to the international data centres for the variables of high-priority to climate is inadequate. This is especially true for many terrestrial-observing networks. Primary causes are a lack of resources, inadequately integrated data-system infrastructure and insufficient or ineffective monitoring of climate data.

It is evident that many of the problems in existing data management systems and capabilities are primarily due to a historic lack of resources and are thus unlikely to be remedied by any recommendations made in this report. However, some deficiencies are primarily the result of poor communication or insufficient coordination and there is reason to hope that these problems could be remedied.

Specific shortcomings primarily due to insufficient coordination

- There are a large number of different applications developed to serve the needs of different Programmes whose development has not been coordinated. There is little connectivity between applications resulting in inadequate/inflexible data exchange standards and incompatible data formats, making integration of data sets technically challenging.
- Inflexible or incompatible data transmission standards, procedures and protocols.
- Diverse and incomplete quality control of observational data and insufficient documentation of procedures for these data to be effectively used by other Programmes.
- No standard practices for the collection, electronic archival and exchange of metadata, both high-level and detailed, especially for stations and instruments.
- Lack of agreed standards needed to effectively identify, acquire and use all of the relevant data, particularly from archives, hampers multidisciplinary application of meteorological, hydrological and oceanographic data
- Poor coordination between observing systems, making it difficult to identify the potential of observation sites established by one Programme to meet the requirements of other Programmes.

6. CURRENT OR RECENT COORDINATION ACTIVITIES

The potential benefits of effective coordination and cooperation between the information management activities of the various WMO Programmes have been recognized for many years and a number of efforts have been made to enhance coordination.

Presidents of Technical Commissions meetings

Meetings of the presidents of the Technical Commissions (PTC) have been held for several years. They are intended to provide an informal forum for the presidents to inform others of significant activities of their Commission and to keep abreast of developments and activities within the other Commissions. In this role the meetings have been successful and have had the added benefit of helping to develop rapport between the presidents.

There has been some hope that these meetings might lead to joint activities and enhanced cooperation. In this role they have not been as successful, since no joint projects or activities have resulted. This is likely because the presidents are not at an appropriate level to initiate these activities. The presidents generally do not have the detailed first hand knowledge of ongoing activities and possible synergies held by experts at the working level. Furthermore, they are generally not in a position to commit resources or staff from other than within their own NMHS. Staff at a high working level with budgetary authority would probably be in a better position to propose joint projects.

There appears to be relatively poor communication between experts within a Commission and their president, both up and down. Experts representing a Commission on inter-programme task teams often do not keep their president (or the Secretariat) informed of important issues or proposals resulting from these meetings. Likewise, the presidents often do not communicate pertinent information from the PTC meetings to their experts.

The PTC meetings serve an important role and are certainly worthwhile. However, they are probably not the most effective means to ensure effective coordination of information management issues and activities between Commissions.

Inter-programme data management coordination meetings

Beginning in the mid 1990s a series of three inter-programme data management coordination meetings were held. Representatives of each of the scientific and technical programmes of WMO, as well as representatives of other international programmes with significant data management components, such as CEOS, IGBP-DIS and the ICSU World Data Centres, were invited and most sent representatives. These meetings provided a useful forum to exchange views and information on the activities of the various programmes. However, they did not produce any concrete results, such as agreement on standards or joint development of information systems.

The most significant outcome of these inter-programme coordination meetings was a proposal for coordinated development of a harmonized information management system for all of WMO. This has been translated into development of FWIS and a WMO metadata standard, which are described in section 4.9 above. Both of these efforts have attempted to identify and respond to the requirements of many Programmes.

<u>FWIS</u>

From the inception of FWIS it has been recognized that its implementation and operation would require the participation of many Programmes and centres. If FWIS is to effectively support the requirements of many diverse Programmes, each Programme must actively participate and contribute its own expertise and resources.

There has been an earnest attempt to include all WMO Programmes and Technical Commissions in the development of the FWIS vision. The presidents of all Technical Commissions were invited to nominate representatives to participate in FWIS planning. Most presidents responded positively and the FWIS task team includes members from five Commissions. The FWIS vision thus recognizes and attempts to address the requirements of these Programmes. Nonetheless, there have been some problems with the effort to date.

There is a perception among several programmes that FWIS is overly controlled by and focused on CBS requirements. Many feel that FWIS considers CBS requirements to be paramount and the requirements of the other Programmes to be secondary, to be addressed only as long as they do not adversely impact on meeting CBS requirements. Unless this perception is addressed other programmes are unlikely to make a significant commitment to FWIS in either manpower or financial resources.

Within some programmes FWIS is seen as a success as a concept but not in application nor as a roadmap toward a common future. Each programme has committed itself to manage and exchange its data in a certain way and has followed its chosen course over many years. There is a large commitment in resources to current practices and changes or adjustments would require additional resources, at least in the short term. Therefore, there is a tremendous reluctance to change current practices and procedures unless there are compelling reasons to do so.

FWIS is not seen to be essential for meeting the critical requirements at the heart of most Programmes. CBS is leading the effort but CBS is not perceived to have the most pressing, un-met requirements and, therefore, is not sufficiently driven to push rapid implementation. Instead, FWIS is seen by some to offer vague promises of inter-programme interoperability or reduced duplication of effort that are not sufficient to justify the significant investment that its implementation would require. Without a dedicated budget or an effective champion behind it, few experts are willing to bet the success of their programmes on the implementation of FWIS.

The comment of one officer within the Secretariat illustrates a serious failure in outreach beyond CBS. He/she stated that they had heard of FWIS but did not have a clear understanding of its goals and did not believe it would effectively contribute to meeting the needs of his/her Programme. Instead it was mainly regarded as a distraction at best. It was feared that FWIS would be imposed upon his/her Programme from above and would divert their limited resources to a bureaucratic paper exercise with no useful outcome. Clearly the efforts that have so far been made to elicit participation and communicate progress in FWIS have not been entirely successful.

<u>Metadata Standards</u>

The need for WMO-wide metadata standards was identified at the second Inter-programme data management coordination meeting in 1999. This requirement was further refined during development of the vision for FWIS. As a consequence CBS tasked its Expert team on

Integrated Data Management to develop this standard. The president of CBS invited the presidents of the other Technical Commissions to nominate experts to participate in this effort and experts from two other Commissions (CCl and CAgM) have actively contributed.

At this point it is too early to determine if the WMO Core Metadata Standard developed by the team will be embraced by any WMO Programme. Nonetheless, it is already clear that the communication of the standard to all relevant groups has not been sufficient. Although the need for a WMO-wide standard was recognized by most of the staff interviewed in the Secretariat, very few were aware that a metadata standard had already been developed. Some were vaguely aware of a team within CBS for integrated data management but few outside of CBS were aware that one of the principal responsibilities of the team was to develop a metadata standard. The proposed standard is even less well known outside the Secretariat.

7. RECOMMENDATIONS

Within WMO, each Programme has traditionally worked independently to develop and implement its own plans and activities. Programmes have worked together towards a common goal only where possible benefits were clear and immediate. An example of success in this area can be found in the cooperation between the WWW and Programmes that use the Basic Systems to generate and transport their data and information. This is particularly true for the Marine and Aeronautical programmes. That all three of these programmes have been managed within the same department in the Secretariat is probably not a coincidence.

There are many reasons for the traditional independence of the Commissions and Programmes.

First and foremost, coordination between departments and between Technical Commissions has historically not been a high priority within WMO. It is clear from the Technical Regulations that it is assumed most activities are to be carried out within a Commission. Although inter-commission groups are possible they are defined as the exception rather than the norm. The WMO budget is defined and administered by Programmes and the departmental structure of the Secretariat mirrors that arrangement. Creation of inter-commission groups thus introduces significant complications in management, budgeting and reporting responsibilities.

Second, many experts and staff see coordination as a cost with little benefit. That is, the effort required to coordinate activities with others can significantly slow progress and may require compromises and introduce complexity that would not be needed if the activity were narrowly focused on the needs of a single Programme. Since major activities within WMO are usually initiated in response to a pressing need, time spent on coordination is often seen as introducing unacceptable delay. This attitude also applies to standards, which are developed only when there is a compelling need and, again, usually within a single Commission.

Third, bureaucratic rivalries between Commissions and departments are an unfortunate reality. This does not imply that directors or staff make deliberate decisions not to cooperate. Rivalry instead manifests itself in the feeling that “we know best” or “we can do it better”. Thus, each Programme sees its expertise as superior and its requirements unique. In such an atmosphere, it is quite rare for one Programme to approach another Programme or department to ask for assistance.

Over the past several years it has become increasingly evident that many scientific and technical problems are inherently interdisciplinary in nature. It has been recognized for more than a decade that data and information management is an area where better coordination and cooperation could lead to increased efficiency and improved services. New crosscutting WMO Programmes (Space and Natural Disaster Prevention and Mitigation) presage increased requirements for well-coordinated information management. Several attempts have been made to address this issue. However, the organizational and bureaucratic obstacles have made this difficult.

7.1. Within the Secretariat:

Although substantial progress has been made over the past several years, current mechanisms for coordinating data management activities within and between departments in the Secretariat could be further improved.

Problems identified

- Poor communication. Often staff members within a department are not aware of related on-going activities in another department. Conclusions or results of meetings are not effectively communicated to others not in attendance. While proposed standards and guidelines are often approved by Executive Council or Congress, staff most often attend only those sessions pertaining to their own Programme. Consequently, they are not aware of standards developed by other Programmes that could be applicable or relevant to their own.
- Budgetary issues. The WMO budget is allocated by Programme and administered by department. There is no explicit mechanism for shared participation in meetings or inter-programme projects. Thus, these activities can only be pursued with funds collected from a number of, often unrelated, budget lines.
- Coordination is a low priority. Cross programme or department coordination has been considered a low priority within the Secretariat. Key issues, such as development of standards or guidance material are normally fractured into programmes/commissions with little consideration of its possible relevance to other Programmes.

Specific recommendations

k. Establish a data management coordination team within the Secretariat. Members of this team, comprising representatives of all of the scientific and technical departments, would be responsible for ensuring the activities within their own department are coordinated with similar activities in other departments. The team should also develop and oversee plans for inter-programme activities to accomplish common goals. It would be worthwhile for the team to meet once every few months to discuss ongoing activities and accomplishments. However, recognizing that conflicting schedules would make it difficult to hold regular meetings, once the team were established it might be sufficient to exchange information via an e-mail mailing list.

l. Budget support for inter-commission work must be made explicit. A formal mechanism for funding cross-programme activities in data management should be defined. The preferred option is for Executive Council to establish, approve and commit funds to a crosscutting programme in data management. There are several models that could be used to manage this programme. It could be managed by dedicated staff within an existing department, an independent office could be created under the Deputy Secretary General, as was done with the WMO Space Programme, or it could be matrix-managed by staff from several departments, perhaps comprising the coordination team recommended above. If it is not possible to create a formal crosscutting programme then each department should be directed to establish a line within its budget to support this work.

m. Improve feedback from experts to the Secretariat. Experts attending inter-commission task teams as representatives of their Programme or Commission often do not provide sufficient feedback of the results to the responsible staff in the Secretariat (or their Commission). These experts should be required to submit reports on their participation, which should include the results of the meeting, with emphasis on the requirements and concerns of their particular Programme and if or how they have been addressed.

n. Improve circulation of meeting reports. By improving the dissemination of the most pertinent results of meetings, communication of important activities could be improved at little cost. All meeting reports should include a concise (less than one page) summary of their most important results. Concrete results, such as proposed standards, are particularly important. These summaries should be distributed to the directors and chiefs of all of the scientific and technical departments. All reports should be accessible via the Web and links to these reports should be prominent and obvious from the front page of each department's Web site. It would probably be worthwhile to standardize the links to this information on each department's home page.

7.2. Within and between Technical Commissions

Problems identified

- Poor communication. There has been poor communication of results at the working level between projects within a Commission and between Commissions. There has also been poor feedback from representatives of joint meetings to their presidents. Related to this, there has been particularly poor communication and coordination between operational and research programmes, which represent different communities, each with their own priorities, requirements and working styles.
- Reporting responsibilities. The progress and results of inter-commission projects need to be reported to Executive Council and/or Congress. To date, this has been accomplished by assigning a single Commission to lead the effort and report on its progress. This arrangement has not always been optimal, since the designation of a lead Commission can have an adverse affect on the commitment of other Commissions to the effort.
- Inappropriate expertise at meetings. Within WMO there is a general problem of people attending meetings as experts who do not have the appropriate expertise. For this study it is more important to note that experts nominated to represent a Programme often are not prepared to perform this role and instead represent only themselves or their NMHS.
- Competing or incompatible requirements. Within projects that involve significant inter-commission participation, there has sometime been some difficulty in balancing competing or incompatible requirements. In these cases the lead Commission often prevails and others feel that their needs are considered of only secondary importance. This can have a serious impact on the level of commitment to the joint project.
- Joint projects are perceived to be slow and unwieldy. The Future WMO Information System is arguably the single most important inter-commission activity currently being pursued within the field of data management. In discussing FWIS with representatives of the various Programmes it became evident that there is widespread dissatisfaction with its rate of progress. A variety of possible reasons for the slow progress of FWIS have been suggested. These include insufficient commitment of staff or resources, poor oversight, poor accountability, and overly broad and unfocused objectives. Regardless of the reasons, it was noted that information management technology evolves very quickly and if projects in this area cannot be implemented quickly they cannot be effective. Unless rapid progress can be made within a joint project, individual, loosely coordinated projects will undoubtedly take its place.
- Poor utilization of data. It has been very difficult for Programmes to find, acquire and use pertinent data from other Programmes. There is no central clearing house for locating the data, there is a plethora or formats used to exchange data, and once data are acquired they are often accompanied by poor documentation or metadata. Restrictive data policies, high costs, and a shortage of users with experience with data from outside their Programme exacerbate the problem.

Specific recommendations

o. Many issues affecting more than one Commission can be resolved using the current mechanism of "weakly inter-commission" teams. That is, those teams that are established within a Commission but with invited representatives of other relevant Commissions. From this point onward within this report these will be referred to as "inter-programme teams". To ensure the most effective experts are nominated, the attendance of the invited representatives should be funded by the Commission he/she represents. Since it is difficult to anticipate all opportunities for such inter-programme teams, every Programme should include some funds within its budget to support its participation in such efforts.

p. Some critical issues can best be addressed by "independent inter-commission" teams. In order to be most effective, these teams need to truly independent of any particular Commission. Membership of these teams should be based on the expertise needed with equal weight given to the requirements of all participating groups. Otherwise there is a perception that the views of one Commission (usually the designated lead) are accorded higher priority than the views of the others. If a crosscutting programme in data management were established (as suggested in recommendation b above), this programme would be responsible for the funding and management of such independent teams related to data management. If an umbrella programme is not established, then each inter-commission team should be directly funded by and report to Executive Council. Suggested mechanisms to manage such teams are discussed under recommendation b above. Since the participation in these teams will be supported by a dedicated budget, the team leader or responsible officer in the Secretariat should have a significant role in confirming team members.

q. Inter-agency, inter-organization coordination or data management activities can provide important benefits. Ocean data management, where JCOMM facilitates coordination with IOC, is a good example. Other Programmes within WMO that could benefit from further coordination with outside organizations include hydrology, agricultural meteorology and the World Climate and Weather Research Programmes. Officers responsible for these activities should reach out to other relevant organizations to investigate possible collaborative arrangements pertaining to data or information management.

r. Currently, FWIS is perceived as catering primarily to the requirements of CBS, while at the same time, moving too slowly to address pressing requirements of other Programmes. An independent inter-commission team, with representation from all of the players on an equal basis, should refocus further development and implementation of FWIS. The team should review the various components of FWIS and concentrate further efforts on a few specific and achievable objectives. Some users have suggested that highest priority should be given to developing an effective means to help users to find, acquire and use pertinent data from all WMO Programmes.

s. Continue to hold informal meetings of the Presidents of Technical Commissions. These meetings provide an effective forum for the presidents to exchange views and to keep abreast of significant developments and activities within the other Commissions. Reports of these meetings should be posted on the Web so that experts not in attendance might benefit from this same exchange of information.

t. Experts representing a Commission should be made explicitly aware of their responsibilities. Experts attending inter-programme or independent inter-commission task teams as representatives of their Programme or Commission should be reminded of their responsibilities in this capacity. Before attending the meeting they should be briefed of the Commission's position on pertinent topics and should be given the authority to speak on the behalf of his/her president when necessary. Furthermore, effectiveness is enhanced when there is continuity in the experts chosen to attend follow-on meetings of the same group. These experts should submit reports on their participation to their president noting the requirements and concerns of their Commission and if or how they have been addressed by the meeting.

REFERENCES

1. Aircraft Meteorological Data Relay (AMDAR) Reference Manual (WMO-958), 2003
2. Application and Presentation Layer Specifications for the LRIT/LRPT/HRIT/HRPT Data Format, SAT-19
3. Best Guide and Principles Manual for the Ships of Opportunity Program (SOOP) and Expendable Bathythermograph (XBT) Operations, 2001
4. GCOS Data and Information Management Plan, Version 1.0 (GCOS-13), 1995
5. Global Atmosphere Watch Measurements Guide (TD-1073)
6. GRDC Policy Guidelines for the Acquisition, Dissemination and Costing of Data, 1997
7. GTSPP Real-Time Quality Control Manual
8. Guidance on Climate Observation Networks and Systems,Version 4, 7 July 2003
9. Guide to Agricultural Meteorological Practices (WMO-134), 1981
10. Guide on the Automation of Data-processing Centres (WMO-636)
11. Guide to Climatological Practices (WMO-100), draft 2001
12. Guide on the GDPS (WMO-305)
13. Guide to Hydrological Practices (WMO-168)
14. Guide to Marine Meteorological Services (WMO-471)
15. Guide on Meteorological Observation and Information Distribution Systems at Aerodromes (WMO-731)
16. Guide to Moored Buoys and other Ocean Data Acquisition Systems (WMO-750)
17. Guide on WMO Binary Code Forms (TD-611), 1994
18. Guide to WMO Table-Driven Code Forms: FM 94 BUFR and FM 95 CREX, Layers 1, 2 and 3, 2002
19. Guide on World Weather Watch data management (WMO-788), 1992
20. Guidelines and Standards for the Identification, Preparation and Implementation of HYCOS Projects within the Framework of WHYCOS, second draft, 2001
21. Guidelines for computerized data processing in operational hydrology and land water management, (WMO-634)
22. HOMS Reference Manual, 2000
23. Hydrological data transmission (WMO-559)
24. Information on Meteorological and other Environmental Satellites (WMO-411), 2002
25. IOC/IODE “Manual of Quality Control Algorithms and Procedures for Oceanographic Data Going into International Oceanographic Data Exchange”, 1989.
26. IOC Manual **&** Guides **#3** - Guide to operational Procedures for the Collection and Exchange of IGOSS (now JCOMM) Data, Third Revised Edition, 1999
27. Joint Data and Information Management Plan, Version 1.0 (GCOS-60), 2000
28. Manual on the GTS (WMO-386)
29. Manual on Codes (WMO-305)
30. Manual on the GDPS (WMO-485)
31. Manual on the GOS (WMO-544)
32. Manual on Marine Meteorological Services (WMO-558)
33. Proceedings of the WCRP Joint Scientific Committee, 24th session, 2003
34. Report of the second Inter-programme Data Management Coordination Meeting, 1995
35. Report of the third Inter-programme Data Management Coordination Meeting, 1999
36. Report of the GCOS/GTOS/HWRP Expert Meeting on Hydrological Data for Global Studies, 2002 (GCOS-84)

37. Report of the GTN-H Coordination Panel Meeting, 2002 (GCOS-83)
38. Report of the first meeting of the Inter-Programme Task Team on Future WMO Information Systems, 1999
39. Report of the second meeting of the Inter-Programme Task Team on Future WMO Information Systems, 2000
40. Report of the third meeting of the Inter-Programme Task Team on Future WMO Information Systems, 2001
41. Report of the fourth meeting of the Inter-Programme Task Team on Future WMO Information Systems, 2002
42. Report of the Expert Team on Integrated Data Management, 2001
43. Report of the second meeting of the Expert Team on Integrated Data Management, 2002
44. Report of the first session of the JCOMM/IODE Expert Team on Data Management Practices, 2003
45. Report of the WMO-NOAA Expert Meeting on GAW Data Acquisition and Archiving (TD-755), 1995
46. Roving Seminar on Data Management for Applications to Agriculture, 1999
47. Satellite Data Archiving, SAT-14, (WMO TD-909), 1998
48. Second Report on the Adequacy of the Global Observing Systems for Climate in Support of the UNFCC (GCOS-82), 2003
49. Software for Agroclimatic Data Management, 2000
50. Strategy for the Implementation of the Global Atmosphere Watch Programme (2001-2007) (TD-1077)
51. The WMO/GAW Programme, information for CG XIV, May 2003
52. Updated Guidelines for Atmospheric Trace Gas Data Management (WMO TD-1149), 2003
53. WMO Guide on Internet Practices, 2000
54. WMO Technical Regulations (WMO-49)

LIST OF ACRONYMS

ACSYS	Arctic Climate System Study
ADIS	ACSYS data information service
ADM	Alternative dissemination methods
ANSI	American National Standards Institute
ASAP	Automated Shipboard Aerological Program
AWG	CBS Advisory Working Group
AWS	Automatic weather station
CAeM	Commission for Aeronautical Meteorology
CAgM	Commission for Agricultural Meteorology
CBS	Commission for Basic Systems
CCGG	Carbon cycle greenhouse gases
CCl	Commission for Climatology
CDMS	Climate data management system
CEOS	Committee on Earth Observation Satellites
CGMS	Co-ordination Group for Meteorological Satellites
CHy	Commission for Hydrology
CIMO	Commission for Instruments and Methods of Observation
CLICOM	Climate computing
CLIPS	Climate Information and Predictions Services
CLIVAR	Climate Variability Programme
CMD	Continuously managed database
CMDL	Climate Monitoring and Diagnostic Laboratory
CREX	Character Representation form for data Exchange
DAC	Data Assembly Centre (WOCE)
DIF	Directory Interchange Format
DMIP	Data management and information panel
DMWP	Data Management Working Group (GEWEX)
DWD	Deutscher Wetterdienst (German Weather Office)
EC	Executive Council of the WMO
ECMWF	European Centre for Medium Range Weather Forecasts
EOSDIS	Earth Observation System Data and Information System
ET	Expert team (of CBS)
FGDC	Federal Geographic Data Committee
FTP	File Transfer Protocol
FWIS	Future WMO information system
GAW	Global Atmosphere Watch
GCC	Global collection centre
GCOS	Global Climate Observing System
GDPS	Global Data Processing System
GEWEX	Global Energy and Water Cycle Experiment
GHP	GEWEX Hydrometeorology Panel
GIS	Geographic Information System
GOOS	Global Ocean Observing System
GOS	Global Observing System
GOSIC	Global Observing Systems Information Centre
GRDC	Global Runoff Data Centre
GSN	GCOS Surface Network
GSNMC	GSN monitoring centre
GTOS	Global Terrestrial Observing System
GTS	Global Telecommunications System
GTSPP	Global Temperature-Salinity Profile Programme
GUAN	GCOS Upper Air Network
G3OS	Global Climate, Ocean and Terrestrial Observing Systems
HDF	Hierarchical data format

HNRC	HOMS national reference centre
HOMS	Hydrological Operational Multipurpose System
HWR	Hydrology and Water Resources Programme
HYCOS	Hydrological cycle observing systems
ICPO	International CLIVAR Project Office
ICSU	International Council of Scientific Unions
ICT	Implementation/coordination team
IDD	Internet data distribution system (Unidata)
IGBP	International Geosphere-Biosphere Program
IGBP-DIS	IGBP data and information system
IMMT	International Maritime Meteorological Tape
IMTN	Improved Main Telecommunication Network project
INFOHYDRO	Hydrological Information Referral Service
IODE	International Oceanographic Data and Information Exchange
ISO	International Standards Organization
JCOMM	Joint WMO/IOC Technical Commission for Oceanography and Marine Meteorology
MCSS	Marine Climatological Summaries Scheme
MTN	Main Telecommunications Network (of the GTS)
MSS	Message switch system
NASA	National Aeronautics and Space Administration (USA)
NCAR	National Center for Atmospheric Research
NCDC	National Climatic Data Center
NCSA	National Center for Supercomputing Applications
NetCDF	Network Common Data Form
NMHS	National meteorological or hydrological service
NMS	National meteorological service
NWP	Numerical weather prediction
OGC	Open GIS Consortium
QA	Quality assurance
QC	Quality control
RBSN	Regional Basic Synoptic Network
RTH	Regional telecommunications hub
SAC	Special Analysis Centre (WOCE)
SAG	Scientific advisory group (GAW)
SPARC	Stratospheric processes and their role in climate
SSUP	Satellite systems utilization and products
TCP/IP	Transport Control Protocol, Internet Protocol
UCAR	University Corporation for Atmospheric Research
VOS	Voluntary observing ship
VOSClim	VOS climate subset project
WAFS	World Area Forecast System
WCASP	World Climate Applications and Services Programme
WCDMP	World Climate Data and Monitoring Programme.
WCRP	World Climate Research Programme
WDC	World data centre
WHYCOS	World Hydrological Cycle Observing System
WOCE	World Ocean Circulation Experiment
WMO	World Meteorological Organization
WWW	World Weather Watch
XML	Extensible mark-up language