Global
Ocean
Observing
System

Intergovernmental Oceanographic Commission

Data and Information Management Strategy and Plan of the Global Ocean Observing System (GOOS)

June 2001

GOOS Report No. 103
IOC/INF-1168

UNESCO

Intergovernmental Oceanographic Commission

Data and Information Management Strategy and Plan of the Global Ocean Observing System (GOOS)

June 2001

Edited by

Dr John Ronald Wilson
Manotick
Ontario, Canada

GOOS Report No. 103
IOC/INF-1168

UNESCO 2002

IOC/INF-1168
Paris, 14 February 2002
English only

(SC-2002/WS/9)

Table of Contents

Executive Summary

The GOOS Data and Information Management System (DIMS) will be a highly distributed system based on contributions by operational agencies, data centres, and research organizations in the oceanographic and meteorological communities. As a measure of its diversity, the various components of the observing system now under development have different data and information management strategies encompassing physical, chemical, and biological observations made in situ and from satellites.

The purpose of this document is to provide basic design information, guiding principles, typical responsibilities for data and information centres, and a strategy to be used by the scientific and technical panels and data managers in planning the development and implementation of the data and information systems for GOOS. It contains practical guidance material and highlights examples of successful systems that now exist. It provides a concept for an on-line single point of entry for data and information and a concept for monitoring data flows and the success of GOOS in meeting its data management goals.

The Strategy and Plan will be read and used by persons in organizations not familiar with international data management and needs to be relatively complete in coverage.

Implementation of GOOS data and information management will be accomplished in an iterative fashion by connecting existing ocean observing programmes that contribute to GOOS requirements. The present implementation is called the Initial Observing System (IOS). The needs for additional data and improved analyses and products are being identified by the science panels as GOOS is further developed. Future data management requirements cannot be foreseen in detail at the present time. However the general characteristics of the systems can be foreseen and are described here.

Because of the diversity and global distribution of the data collection programmes and the expertise to carry out the data management and produce the relevant analyses, the system must be highly distributed and evolving.

Practical implementation of GOOS data and information management will begin by rationalizing and modifying the operation of the elements of the IOS, where necessary, to conform to the guidance, standards, and strategies of this plan. Then as GOOS is further developed, mostly through the vehicle of pilot projects, the data management requirements will be developed in parallel with the new observation networks and with full understanding of the specific details of the requirements.

This document begins (section 2) with a discussion of ten general categories of applications to be served by GOOS, ranging from climate forecasting to fisheries productivity. Each category is described in terms of a common set of headings described in Annex D. The headings include such items as inputs, input delivery requirements, quality control/assembly requirements, etc. Following the presentation of the ten categories there is an analysis of the general requirements and a compilation of the common features and guiding principles that can be foreseen. (section 3)

The plan also provides an overview of the status of development of the GOOS IOS as of publication of this document. (section 4) Since the IOS will be under continuing development as other GOOS services and programmes are implemented the latest information will be maintained on-line in the Global Observing Systems Information Center (GOSIC).

In addition to actually managing and moving the data and information, there are two other important functions that must be included in the design of the GOOS DIMS. There is a need to connect these programmes and the participating data and science centres under a unified and centralized information services system, where information about the programmes and observations may be obtained from a single source and where access to the data holdings or holder is provided. This function is provided by GOSIC. (section 5.1)

The second important function is a carefully designed, automated (or at least semi automated) tracking system for the data and information so that it can be demonstrated that the system is working or if not where the problems lie. This plan contains strategies to address these needs. In particular the idea of data coordinators supported by an on-line information system is presented. (section 5.3)

There is a need for cooperation with other international programmes in implementing GOOS data management systems, in particular with the data management programmes of IOC and WMO. There are benefits in partnering with other programmes and in development of data management systems through the GOOS pilot projects. (section 6) Development of archiving practices and standards in cooperation with the ICSU World Data Centre System and the IOC Committee on International Oceanographic Data and Information Exchange (IODE) will make the GOOS data and information available for future scientific and engineering needs. (section 7)

The Plan and Strategy proposes a cooperative and focussed approach to capacity building. In particular the development of underpinning infrastructure in the form of national GOOS coordinating committees and national oceanographic data centres is seen as pivotal. Various strategies for a focussed, effective approach include building programmes around identified projects with all interested national and international agencies involved. (section 8)

The way forward for further development of the GOOS data and information management systems is seen as focussed on the IOS and the pilot projects. In addition a GOOS initiative for the development of necessary advanced data management technologies has been approved by the GOOS Steering Committee. (section 9)

The plan will be kept up to date as GOOS develops. The latest version will be available on-line.

1. Introduction

There are three general classes of user for GOOS data and information. The three classes include

- operational users in the ten categories identified in section 2 below;
- managers and developers of the national and international data collection and information dissemination programmes; and
- scientists, engineers, and economists doing special research, strategic design studies, and other studies to advance the application and usefulness of GOOS data and information.
-

More specifically GOOS data, products, and services will be used by or for;

- Intergovernmental conventions;
- Government agencies, regulators, public health, certification agencies;
- Environmental management, wildlife protection, amenities, marine parks;
- Operating agencies, services, safety, navigation, ports, pilotage, search, rescue;
- Small companies; fish farming; trawler skippers, hotel owners, recreation managers;
- Large companies, offshore oil and gas, survey companies, shipping lines, fisheries, dredging, construction;
- The single user, tourist, yachtsman, surfer, fisherman, scuba diver;
- Scientific researchers in public and private institutions.

The operational users have a responsibility to their clients to analyze data that have been collected and produce a prediction about weather or climate; issue a warning of an impending condition such as a severe storm at sea or a coastal storm surge; or implement a regulation such as the closure of a fishery for a specific health danger. These users will have time critical applications requiring data collection and distribution in an operational time frame.

For managers of the data collection programmes, it is information that is important rather than the data itself. The managers must be able to evaluate progress against plans on a continuing basis to ensure the programme is succeeding in collecting its data. The managers must be able to identify gaps in observations before they can damage the success of the programme. They must know that the quality of the measurements are meeting the standards set for the programme and that the analyses and data products are being produced and distributed and meet the needs of the clients. Thus this class of user needs a broad range of timely information to manage the collection programme to meet its goals.

The scientists, engineers, and social scientists will have a requirement for accurate, long term data sets for research into physical, biological, and chemical oceanographic processes; model development and testing; design criteria for ships, structures, and marine facilities; studies of the effects of climate change on economies and populations, etc. For these types of work, accuracy and completeness of the data sets are more important than having the data in real or near real time. Many of these studies may be doing work not foreseen in the design of GOOS and for which specific data collection has not been implemented. However, such studies may well lead to other application to be incorporated into the GOOS programme at a later date.

The end-to-end data management systems that are imp lemented for GOOS must have the flexibility to serve this large variety of requirements. The development of a sufficiently comprehensive strategy and plan to meet this variety of requirements begins with a study of the general characteristics of the applications to be served. An analysis of these general categories of applications and of the existing successful programmes that will form the basis of the GOOS IOS was used to specify guiding principles and common characteristics of the GOOS data and information systems of the future. The Plan and Strategy also discusses provision of access to the data, data products, and services; cooperation with other relevant national and international data programmes; archiving strategies and standards; capacity building; and the way forward for implementation of the GOOS D&IMsystems.

2. General Categories of Applications Served by GOOS

In order to understand the requirements of the GOOS D&IM system, the general data management characteristics have been defined for ten categories of applications that will be addressed by GOOS. These categories are

- Operational marine coastal and ocean short range forecasting and analyses
- Seasonal-to-interannual climate prediction
- Numerical weather prediction
- High-quality procedures for climate products
- Biodiversity and habitat
- Natural and man made hazards
- Environmental indices
- Fisheries productivity
- Satellite data processing systems
- Regional, integrated systems

Ten categories were judged to cover all the applications that could be foreseen at this time while keeping the number manageable. This section provides the D&IM equivalent of the scientific objectives for the two main GOOS themes represented by the Ocean Observations Panel for Climate (OOPC) on the one hand, and the Coastal Ocean Observations Panel (COOP) on the other hand. Annex D contains an explanation of the headings used to describe the applications in the following sub-sections.

Section 3.1 contains a general analysis of the requirement and a preliminary evaluation of the state of the existing systems relative to that requirement.

2.1 Operational Marine Coastal and Ocean Short Range Forecasting and Analyses

General Overview of the Requirement

Operational coastal and ocean short-range forecasting in the 0 to 10 day range is for protection of human life, protection of property, protection of the environmental, and for economy of operations. The forecasts and warnings are usually provided by National Meteorological or Oceanographic Services using data collected by buoys, ships and satellites and reported in real time on electronic networks. GOOS participation is to provide ocean surface and sub-surface data to the networks in real time, assist in the implementation of operational forecasting systems in Member States of IOC that do not have such systems, and ensure the maintenance of the data and information in databases for future research and strategic planning. Examples of ocean forecasts include storm surges, harmful algal blooms, sea-ice distribution, and storm tracks for ship routing.

Products, Product Delivery, and Product-Service Chain Characteristics

Data and information products that are required begin with data reported in real time or near-real time on computer networks to centres that perform quality control and produce the relevant analyses. The analyses are in the form of synoptic global or regional representations of the physical parameters of the ocean. These analyses are either examined directly or more often serve as input to other local and regional computer models that make the predictions from which the forecasts and/or warning are prepared and issued.

The service chain involves data collected in real and near real time flowing to national weather services and to short range forecast centres (which may also be in the national weather service or oceanographic institutions). Products in the form of forcing fields will flow from NWP activities to short range forecasting centres for input to the models for short range prediction. Products from these models flow to users. There is increasing emphasis on also producing error estimates.

Note that in addition to the physical data, ocean colour is beginning to be used in operational products and services.

Character of the Inputs

Inputs are focussed on the principal physical variables of surface and subsurface temperature, wind velocity, ocean colour, and sea level with an emerging need for salinity in some regions (e.g., Norwegian coastal current).

Input Delivery Requirements

The delivery time cut-off is short (12 to ~ 60 hours). There is however an emphasis on providing those data also used for numerical weather prediction in real time for use by the NWP centres. Cooperation with the NWP community is essential for acquisition of the surface forcing fields for the short range forecasting models.

Quality Control/Assembly Requirements

Data assembly requires the integration of data from several sources including in situ observations from ships, and moored and drifting buoys; and from satellites. Quality control of this real and near real time data must be highly automated because of the few hours to 2 to 3 day cutoffs. There is time to acquire data that did not make the meteorological cutoffs or is available in near real time from other networks to augment the data sets used for NWP. Thus there will be a second task of integration of data from these other sources and from a similar variety of observation platforms.

Character of the Processing

Processing is generally automated and the QC is of a modest standard because of the short cutoffs. Models and data assimilation are usually involved and range from simple to complex. The products that are generated need to be available in electronic form for distribution, although in some cases the results may be distributed as ASCII text or by radio broadcast.

The Level of Feedback

The feedback from product users to data originator is not well developed. Formal evaluations of the needs of the users, the relevancy of the product to those needs, and the accuracy of these forecasts are not yet routinely produced on a broad basis for many of the products.

State of Permanent Archives for Data and Products

There are established archives for much of the data that generated the short-term forecasts and many of the products that are produced. However these archives are not necessarily comprehensive or standardized. They are often maintained by the organizations that generated the products and may not be readily available, or up-to-date. The organizations may not have accepted a long-term responsibility to maintain the archives and provide services to users from them. At the same time it must be said that the requirements for archival of these data and products have been defined. Much if not all of the data will have been archived elsewhere. Is it sensible to archive wave forecasts for a given region when the forecasts are done using a variety of national models? There is a need to discuss and define the archival needs and policies.

Strengths and Building Blocks

A major strength for this type of service is that IOC and WMO have been working to develop the basic infrastructure for these types of services for a number of years. This infrastructure will need some improvements but a good starting point for these additional GOOS applications has been established.

Building blocks include:

- The marine systems of the WWW, JCOMM, and IODE for data management.
- Systems emerging from operational meteorological and oceanographic centres, both national and regional.
- GODAE and its data pilot projects and systems.

Weaknesses and Needed Capacity

There are few prescribed standards for these products. Many have been developed locally and could be improved by adoption of the latest knowledge and techniques. Validation techniques are relatively undeveloped and are mostly based on statistics or case studies. Most of the sophisticated short range forecasting operates in developed countries. There is a significant need for capacity building in this area in developing countries.

Weaknesses also include the inadequacy of existing communications and protocols for newer data types and for rapid, automated exchange of the data to all types of centres involved in the generation of the forecasts and to the users themselves. There is also a problem in that the modeling and assimilation systems in operation are often not as mature as those in the research community.

Opportunities

There is an opportunity to improve the situation with the formation of the JCOMM with its focus on a significant increase in cooperation between the meteorological and oceanographic communities and on improved communications with the users as part of building end-to-end systems.

2.2 Seasonal-to-Interannual Climate Prediction

General Overview of the Requirement

The first and most familiar examples of seasonal to interannual climate prediction are the forecasts for the El Nino/La Nina phenomena. Some skill in forecasting has been demonstrated in recent years by some of the models. The models forecast the SST fields. The SST fields are then used to infer climatic conditions by various means. The means can be as simple as basing climate forecasts on similar situations that have occurred in the past.

In general this type of prediction is in an early stage. In particular the use of the predicted state of the ocean to then predict the resulting local climate conditions has only been demonstrated for only a few applications. Much more work is required to broaden the applications for the predicted SST fields. However it is important to have operational systems in place to support the operational predictions being made so that both the predictions and application of the predictions can be further developed in a realistic scenario.

Products, Product Delivery, and Product-Service Chain Characteristics

The products are normally forecasts for climate variations with SST as the principal field. Value-added products include inferred precipitation and the generation of inputs to agricultural and energy sector models.

The service chain involves data flows at several time scales. Products and data sets may be acquired on a near real time basis for evaluation of predictions to detect when the models get off track. Other data that was not available in real time will be acquired and assembled to be used in developing the initialization data set for the next cycle of the model. Products will include evaluations of model performance as well as the ones listed above.

Character of the Inputs

Inputs are focussed on the "standard" physical variables, mostly ocean temperature and sea level. There are emerging requirements for salinity, short wave radiation and ocean colour and transparency. The data are observed by ships, moored and drifting buoys, and by satellites. A significant portion of the in situ data is collected as part of international experiments such as WOCE and CLIVAR. The real time data from ships and buoys for both operational and research requirements are handled mostly by the operational systems established by IOC and WMO with assistance from the research organizations for special analyses and quality control.

Input Delivery Requirements

Input delivery is a problem of assembling the data from these diverse sources, which report through different agencies, and at different time scales from real time to months after the observation. Users need to be able to acquire the most up-to-date data set at time scales from days to months after collection. The data cut-off is thus not demanding allowing for acquisition of data that are not transmitted in real time to augment the data sets.

Quality Control/Assembly Requirements

Data assembly is complex due to the diversity of the data types and the fact it is widely distributed among collecting organizations. Satellite and in situ data need to be merged.

Quick-look quality control with some manual intervention, duplicates removal, comparisons to climatology, and/or pre-processing of some inputs is used to provide an internally consistent data set. As an example of the complexity and diversity of the systems value added products are now provided by the GTSPP, the Hawaii Sea Level Centre, Topex/Poseidon, Florida State University, and others.

Character of the Processing

Some processing systems are automated (e.g., BMRC, ECMWF) though the level of sophistication (in models, assimilation, QC, etc.) is modest. The spin-up toward the initial conditions remains an issue for assimilation and coupled model initialization.

The Level of Feedback

The operational systems of IOC and WMO provide feedback on data problems to data originators on a regular basis for the data reported in real time and some delayed mode data.

The feedback from products (not value-added) to data originators is also good, mostly because the research community is involved. Validation remains an issue. Testing is also problematic because of the lack of long records.

State of Permanent Archives for Data and Products

The permanent archives for the data required for seasonal to interannual climate prediction are relatively well developed and have been operational for many years. In situ data are archived in the systems operated by the IOC and WMO. Satellite data are archived by NASA, NOAA and other space agencies.

The status of the archival of products is less well developed. Some relevant data sets and products have been published on CD-ROM. However a review needs to be conducted to define comprehensive archival needs and ensure the continued availability of the data products.

Strengths and Building Blocks

Strengths include the relatively long history of international data management and exchange that has been developed by IOC and WMO. This has led to needed standardization and a robust system that can respond to the growing needs of the GOOS programme.

Building blocks include the systems of JCOMM, IODE, and those developed in the research programmes TOGA, CLIVAR, and WOCE. It should be noted that the support of organizations in the research community to the management of the operational data will be required for years to come.

Weaknesses and Needed Capacity

Weaknesses include the lack of connectivity between the impacts sector, the observers and the modellers. In other words the end-to-end process has not been adequately implemented. However as noted these applications are in an early stage of development and this connectivity has not had a chance to develop.

Another weakness is the limited ability of the existing systems to deal with biological and chemical variables and with some coastal physical variables that will be needed by the Coastal Ocean Observations Panel (COOP). The existing systems should begin to address the requirements for these other variables at an early date.

Opportunities

The prediction of the El Nino Southern Oscillation (ENSO) is at present the most visible example of a seasonal to interannual system for climate prediction. The system has a well supported operational observing system, a high level of research and indeed public interest, and the basic elements of an operational prediction system. It can be used to demonstrate the feasibility and utility of such climate predictions and the fact that we have reached the point that we have the knowledge and technology to begin making such useful predictions.

The ENSO systems also provides the opportunity to learn from experience in the development of such operational systems.

2.3 Numerical Weather Prediction (NWP)

General Overview of the Requirement

NWP models are used by national, regional, or global centres to predict the weather over the next one to five days. The models generally run on a 6-hour schedule. Thus the data management system must move the observations from the collection point to the weather prediction centre in minutes to a very few hours using electronic networks. These centres are operational and ingesting and processing data 24 hours per day, 7 days per week. Quality control is usually left to the weather centres although some data types are beginning to have automated QC or even some subjective QC applied before the data are supplied to the weather centres.

GOOS does not have any direct responsibility for NWP. However GOOS does have an interest in the NWP data and a responsibility to contribute relevant surface and subsurface data to improve the database for NWP. The data management system should supply these data to the weather networks in real or very near-real time. Secondly much of the data used for NWP are only available in real time or will not be available except in the real time stream for months to years. Some GOOS systems will therefore have to capture at least parts of the real time stream for such things as seasonal to inter-annual forecasting. Also some of the NWP model outputs will have to be used for such things as boundary conditions for regional and local marine and ocean short-term forecasting.

The GTSPP provides an example of capture of the real time stream for other applications.

Products, Product Delivery, and Product-Service Chain Characteristics

The types of products needed for GOOS purposes are

- data sets of observed data that are assembled by the weather prediction centres with duplicates removed.
- global or regional model fields in digital form for various areas of the oceans.
- supporting meta data on data quality including changes made to observations.

The service chain once again features World Weather Watch, GOOS, and other real or near-real time data flowing in a timescale of hours to the NWP and short range forecasting centres for assembly, QC, and issuance of forecasts and warnings to users. Data sets and model fields from the NWP are then made available for other GOOS applications and their users as required.

Character of the Inputs

The input data of use to the weather services include sea surface pressure, temperature, and salinity, wind velocity, precipitation, and subsurface temperature and salinity observed by instruments that meet the standards approved for such meteorological instruments for WMO.

Input Delivery Requirements

The data must be delivered on the global weather circuits in a time frame to satisfy the meteorological cut-offs for such data. These cut-offs vary from 3 hours before the model cycle begins for pressure, winds, etc and up to 30 days for ocean sub-surface data.

Quality Control/Assembly Requirements

The weather services generally carry out the quality control of the data with automated checks for possible values, comparison to first guess fields and near neighbour observations, and some subjective checking of computed fields.

For some types of data some quality checking is done at source as long as it does not delay data delivery beyond the cutoff.

Character of the Processing

Processing at the weather centres includes assembly, quality control, and analysis using numerical prediction models.

The Level of Feedback

Feedback of problems with data quality and timeliness is routine. Certain operational centres are designated as "lead centres" for various types of data and pass information on persistent lateness of data or quality problems to the collectors of the data and to the WMO Secretariat for action.

State of Permanent Archives for Data and Products

The WMO has a variety of permanent archives for the atmospheric and marine data. There are WDCs for Meteorology in the USA and the USSR. A number of national weather centres participate in the Distributed Databases (DDBs) system of WMO. There is also a scheme for maintaining archives of marine climatological data (Marine Climatological Summaries Scheme). Some ocean surface and most ocean subsurface data are managed in the archives of the IOC RNODCs and the ICSU WDCs. The GCOS has established a number of land-based stations that collect and archive high quality observations for detection and evaluation of climate change. These archives will likely be of interest to the Coastal Ocean Observations Panel of GOOS.

Strengths and Building Blocks

The NWP activity already exists. It has been built from a network of national weather services in member countries of WMO over the past several decades.

Weaknesses and Needed Capacity

Weaknesses and needed capacity for NWP are not subjects that can be addressed by GOOS bodies. The main problems are related to cut backs in nations that are Members of WMO resulting in less data for NWP. GOOS should make special efforts to deliver copies of any relevant data that is being collected to the GTS before the meteorological cutoffs for use in NWP.

Opportunities

The major opportunity here is for GOOS and the NWP community to share data and products to the benefit of both. There will of course need to be some tuning of systems and outputs to accommodate the other. Progress in understanding ocean-atmosphere processes can in turn result in improved results for NWP.

2.4 High-Quality Procedures for Climate Products

General Overview of the Requirement

There are a number of products that need to be generated on a continuing basis to detect climate change, to evaluate the magnitude and rate of change, and then to determine the performance of the models and other analysis procedures that predict future behaviour. These products will have to be based on carefully calibrated data sets collected and managed according to high standards. Examples include sea level change, sea surface temperature change, heat storage and transport characteristics, carbon budgets, etc.

Systems managing these types of data will have to be more carefully designed to identify suspect data and prevent errors creeping in. Meta data will be extremely important. Archival will be more challenging because of a need to continually update values that are found to be suspect or in error.

Products will be subject to regular updating due to reanalysis, model improvements, etc. These products and data sets will be the ones that provide definitive information on climate change.

Products, Product Delivery, and Product-Service Chain Characteristics

As stated the products will include such things as sea level change, sea surface temperature change, heat storage and transport characteristics, carbon budgets, etc. The databases of high quality observations along with very complete meta data will be also be vital products of the data management system because of the need for continuing analysis and reanalysis. The more active the databases the better the quality should be as inconsistencies and errors are discovered in the preparation of products by scientific organizations.

The product-service chain will feature the development and maintenance of high quality, very well documented databases that are analyzed on a regular basis to produce climate products for users. The users will typically be organizations such as the IPCC and the end result will be evaluations of the state of the climate.

Character of the Inputs

Inputs will consist of the most accurate and highest resolution versions of the physical, chemical, and biological data that are available with the most careful quality control applied, usually by the originator. Meta data standards will have to be met for the data to be added to the databases. An audit trail of the processing and analysis of the data will have to be kept.

Input Delivery Requirements

Input delivery requirements will not be stringent, of the order of months to 2 years. This is the case where it is better to get the data later so that proper processing has been done. If re-analyses are done then delivery will be when it has been established that the quality of the new products meets requirements.

Quality Control/Assembly Requirements

Assembly includes an active procedure to seek and acquire the data collected to the required standard, conducting evaluations of the data, reformatting the data from diverse sources to first produce the high quality data sets, and then preparing the products. High standards of quality control that are well documented will be a necessary element of the data management system for these data.

Character of the Processing

Processing will include procedures for quality control analyses, gridding for models, processing and assimilation by models to generate products and to estimate errors and detect suspect data by comparison of model fields to data fields.

The Level of Feedback

Feedback will be absolutely necessary at all stages. It must be ensured that problems in data collection are identified and solved. Errors discovered in data values must be fed back to the archival centres so that the data or meta data can be corrected.

State of Permanent Archives for Data and Products

Permanent archives of data and products probably only exist at this point in a few research organizations and will be related to the work of groups such as the IPCC or the Hadley Centre. Requirements for long-term management and archival will need to be addressed as these systems develop.

Strengths and Building Blocks

Experience with some of the types of data that will need to be managed does exist in many research centres and in the IOC and WMO. However there will have to be a careful review of the requirement before these sorts of databases are further developed by the existing systems. In particular care will have to be taken not to mix the more precise data with lower resolution or less accurate data in such a way that they can not be distinguished. The meta data will have to be handled in a completely satisfactory manner.

Weaknesses and Needed Capacity

No formal examples yet exist of this type of operational data management system. It will likely develop within the research community and then move to the operational community as the application becomes routine.

Opportunities

The opportunity here is for the operational community to support the research community by improving its capabilities in dealing with the more precise data sets and with extensive meta data.

The SST project under way in the AOPC and led by the US provides an opportunity to implement operational production of a good quality SST data set from satellite and in situ sources.

2.5 Biodiversity and Habitat

General Overview of the Requirement

Biodiversity is a sensitive and important measure of the health of an ecosystem and needs to be conserved to maintain that health. As such it is important to the work of many organizations in most countries. The organizations collecting and using the data are those concerned with human health, agriculture, forestry, fisheries, environment, etc. The information holdings are vast and complex and are distributed throughout large numbers of agencies around the world. There is a large requirement for digitization and computerization of these data to make them accessible. The need for harmonization, standards, and protocols is great.

The present focus for much of the biodiversity effort is related to developing interoperable data networks, digitization of data, etc., in order to improve the availability and usefulness of the existing data. Specifications for global or regional monitoring networks for biodiversity are only now being addressed. Data collection will continue to be driven by national organizations for national purposes.

In terms of monitoring biodiversity, a specific working group in the framework of the Convention of Biological Diversity (CBD) will define key variables. A review paper 'Suggestions for Biodiversity Loss Indicators for GTOS' was prepared and distributed for review. A literature and Internet search was undertaken and some 300 indicators were found which are being reviewed, sorted and short-listed for the first meeting or the working group. For GOOS, COOP will be studying relationships between habitat, biodiversity, and living marine resources. The pilot projects Phytonet, and the Coordinated Adriatic Observing System (CAOS) will both collect data relevant to biodiversity. For example CAOS will investigate biodiversity of native species of marsh grass, submerged vascular plants, macro algae, fish, shellfish, mammals, and birds. COOP will meet GTOS in the coastal zone.

Until monitoring networks have been designed and appropriate applications and products have been developed and proven, there is no need for systems that manage data on biodiversity on an operational basis.

Products, Product Delivery, and Product-Service Chain Characteristics

The basic data for purposes of biodiversity is the identification of a species at a given location and time. Products include analyses of temporal and spatial distributions of species, population trends, and indicators of the magnitude of any temporal change and loss of diversity.

The product-service chain will be based on collection and analyses of the data into products that will be made available in delayed mode to users via the information networks being developed as described above, the Global Biodiversity Information Facility (GBIF) being one example. Other examples will include regional and national networks that join GBIF or provide access to their data by other announced arrangements. Access to biodiversity data collected by COOP will undoubtedly be available through GOSIC as well as being referenced in these other systems.

Character of the Inputs

As discussed above, the basic input data for biodiversity consists of identification of species at a given location and time. These data will be managed locally by national and international programme centres. The data bases will have to include pertinent and well defined meta data in order to document methods of identification, standards and protocols used, taxonomic codes, etc.

Input Delivery Requirements

Input delivery requirements are not time critical. There is no need for even semi-operational time frame data at this time. For submission of data and meta data and preparation of products that include old and new data, perhaps the IODE standard of one year after collection would be an appropriate goal.

Quality Control/Assembly Requirements

The most important QC element is confirmation of the correctness of the identification of the species and sub-species and inclusion of appropriate meta data that is a measure of the confidence in the identification. The data will need to be assembled into databases that permit the generation of comprehensive products that demonstrate the state and trends for biodiversity in a given place or region.

Character of the Processing

Processing is difficult because of the manual nature of identification of the species and sub-species. There will be significant problems in standardizing existing data because of the use of variety of taxonomic coding schemes that cannot be routinely mapped from one to another. The databases will have to carry significant amounts of meta data to enable the use of the data with confidence.

The Level of Feedback

Feedback and consultation/cooperation is most important on problems of misidentification, and the development of standards and protocols. Once products and services become available, there will be a requirement for evaluation and feedback of the usefulness of the products and services to ensure the right data are being collected, and the standards and protocols meet the need.

State of Permanent Archives for Data and Products

As already stated the existing archived information is not readily accessible and thus not as useful as it should be. There are a variety of standards, massive needs for digitization, needs for protocols, needs for harmonization to make data intercomparable, etc. Many of these deficiencies will have to be addressed within the present projects being implemented both nationally, regionally, and globally to create interoperable networks supported by high quality databases.

Strengths and Building Blocks

The GBIF is a strong programme that is being developed to build a distributed access to a system of distributed interoperable databases in member countries. This is one possible building block that can be used to make progress in data management for biodiversity. Specifically GBIF will

- provide coordination of new software development that will link data bases which embrace the full range of biodiversity information (including geographical, ecological, genetic and molecular data);
- assist in the organization and digitization of all biodiversity information; and
- assist in the compilation of definitive lists of species.

At least 50 biodiversity databases have agreed to affiliate with GBIF

Weaknesses and Needed Capacity

There is a strong need for digitization and computerization of data and development of standards and integrated taxonomic information systems to improve and make identification of species and sub-species comparable.

There is a problem that most data is in developed countries. There is also a strong need for capacity building to include the needs of developing countries.

Opportunities

There is at present a recognition of a need for improved information on biodiversity and several national and international programmes are being developed to address the issues, particularly those related to data and information management. At the same time the global observing systems, specifically GTOS and GOOS, are defining needs for observations that include data that can be used to monitor biodiversity. It is important to take advantage of these activities by taking steps towards standardization, and towards the coordination and development of networks and interoperable databases on biodiversity.

2.6 Natural and Man Made Hazards

General Overview of the Requirement

The decade of the 1990's has clearly demonstrated the need for warning systems for natural disasters including flooding from severe rainstorms, severe wind events, storm waves in the coastal zone, and tsunamis and storm surges. Another type of natural hazard is the harmful algal bloom or some other type of biological phenomenon or contamination that can adversely affect seafood and aquaculture operations.

Forecasts and warnings for the physical hazards of severe rain, wind, and wave events are usually issued by national weather services as part of their short-term weather forecasting (hours to days) activities. Data management for this purpose has been described in sections 2.1 and 2.3 above. Tsunami warning systems are in place in the Pacific Ocean and are under development for the Atlantic by the International Tsunami Warning System (ITSU) which operates under the auspices of the IOC. Although there is sufficient knowledge of the processes for useful predictions, forecasting of storm surges is not broadly and routinely implemented globally. In particular the timing of the surge in relation to high tide is critical and can still be a problem for accurate forecasting. This type of forecasting generally falls under the responsibility of weather services because they have the necessary parameter fields and operational capability to do the job.

Warning for other types of natural hazards such as harmful algal blooms are at an early stage. Some countries have embryo systems in operation, but much work remains to be done.

There is also the problem of man made hazards such as spills of oil or toxic chemicals and the requirements for data and predictions to deal with clean up and the effects of the spill.

Products, Product Delivery, and Product-Service Chain Characteristics

The product or service that is being delivered to a user or community is some sort of warning or advance notice of adverse conditions to come, whether it is flooding from a storm surge or conditions at an aquaculture site that could precede a harmful algal bloom. The product-service chain will involve acquisition of a variety of data, analysis and inspection of these data (usually including some modeling), development of a product and delivery of the product or a warning to users or those to be affected by the hazard.

For hazardous spills the products required include trajectory modeling and prediction, inventory and distribution of the pollutant, access to sensitivity or habitat data and information, and long-term data on effects.

Character of the Inputs

There will be a variety of data inputs including satellite and in-situ data. In many cases the hazards will be addressed under 2.1 or 2.3 above by the meteorological centres. For tsunami warnings the initial data, location and intensity, come from the earthquake monitoring network. This provides the first alert. More definitive warnings depend on a network of automated water level stations that detect sudden increases in water level and report to the warning centre in Hawaii. This system has been in operation for many years. It has not been identified as an element of GOOS.

Other systems for storm surge prediction, harmful algal blooms and contamination warnings for aquaculture and habitat, and man made hazards require a variety of in situ and satellite data to be made available in real and near real time. These requirements will be further developed in the Coastal Ocean Observations Panel.

In any case the data and products produced in the NWP and short-term forecasting activities will be important inputs for all the applications related to natural and man made hazards.

Input Delivery Requirements

Inputs for applications dealing with hazards will mostly have stringent requirements for timeliness. The time scales will mostly be similar to those for NWP. Many of the more important products will come out of the NWP and short-term forecasting chains. Warnings will have to be issued within hours or at most a very few days.

Quality Control/Assembly Requirements

Data quality control and assembly will have to be highly automated for the most part as for NWP. In some situations it will be necessary to establish additional data collection at the location of the hazard. To be able to do this instrumentation and procedures will have to be available beforehand for fast deployment. Databases with habitat information and sensitive areas will have to be available beforehand as well. This will all involve considerable pre-planning and development and will only be practical for important or sensitive fishing, aquaculture, or coastal areas.

Character of the Processing

Processing requirements will range from simple to complex. For tsunamis for example, the speed of the shock wave in deep water is known. Thus the time of expected arrival can be easily predicted. Once the nearest landfalls are reached and it is known whether the quake has generated a tsunami, warnings can be issued as to when and how severe it will be.

Other applications will require complex modeling of a physical nature, and, as knowledge develops, biological nature for effects on species and habitat.

The Level of Feedback

It is important that information on the development of the hazard be fed back to forecasters so that important characteristics that are being predicted, such as the severity or timing of the event, can be adjusted if they start to miss the mark.

State of Permanent Archives for Data and Products

It is unlikely that formal archives exist for many of these sorts of data. However it is important that as systems are implemented for dealing with hazards, the appropriate archives are defined and implemented so that the necessary research can be done to improve forecasts and warnings. Archival of data and products may also be important in the event of legal action.

Strengths and Building Blocks

The existing IOC and WMO systems as well as those established in WOCE, TOGA, and CLIVAR all have examples of operational services that can contribute to the proper implementation of systems to deal with natural and man made hazards.

Weaknesses and Needed Capacity

It is not practical to build a system or systems that can be applied to all natural and man-made hazards. It is necessary to pick the most important hazards and put in place the data collection, management, modeling and prediction infrastructure to deal with that hazard when it occurs. This has been done for some developed countries and for some of the most dangerous hazards in some developing countries. It needs to be done on a more comprehensive basis for most countries.

Cooperation between the various agencies to share data and expertise could also be improved nationally and internationally.

Opportunities

Opportunities for significantly improved cooperation or implementation of improved services are not obvious due to scarcity of resources

2.7 Environmental Indices

General Overview of the Requirement

Environmental indices are a measure of the severity of some condition that requires monitoring followed by some sort of action if the index exceeds or falls below some critical value. An example would be the significant wave height at an offshore oil production facility. Regulations may require that if the wave height exceeds a certain value, then re-supply or off loading of oil operations must be halted. Another example would be the air quality index in a city. If the amount of smog exceeds a certain value, then a warning is issued that persons with breathing problems should stay indoors.

Environmental indices are also used by countries as a means of reporting on the state of the environment. When used for this purpose the reporting frequency will usually be measured in years. The purpose of the reporting will be to show by some relatively objective number that the environment is improving with time as a result of some action that has been taken by the government.

The data management system for such an activity must collect and report the data on a regular basis and provide for notification of the proper authority when the threshold is exceeded. The authority then initiates the prescribed action.

This is a different class of activity than natural hazards. Although the two data management systems may have some common features, indices are usually associated with preventative measures. Generally speaking the use of environmental indices is in an early stage.

Products, Product Delivery, and Product-Service Chain Characteristics

One type of user for an environmental index will be a person with the responsibility to initiate an action when the value exceeds or falls below a threshold. The threshold may be a single number as in the example of the significant wave height above, or it may be a more complex index based on a combination of conditions.

Certain indices of this type may only be needed at certain times of year or while certain operations or events are in progress.

Another type of user will be one who is trying to demonstrate improvement or worsening of the environment as time goes on. This user will usually be in government and will use the values of the index to formulate or modify environmental laws or regulations.

Most of the indices that will be developed in the earlier stages of GOOS will comprise physical processes. However indices of biological variables such as plankton abundance will be needed for coastal zone and fisheries applications.

Character of the Inputs

The data to produce an environmental index may come from a single instrument, from a network of instruments, or from a computer model. Generally the requirement will be more complex rather than less, and will involve a variety of variables, significant processing including modeling, and some interpretation and evaluation before being acted on. Some indices will include complex measurements of chemical and biological parameters that require laboratory analyses of long duration. In some cases anecdotal information may also be useful.

For some indices the products from NWP and short-term forecasting or indeed from seasonal to interannual prediction will be required to calculate a predicted index for consideration.

Input Delivery Requirements

In some cases input delivery requirements will be relatively stringent as the actions will have to be initiated in hours to a few days. Indices for environmental reporting will be associated with longer-term processes and inputs will have a much longer time frame.

Quality Control/Assembly Requirements

Quality control and assembly requirements will vary greatly depending on the index being calculated. The data flows for the ones dealing with operational time frame phenomena will have to have automated QC and assembly. In other cases there will be time for other more careful QC methods and incorporation of various other information. Data assembly can range from single parameters to a complex mix of satellite and in situ physical, chemical, and biological parameters.

Character of the Processing

Processing requirements will vary from very simple to very complex. In the wave example only a single number from an instrument on the oil platform need be reviewed. In other cases the processing will be quite complex and will include data assimilation into models.

The Level of Feedback

Some of the data used in calculating environmental indices will be collected for other purposes and feedback will have been established in the other data flow. For data that are being collected for just the purpose of the index it will be important to feedback information on data quality problems or missing reports to those collecting what is essentially monitoring data.

State of Permanent Archives for Data and Products

As noted above much of the data used for indices may be handled and archived in other systems as described above. It will be important that the relevant data and information produced in calculating the index be archived for the future, at least for potential legal requirements. For environmental reporting it is important that the data and analysis methods be documented and kept for re-examination when future indices are calculated.

Strengths and Building Blocks

The existing and developing data management systems and products will be available for use in developing systems to produce appropriate environmental indices.

Weaknesses and Needed Capacity

The use of environmental indices is at an early stage. The requirements for indices and their specifications will need to be developed by the science panels as GOOS develops.

Opportunities

No specific opportunities have been identified for the development of environmental indices as products in GOOS.

2.8 Fisheries Productivity

General Overview of the Requirement

In general there are two types of fisheries stock assessment. The first is a year to year stock assessment to monitor and manage the long term sustainability of commercial fisheries. Most countries that have large fishing industries do these sorts of assessments. They are done on an annual basis leading up to meetings that set quotas for species, fishing fleets, or countries. These systems use the most up to date stock assessments and environmental data that are available up to the time of the meetings. Thus there is an “operational” character required of the data management systems, although the time frame for data delivery is weeks to months rather than hours. It should also be noted however that where possible the relevant supporting environmental data collected for this application should be reported within hours for use in NWP and short term forecasting operations.

It should be understood that this assessment activity, which supports operational fisheries management, is a science activity. Because of the complexity and the state of knowledge of the various contributing factors, this activity will not become a routine job for an operational centre for the foreseeable future.

The other types of assessments are done as a part of long term research studies of the variability of fish stocks. The California Cooperative Fisheries Investigations (CalCOFI) is an example of this type of long term programme. CalCOFI celebrated it 50th year of continuous operation in 1999. The programme has operated more than 300 cruises resulting in hundreds of thousands of ocean measurements and numerous net tows.

CalCOFI has demonstrated that fluctuations in fisheries yields could only be understood by long term studies that included the links between the ocean and the atmosphere and the variability of these systems within different scales of time and space. Thus the approach to understanding the long term fluctuations became interdisciplinary and ecosystem-based.

Dramatic changes in sea-surface temperature, ocean circulation, and climatology have been observed including a regime shift that impacted the climate and biodiversity along the California coast. The consequences of an El Niño to nutrient, chlorophyll, and zooplankton patterns have demonstrated the need for a close look at the links between ocean physics and biology.

Thus data programmes that support management of the fisheries for long term sustainability will collect a mix of physical and biological data, have some operational data flows, and support both operational fisheries management and long term research programmes.

This application will share many issues with the biodiversity and regional integrated systems applications described in this section.

Products, Product Delivery, and Product-Service Chain Characteristics

Commercial fisheries operations are managed to achieve a maximum sustainable yield for each species in the fishery over the long term. The basic data needed include such items as the distribution of the species over age and size as a function of time, abundance of species from catch per unit of effort data, estimates of fishing mortality, spawning stock biomass, plankton abundance, etc.

Products that are measures of fisheries and ecosystem productivity include such analyses as

- indices of abundance generated from commercial landings and fishing effort data, and/or research vessel survey data
- trends in stock size, recruitment and fishing mortality
- indices of incoming recruitment developed from survey data
- relationships between effort and yield
- yield and spawning stock biomass-per-recruit curves based on data from biological sampling or other sources of information
- measures of ecosystem biomass such as plankton, mammal, seabird, and benthic organisms biomass.

There will be a variety of products generated for various purposes by the long term research agencies in pursuing understanding of the processes involved in determining the long term success of fisheries species.

A service chain for this type of application will thus be complex supporting multidisciplinary data collection and management, having some operational and semi-operational aspects, and supporting long term research programmes with data exchange, and data archival and access facilities.

Character of the Inputs

For fisheries management the data required typically include;

- commercial landings, gear, and fishing effort data,
- detailed information on trends in stock size, recruitment and fishing mortality,
- research vessel surveys or dockside surveys that produce age/size distributions and other biological information including bi-catch statistics,
- physical environmental data including temperature, salinity, currents, etc.,
- productivity information from ocean colour satellites, CPR surveys, and other data sources.

As an example of inputs to a long term research programme, CalCOFI researchers have collected samples of larval fish, fish, fish eggs, and plankton in net tows. They also have collected data on ocean circulation, temperature, salinity, oxygen, nutrients, chlorophyll, bottom core samples, and even pelagic birds.

Input Delivery Requirements

As for many of the ten applications described here the data gathered for fisheries stock assessment will support other of the applications. Thus while the input delivery for stock assessment is not time critical, some of the data should be reported in operational time frames for these other applications. Input delivery for fisheries management requirements will have a time scale of weeks to months with the shorter time frames applying as annual assessment times approach. There will be a significant requirement for sharing some of the data and cooperating with regional organizations such as the North Atlantic Fisheries (NAFO), PICES (its Pacific equivalent) and with bilateral organizations that jointly manage stocks that straddle national borders.

There will also be a requirement to share data with the fisheries programmes operated by the Food and Agriculture Organization (FAO).

Quality Control/Assembly Requirements

As more of these data are shared and exchanged, the quality control of fisheries biological data will require examination and agreements on standards and protocols and on the meta data necessary to document the collection, species identification methods, and analysis procedures. Species lists and taxonomic codes will have to be agreed and work will need to be done to harmonize and make comparable older data that use other codes.

Data assembly for this application will continue to be challenging as it will continue to bring together research and monitoring data, and the commercial fishing data from a significant number of sources, that deal with the data through diverse systems and on differing time scales.

Character of the Processing

Processing of fisheries management data involves a number of techniques for calculating indices, for modeling interactions and relationships between factors such as yield as a function of fishing mortality and age of recruitment. Some models are based on biological rules of increase or decrease, or on rules such as size and weight at age and maturation rates. The assessments must take into account the state of stock as the rules and methods will vary depending on the abundance of the stock. It should be remembered that this is a science activity and the procedures and methods used will vary from year to year depending on the state of knowledge as well as on environmental and biological factors.

The Level of Feedback

For the assessments developed for management of the fisheries there is an almost automatic feedback. Last years predictions will be examined in preparing this years predictions. For the data management it has been pointed out that there will be many organizations involved in using the data for a variety of purposes. These organizations will have a variety of experience and expertise in different data types. For the data sets that continue to be used and those archived for future users, it is important that any problems or deficiencies found in the data be fed back to the holders of the “official” copies of the database so that data can be corrected and documented with appropriate meta data flags.

State of Permanent Archives for Data and Products

There are a variety of databases held by national, regional and international organizations. These archives will be in various states regarding quality, consistency, and having adequate documentation in terms of meta data and supporting information. It is assumed that GOOS activities for this application will be developed by building on existing systems and databases and that this will be done by expert groups that include the requisite scientific and data management expertise. As development proceeds, the databases can be examined and actions to make improvements and standardize can be implemented.

It should be noted that databases on fisheries stocks have been developed and operated by many years by experts in the FAO. These databases are intended to provide a global perspective on the state of the various fisheries. Fisheries systems identified as elements of GOOS should be coordinated with the activities of FAO to be mutually supportive.

Strengths and Building Blocks

Some countries and regional programmes such as NAFO have several decades of experience in dealing with fisheries stock assessment. There is thus much expertise and capacity available to build upon. For research programmes the experience, benefits and successes of CalCOFI can be demonstrated and incorporated.

Weaknesses and Needed Capacity

The major weaknesses for this activity will be the diversity of data and methods, and the fact that fisheries stock assessment is not a solved problem. Data collection, analyses, products, data requirements, and interpretations will continue to evolve and the data management systems will have to evolve with them.

Opportunities

There may well be opportunities for closer cooperation between the management and research programmes. There is also a significant possibility of more cooperation amongst data collection and management agencies nationally and internationally to share technology and the workload to everyone’s benefit.

2.9 Satellite Data Processing Systems

General Overview of the Requirement

Observations from satellites will comprise a significant portion of the data that are required by the global observing systems for atmosphere, ocean and terrestrial applications. The data are particularly valuable because of the global coverage and because satellites can provide a variety of measurements that are not otherwise available or affordable.

For oceanography the variables of interest are sea-surface temperature, wind speed and direction, ocean topography, ocean colour, wave height and direction, sea ice extent and concentration, and in the future, hopefully, sea-surface salinity.

The various sensors generally observe radiances that need to be corrected for effects such as atmospheric conditions and orbital variations, and then converted to physical variables. Many of these corrections rely on in situ data for sensor calibration or on orbital characteristics calculated from ground stations monitoring the satellite position. The algorithms for conversion of the radiances to physical variables are generally complex and require significant computer resources in the processing chain.

Products, Product Delivery, and Product-Service Chain Characteristics

The products that are desired for oceanography are synoptic fields of the variables given above. Satellite only fields and fields created by merging satellite and in situ data are both required at different stages and by different users. High quality merged fields would be most desirable for all applications, but it is not possible to produce them soon enough for some of them. It takes time to acquire the in situ data and perform the necessary QC to obtain the best quality fields. There will probably be a requirement for re-analyses from time to time.

The service chain will have to supply satellite data sets on time scales of the meteorological cutoffs, and then improved ones for short-range forecasts for operations, and for strategic seasonal-to-interannual forecasting.

For application to NWP, most weather services have receiving stations to get the satellite data directly. For others much of the preliminary processing takes place in the satellite services (in the US also in EOSDIS) which relieves the organizations producing the products of handling the large bandwidth of the raw data and the heavy processing load of the conversion to accurate earth located physical variables.

Character of the Inputs

Inputs to the system at the satellite centres and meteorological and oceanographic services will consist of the satellite radiances received at the ground stations, in situ data for ground truth, QC, and merging. Also required is various other ancillary information needed for earth location, corrections, and conversions.

Input Delivery Requirements

The inputs are voluminous and must be moved by electronic networks because of the time factors involved. Since satellites provide data for most of the applications discussed here, the input delivery requirements will usually be those of the most demanding ones.

Quality Control/Assembly Requirements

Quality control requirements will vary with the application. For the fast delivery mode for NWP and short-range predictions QC will have to be automated and of limited sophistication. For less time constrained applications more sophisticated QC involving models and assimilation will be needed in order to produce the necessary high quality fields.

Assembly will occur in stages as more data and information become available.

Character of the Processing

The processing will be complex and CPU intensive. Large fast computers will be required for both the preliminary processing and for assimilation of the data into models as part of the generation of the fields. The development of error estimates for the product fields will be important.

The Level of Feedback

Feedback mechanisms for the in situ data will already be in place.

State of Permanent Archives for Data and Products

The meteorological and oceanographic services rely on the satellite services for the archival of the raw satellite data and the supporting information for orbital corrections and for sensor corrections for atmospheric conditions. The status of the archives for the derived earth located parameters is not clear and needs to be reviewed. It is unlikely that formalized archival procedures exist for some of these data or that a well-developed strategy has been thought out other than for EOS DIS. There would seem to be a requirement to review the needs for archival of satellite and supporting data. This could be carried out using the GODAE pilot project to set the context for the task.

Strengths and Building Blocks

There is a by now long history of management of satellite data so many of the pieces for management of these data for GOOS do exist and have been exercised on a continuing basis. Some systems exist in weather and oceanographic services that ingest satellite data and in situ data and produce data sets and data products for users.

Weaknesses and Needed Capacity

The integration of satellite and in situ data is not as routine a procedure as is eventually required for GOOS services and products. More work needs to be done on developing the application of satellite data to GOOS requirements.

Opportunities

The development of the EOSDIS has addressed in a significant way the problem of turning the satellite into useful products for users. The CEOS and its Strategic Implementation Teams with their pilot projects (includes GODAE) are also doing significant work towards implementing improved products and services. Thus there is a large body of continuing and very active work of which GOOS can take advantage and with which to cooperate in implementing the satellite aspects of the programme.

2.10 Regional, Integrated Systems

General Overview of the Requirement

Regional integrated systems in GOOS will be initially be developed within the COOP, mostly in cooperation with other organizations and based on enabling science from other international programmes. The systems will be diverse and complex and will depend much on the local problems being addressed and on the missions of the national and regional organizations involved. There is broad international interest in the coastal zone and other international organizations will also often be involved, e.g. LOICZ, CLIVAR, IGAC, GEOHAB, and GLOBEC. Thus while regional integrated systems are required for efficiency and effectiveness, because of the diversity of problems and the multitude of organizations participating, the systems will each present a unique and complex challenge in developing and implementing the necessary cooperative nested observation systems and data management strategies.

It is thus difficult to talk here in other than generalities. To come up with some reasonable ideas for this data management application it was thought best to base this section on the five pilot projects identified in the Strategic Design Plan for the Coastal Component of GOOS dated October 2000.

The five pilot projects are;

- The Vietnamese Forecasting System
- Coastal Observing System for the Eastern South Pacific Ocean (COSESPO) (Colombia, Ecuador, Peru and Chile)
- SW Atlantic Network: Quickly Integrated Joint Observing Team (QUIJOTE)
- PHYTONET: The Phytoplankton Network
- The Coordinated Adriatic Observing System (CAOS)

Products, Product Delivery, and Product-Service Chain Characteristics

These systems will serve so many applications that it is not possible to predefine the products, and delivery schedules. For example PHYTONET will provide products for;

- aquaculturists: to take timely decisions on management of aquaculture activities;
- health authorities: to optimize human health protection;
- scientists: who are both contributors and users of the data sets for scientific purposes;
- monitoring authorities: to produce data in a more timely way and reduce the cost/benefit ratio;
- resident populations: to diminish the risk of accidents caused by ingestion or exposure to algal toxins;
- tourist industry: to protect tourists from exposure and to mitigate the effect of HABs;
- decision-makers: to plan a safe and effective use of the coastal zone;
- regulatory bodies: to lay down common regulations for monitoring operations.

The service chain will have to involve scientific organizations with expertise in the various data and applications, data management organizations that integrate data sets to produce products, operational organizations providing services related to the mandate of their organizations as well as to the COOP pilot project, and archival organizations to save the data and products for future users. It is important to note the service chain will involve many other programmes. CAOS involves EuroGOOS and MedGOOS.

The users of the proposed observing systems include, local authorities (e.g., harbor authorities), government agencies (e.g., environmental agencies), coastal managers, industry, and research organizations.

Character of the Inputs

Most of the pilot projects will collect physical data on variables such as water levels, waves, currents, temperatures as well as chemical or biological data. Some of the physical data will be appropriate for use in local numerical weather prediction and associated services such as wave and storm surge forecasting. These data will need to be forwarded in operational time frames to the appropriate agencies.

Other physical, chemical and biological data will not have existing supporting national or international data management systems. These systems will have to be designed and built as part of the development and implementation of the pilot projects. There will be a requirement for the harmonization of different data bases, built on agreed protocols and taxonomic lists of species and common procedures for data archives. The recovery of data gathered with old sampling techniques in the transition towards new and improved ones will have to be done carefully, in order not to interrupt or distort existing data-series.

Thus there will be a mix of data that can be routinely processed in relatively short time frames and other data that will likely have to be processed and quality controlled in research mode by research organizations.

Input Delivery Requirements

As discussed above regional systems will collect some data that has that has to be provided to user organizations in hours to days. Other data will have to be worked over by research organizations and will require weeks to months for analysis and product generation.

Quality Control/Assembly Requirements

For regional systems more of the physical data will have established and standardized quality control procedures that can be adopted. For much of the other data to be collected in the COOP pilot projects, these procedures and standards will have to be developed.

Data assembly will be challenging. Typically several organizations will need access to subsets of the total data set. These organizations may well do quality control and analysis on the subsets which will all need to be reincorporated at some point into a final data set for long term archival. The other problem will be that some parts of the data set will be available quite quickly and others will be delayed because of the differences and complexities of the analysis procedures. Keeping available a “best” copy of the project data set in the “continuously managed database” sense will not be straight forward.

Character of the Processing

The processing systems will vary from relatively simple ones for variables that are well understood and measured by standard means to processing in a research mode that will require analyses, evaluation, and perhaps more analyses.

The Level of Feedback

Feedback will be a very important design consideration in building these regional integrated systems. In particular because of the variety of data and the number and diversity of the organizations involved there will have to be more or less continuous contact between the various data management groups to ensure that the most consistent, harmonized, and high quality data set is available to the users at all stages of the project.

State of Permanent Archives for Data and Products

Some data will be archived in the existing systems of IOC and WMO. Much will not. Diverse formats and standards can be expected. There will be a need for distributed archives to provide access to the data. There will also be a need to recruit organizations with expertise in the various new (for exchange) data types to become associated with IOC and WMO subsidiary bodies such as IODE and JCOMM to support the data management activities of these regional systems.

Strengths and Building Blocks

The pilot projects are presently supported by enthusiasts and organizations, both national and international, that have the necessary expertise to develop the end-to-end systems from data collection to user products. The IOC and WMO systems and their network of data and processing centres are also available for expertise and advice and for processing, archiving, and disseminating data and products.

Weaknesses and Needed Capacity

Standards, protocols, interoperability and harmonization of databases, data communications according to schedules, are all important issues for these systems to be successful.

Opportunities

A major opportunity for the development of the data management aspects of regional integrated systems is to work with the COOP pilot projects to build some example systems which can be evaluated and further improved until the data management goals for such systems are better defined.

3. Definition of Requirements and Guiding Principles

GOOS has established some overall guidelines and principles for inclusion of programmes in GOOS. Guidelines include recommended practices for observations, QC/QA, standards for data and products and for meta data, and archiving. Several general guidelines suggested include:

- Observation programmes must be long-term, sustained programmes that will continue into the foreseeable future.
- Observations should be systematic and relevant to the observing system – programmes should be designed with careful scientific attention to the necessary spatial and temporal sampling, precision, accuracy, and calibration, and tuned towards users' needs.
- Programmes should be subject to continuous examination including scientific evaluation of the observation programme, QC/QA practices, quality of data and products, etc.
- Measurements should be cost-effective, timely, and treated routinely by dedicated staff.
- Data quality must meet standards acceptable to the peers and users of the data being submitted.
- Data must contain documentation (metadata) of a standard acceptable to peers using similar data.
- Meta data must be archived and made available with the data.

At a lower level of detail, section 3.1 provides an analysis of the general requirements for a GOOS DIMS, and comments on the abilities of the existing systems to meet these requirements. Section 3.2 summarizes some of the characteristics that will be common to many of the end-to-end D&IM systems that will be required by GOOS and is based on successful characteristics of existing systems as well as the requirements identified in section 2.

3.1 Analysis of the Requirements

A discussion of the requirements is given for each of the headings of section 2 across the ten categories. The ability of the existing systems to respond and the requirements for development are also commented upon, but not in detail.

General Overview of the Requirement

The requirements for GOOS D&IM management are for a series of intersecting end-to-end systems with the characteristics that follow. The systems must

- ingest various versions of in situ and satellite data and meta data (satellite data processed to earth located physical parameters at least),
- apply the appropriate level of quality control based on the delivery time for each data set or product,
- provide the best available copy of the data, meta data and products at several elapsed times after observation to a variety of users,
- provide for the archival of the best available final copy of the data, meta data and products for future users with updates of the archives if re-analyses are done or better copies of the data and information become available.
- provide comprehensive feedback on (i) data problems to collectors, and developers of the systems, (ii) on the usefulness of the data and products to users, and (iii) to the developers of the systems, on the timeliness and completeness of the data flows.

Each of these systems is likely to involve several organizations with varying expertise and emphasis in operational data assimilation, modeling, data dissemination, meta-data standards, archival, and product development and distribution. The challenge will be to develop integrated networks that can adapt to accommodate development of new requirements and capabilities as the full end-to-end systems evolve.

Only the most mature of the IOC and WMO systems that have been implemented by IODE and JCOMM approach this standard for an end-to-end system operating on a global basis. The World Weather Watch operates such a mature system for surface and upper air observations. For the ocean the IODE and JCOMM operates the GTSPP, a relatively mature system. Other ocean systems that have been identified as part of the GOOS IOS are in various stages of maturity measured by the requirements above.

Not all of these systems, regarded as mature or otherwise, necessarily fully reach the standard required by GOOS. For example meta data and quality control standards are not yet sufficient for use of the data for precise calculations on climate. There needs to be more emphasis on quality data products for the needs of GOOS. Timeliness of submission of the original data or the fully processed delayed mode data needs to be improved.

For the coastal zone, the observing subsystem will be multidisciplinary and multidimensional. The data streams will include physical, biological, chemical and geological data from *in situ* and remote sensors, and from laboratory analyses of samples collected from fixed platforms, drifters and ships. Some applications will require real time or near real time inputs.

Products, Product Delivery, and Product-Service Chain Characteristics

Most of the systems in operation today have been targeted for one class of user, be it the weather services, or the physical oceanographic researchers. The generation of products has generally been left to these users. GOOS needs a different paradigm, the service chain.

The service chain embodies the idea of an end-to-end distributed system that takes in data, performs sophisticated QC, analyses it as necessary, delivers products to a broad and varied user community at various times after collection, and archives the data for continuing service to the research and engineering communities. Service chains will intersect as the data from one chain is passed at some point into another service chain for use by another application.

The data communications and management structure is envisaged as a hierarchy consisting of three levels: local, national, and supra-national. Data and information are disseminated laterally within levels and vertically among levels.

In particular the service chain considers the needs of all the users from the planning stage. A review of the requirements in section 2 will reveal the broad needs across most of the applications for the physical meteorological and oceanographic data. That service chain will have to address all those requirements and will need to be very well coordinated between the applications.

It is important, as GOOS develops other applications and pilot projects, that the development of the products and the service chain that delivers them receive a lot of attention right from the beginning. For the existing systems identified as part of the IOS, the service chains that exist need to be re-examined and improved.

Character of the Inputs

The inputs to the various GOOS systems vary significantly in timeliness, completeness, and quality control requirements. Some inputs will be raw data. Some inputs will already be a product or product data set. Satellite data are an example. Other inputs will include messages in the WMO codes on the GTS. There is the whole question of routine input of the meta data.

National programmes of environmental research and monitoring were developed independently, case-by-case, by different groups to address specific issues. There are a plethora of programmes in coastal ecosystems that employ different platforms and methods, make measurements on different time and space scales, and use different data management schemes designed for the purposes of a particular agency, institution, or programme. This will result in a significant number of diverse inputs in terms of formats and standards that will only gradually disappear as the needs for standards become clear and action to develop them is taken.

Centres that participate in GOOS systems are going to be faced with dealing with an increased number of data types and formats. It will be important when developing the end-to-end systems to give consideration to opportunities for adopting existing standards. When such standards do not exist for a given data type, then formats and standards will have to be developed. In undertaking this development, requirements for other systems should be considered so that a common standard can be established which can then become more or less a standard at least across several systems. At the same time the development and implementation of GOOS systems cannot be allowed to stall over the standards and formats issue.

The opportunities for sharing software or software development activities should also be kept in mind.

Input Delivery Requirements

As can be seen from section 2, delivery of data for GOOS must be considerably more timely than in the past. In particular the delivery of data that was not available in real time and of the fully processed delayed mode data must be accelerated in order to create the applications (products and services) in a time frame that meets the users' needs and expectations. As the new applications are developed for each set of users it is important to consider the needs of other users for early versions of the data. For example any applications that collect physical surface or subsurface data should consider making it available on the GTS for NWP and short-range forecasting.

Electronic networks will play an increasingly important role in the development of GOOS systems. The two networks of choice will be the GTS and the Internet. The GTS should be used for physical data circulated in operational time frames for NWP and other applications operated by the weather services. For applications not associated with weather and short-term forecasting, the Internet should be the network of choice. All organizations that will participate in GOOS will likely already have Internet connections. This will avoid overloading the GTS, and having to develop the complex formats and encoders/decoders that are required on the GTS.

As integrated satellite and in situ systems develop, the use of special high speed networks will become increasing important. These networks (HPRENs for high performance research and education networks) can not be used for operational purposes at present. As operational needs for them emerge, an operational equivalent will need to be found.

Quality Control/Assembly Requirements

The need and tendency of the past decade has been for more quality control, and for that function to move closer to the observations. Automated procedures have been introduced early in the data flows and some data streams are routed through semi-automated QC procedures before being made available to the GTS (still in time for the meteorological cutoffs though).

As coastal zone applications develop the quality control and assembly function will see systems with more and more variables from many more sources that will have to be assembled in various combinations for differing analyses for a more diverse user community.

For GOOS this tendency needs to continue. As was seen in section 2, more quality control is needed sooner to support the generation of products sooner after data collection.

Data assembly has become more complex with the need for replacement of early low resolution copies of data with fully processed higher resolution data. Also there will be an increasing need for interdisciplinary data sets as the other GOOS modules are implemented. Assembly of integrated satellite and in situ data sets has added another dimension of complexity.

Character of the Processing

In the past the main deliverable of a data management system was a data set with perhaps a few products of an inventory nature, such as maps of the data locations. All sections of this report have noted the importance of delivering complex predictions and useful scientific products to groups of users as well as the data. Once this is decided then the character of the processing in an end-to-end system becomes far more complex involving modeling and data assimilation. The data must pass through organizations that add value to the service chain by doing data assimilation and complex modeling, including biological modelling, on a routine basis as a service to the operational community. All this must happen in an appropriate time frame. In addition the organizations must produce archives of high quality data and products.

The Level of Feedback

The review of the 10 application areas has identified the need for improved feedback at all stages of the data and information flow. The two types of feedback are feedback on problems with data collection or data flow, and feedback on the accuracy and usefulness of the products to the user and/or user satisfaction.

Feedback should be provided for:

- correction of problems in data collection, instrumental and otherwise;
- correction of timeliness problems at all stages of data flow;
- correction of databases or at least flagging of values for erroneous, suspect, or missing data or meta data;
- correction of model deficiencies and redevelopment of models to produce more reliable analyses and predictions;
- improved design of products to meet user needs;
- improved knowledge of user satisfaction and the impact and value of the services being provided.

Although feedback now exists and is effective for parts of some of the existing systems, there needs to be a more comprehensive approach to this function. In particular the QC must feed back to correct or flag the databases. User satisfaction and the usefulness of the products and services needs to be demonstrated.

State of Permanent Archives for Data and Products

For those systems that compose the GOOS IOS, mostly adequate archives exist. Some improvements in timeliness and completeness are required and meta data standards could be improved, but for the most part the data are being preserved for future users. However this is only true for physical in situ data. Much work remains to be done in the existing systems to deal with the management and archiving of biological and chemical data. The archiving of satellite and integrated satellite and in situ data and of data products is not well organized or documented. As the pilot projects proceed these aspects of the permanent archival strategy must be addressed.

For the historical data sets of interest to GOOS users there are also problems. IODE needs to develop an active programme to expand the expertise of its data centres to include chemical and biological data. There are centres in the IODE system that have these capabilities now, so the expertise to proceed is available.

Strengths and Building Blocks

There is a lot of the necessary data and information management expertise and experience available in the existing systems of IOC and WMO, and in the research community. If cooperative programmes are developed around the pilot projects where research and data management organizations work together to design and build the D&IM systems then considerable progress can be made, even in the existing financial climate. However it must be recognized that there is need for significant additional resources to be devoted to international data and information management for GOOS to be successful. Some of these resources will have to be new resources.

Weaknesses and Needed Capacity

Consideration of the requirements in this section and section 2 demonstrate the shortfalls of the existing systems to deal with the broad range of applications, the operational requirements for data and information, the integration of satellite and in situ data, and an increased variety of physical, chemical, and biological parameters. There is a need to improve our ability to integrate regional and global data systems. Another concern is the state of data management software in the centres now managing ocean data and information. Many of the software systems that are now operational in the data centres are not adequate to the job. There is a need to take advantage of more sophisticated algorithms and software technologies to increase the amount of automation. There is a need to provide the centre staff processing the data with graphic presentations and the power to interact effectively with the presentation and the databases behind them to significantly increase the sophistication of the processing and the volume of throughput. Progress in this area will allow increases in the volume of data handled with a minimum of increase in resources.

Opportunities

Data management at this time has a high profile in the international area and, in some member states of IOC, at the national level. The willingness of other national and international organizations to participate in cooperative ventures is also at a high at this time.

It is a good time for JCOMM and IODE to take a highly active role in expanding the capabilities and role of the existing systems by

- dealing with satellite data, preparing integrated data products, implementing improved meta data directories and improved services on the web, and handling a broader suite of physical, chemical and biological parameters,
- developing cooperative programmes with the research community to implement end-to-end systems for all the modules of GOOS, and
- developing improved capacity building programmes in cooperation with other agencies such as the World Bank, UNEP, and UNDP.

3.2 GOOS Data and Information Management Guidelines

This section summarizes some of the characteristics that will be common to many of the end-to-end D&IM systems that will be required by GOOS. Some of the sections are based on requirements identified in sections 2 and 3.1. Others are characteristics of existing systems that are proving successful at present.

3.2.1 Planned and Directed Data Systems

It has been stated earlier that GOOS is being designed as a series of end-to-end data collection/management systems that will deliver data sets and products to specific users for specific purposes such as those defined in section 2. The observing network, the frequency of observations, the capabilities of the data management system, the products that are produced, the schedules the products are produced on, and the distribution to the various users at various time are pre-planned. Thus the data that are collected should by definition satisfy the need, provided that the sampling and processing are done according to the standards and specifications for the systems (which needs to be monitored). This characterizes an operational system as opposed to an opportunistic system where “secondary users” look for any available data that has been collected for various purposes and then examine it to see if it can be used for their need at the moment.

This does not mean that everything will work as planned or that the products will be adequate at the beginning. Further development of the systems will obviously be required both in the medium and long term as knowledge and applications develop.

What it does mean is that GOOS data management has a different starting point than, for example, traditional international oceanographic data management that was based on data of opportunity being provided to "secondary users". The GOOS D&IM systems will be based on specific responsibilities accepted by participating organizations to do specific tasks with specific data according to agreed standards and schedules. If data collected by others for other purposes are to be used, then its applicability and availability will have to be reviewed when the GOOS systems are being designed or further developed.

It should be understood that

- GOOS cannot depend on opportunistic data sets from research for operations.
- GOOS data must meet predefined standards and specifications based on the application being supported.
- GOOS data will be a subset of the ocean data that are collected globally (hopefully a large subset).
- GOOS data should not be merged with other data unless it can be separated from these other data and, along with its meta data, be made available to users.
- All GOOS data sets, operational and historical, must receive scientific approval as meeting a GOOS standard.

All ocean data can not be collected and managed to the GOOS standard because of the associated costs. At the same time some data for specific purposes may be collected to a higher standard.

3.2.2 Operational Components for Most Systems

Most data systems implemented in GOOS will have an operational data flow. There are two reasons for this. The systems are generally supporting operational programmes that require the data in a time frame that is not consistent with delays in submission of the data. The second requirement has to do with management of the data collection programmes. Early availability of the data is necessary to address any quality or collection problems and avoid data gaps or long periods collecting bad data.

3.2.3 Extensive Use of Electronic Networking

Electronic networks will be a common feature of all GOOS systems. They will be necessary to provide the operational data flows. All the participating data centres and associated scientific centres will be required to operate web sites for the provision of data sets and products to users.

It will also be necessary to move the data between centres participating in a data flow on a regular basis. For efficiencies, the data will usually be moved to a centre that has the software and expertise to do a certain task rather than developing the capability in a centre that does not. Networks will also be used to move the products to users on the agreed schedule and for making management information available to those responsible for monitoring operations.

3.2.4 Products for Users and for Managers

The GOOS D&IM systems will each have a number of centres that are providing data products for users and data management products for managers. Products for users will include those described under the heading "Products, Product Delivery, and Product-Service Chain Characteristics" in section 2. In addition the centres will provide on-line products concerning the data flow at various points in the system so that a manager (or data coordinator) monitoring the data flow can identify problems with data quality, completeness, and timeliness and know what action to take to rectify the situation. The centres that operate databases should also provide detailed on-line inventories of such things as positions of data points, distributions of data in time, distributions of parameters, and other types of presentations useful to users not familiar with the data collection programmes.

3.2.5 Web Access to Data and Products

User access to data and data products will be enhanced by requiring participating centres to operate web sites and by publishing CD-ROM sets on a regular basis for users not on the Internet. Each centre that provides data sets or data products will be expected to maintain a special Web page that allows a user to navigate to the GOOS products and view them, view the data holdings, order or download data, or identify contact persons for additional information.

GOSIC, as an entry point for access to GOOS data and products, will provide data flow diagrams for the GOOS data flows that will identify all centres that have data or data products. Clicking on the centre in the data flow diagram will transfer the GOSIC user to the centre's special Web page that points to the GOOS relevant data and information without the user having to navigate complex and varying organizational Web pages to try to identify the GOOS material. If the centre operates a database, then one of the entries on their GOSIC interface page should be the detailed data inventory for that database. See section 5.1 for information on GOSIC.

3.2.6 Continuously Managed Databases

The continuously managed database (CMD) is a concept implemented by the GTSPP and WOCE for the Upper Ocean Thermal Experiment. The idea is that the database at any given time holds the "best", highest resolution copy of an ocean profile. The first copy received may be the low-resolution copy from a BATHY or TESAC message from the GTS. When the delayed mode, quality controlled, high-resolution copy of the data is received after some weeks to months it replaces the low-resolution version. Any time there is a "better" version of the profile available, such as after "better" quality control, the better version replaces the version in the database as the "active" copy. Each copy of an observation has the necessary meta data stored with it to identify which version of an observation is of what type so that, for example, high-resolution data can be distinguished from low-resolution data. When other profile data become available than were available in real time they are added to the database. Thus at any given time the database holds the best data and is as complete as it can be.

An important tool in managing a CMD is a duplicates identification algorithm that allows the automatic detection of duplicate observations so that the older version of an observation can automatically be located and flagged as an inactive copy.

Many of the databases that will be operated as part of GOOS will be of the CMD type to accommodate replacement of the early operational versions by later, better versions and to accommodate phased QC operations. Also known errors must be corrected or flagged in the database when they are discovered.

3.2.7 Scientific Participation in the Data Management Operations

One goal of a good end-to-end data management system is to produce at the end of the day a high quality and complete database along with a variety of useful and informative analyses and presentations of results. At the present state of knowledge of ocean products and measurement instruments this goal can not be realized without significant participation by the scientific community. To this end the twinning of a scientific organization with a data centre to provide a joint Data Assembly Centre (DAC) has been a useful arrangement in the past. This arrangement provides the data centre with the necessary scientific support and the scientific organization with data management support to the benefit of both in managing the data flow.

3.2.8 Scientific Quality Control of Data

The more sophisticated quality control procedures that need to be applied to data in order to ensure a high quality, long term database can generally not be applied in operational time frames. These procedures generally feature use of the data in a model or analysis in which the more subtle errors will be detected. Some of the recent global experiments implemented a data flow that involved the data going on a semi-annual or annual basis to a scientific organization for this type of QC. Data that were identified as suspect or erroneous were flagged to identify the problem and returned to the data centre. The new QC flags were then attached to the observations in the database.

3.2.9 Meta Data Standards

The importance of meta data attached to the observations in long term data sets has been recognized for some time. Most observation programmes now provide for meta data that describe data collection methods, instruments, quality control procedures applied, analyses done, etc. Meta data must be stored with the data and included when the data are provided to users.

The development of effective meta data standards has been much improved by having both the scientists and data mangers involved in their specification. (One difference between GOOS and the ocean experiments is that GOOS users will have less familiarity with the data collection and processing. Thus there is a more onerous requirement to collect and maintain an adequate set of meta data to ensure proper application now and in the future.)

3.2.10 Historical Data Sets

There will always be oceanographic, meteorological, and terrestrial environmental data sets and products not produced by GOOS available from scientific and operational organizations around the world. Many of them have overlapping content, are versions of the same data set, are combinations of several of the same data sets, have been quality controlled and analyzed by different and unspecified techniques, or have varying and often incomplete meta data attached to them. These data sets and products will have been developed for purposes other than to address the questions that GOOS has been designed to answer.

The data requirements for GOOS have been developed to address specific scientific questions and operational needs. The quality control and analyses have been specified to ensure that the resulting data sets can answer these questions. These other data collected for other purposes (for GOOS this has been called opportunistic data) may or may not be applicable. How does a user looking for historical data know?

GCOS and GTOS have discussed this issue in depth. Both these observing systems have chosen to have their science panels review the more relevant historical data sets for quality and applicability to the applications being served. If a data set meets the observing system standards, then it is listed in a “data set registry” that is part of GOSIC.

It is recommended that GOOS addresses the subject of historical data sets as the existing data management sytems are reviewed and improved and as new data management systems are developed.

3.2.11 Data Archaeology

In the context of GOOS data archaeology and rescue represents an activity that identifies and retrieves ocean data that have not previously been captured for the international data banks. In fact most of the data that are generally available is physical data related to the programmes of IOC, WMO, ICSU, and SCOR. Much needs to be done to locate and rescue the balance of the physical data and the biological and chemical data.

The Global Oceanographic Data Archaeology and Rescue Programme (GODAR) is an IODE programme that has very successful and more than doubled the international databases of ocean temperature and salinity data. GODAR has also taken available opportunities to rescue some chemical and biological data but has not yet addressed these data types in a comprehensive way.

It is clear that for GOOS there is a need for a significant effort related to data archaeology and rescue. It is even more clear that this effort will have to be focused as there are not enough resources in the oceanographic community to effect a complete rescue of all available data. The question is how to develop a focused and effective data rescue programme.

It is assumed that implementation of the D&IM strategies described in this document be accomplished by focusing on the IOS and the pilot projects. Groups of experts consisting primarily of involved scientists but with strong data management support, presumably from IODE and JCOMM and well developed participating data centres, would address the adequacy of the components of the IOS, such as GTSPP and GLOSS, and the pilot projects. As each component or pilot project was identified, the need for various databases and data archaeology and rescue would be identified and defined.

Such an approach would ensure the necessary focussed effort and also set the priorities for action based on the scientific priorities and needs. As needs are identified, then cooperative data rescue ventures can be sought with other organizations including GODAR, and perhaps with other international players such as UNEP or UNDP.

3.2.12 Characteristics of Satellite Data Management

The data flows for satellite data are more complex than those for most of the in situ observations. The complexity results from two sources. The first source of complexity is the problem of turning the satellite sensor output into environmental variables. This can take several steps. The second complexity is a result of the large volume of sensor data that must be processed.

Figure 1 represents the steps necessary for a data management system to turn satellite observations into a product for an end user. This figure has been reproduced from the report of the Fourth Session of the Terrestrial Observing Panel for Climate (TOPC). It applies equally well to management of the data from ocean observing satellites.

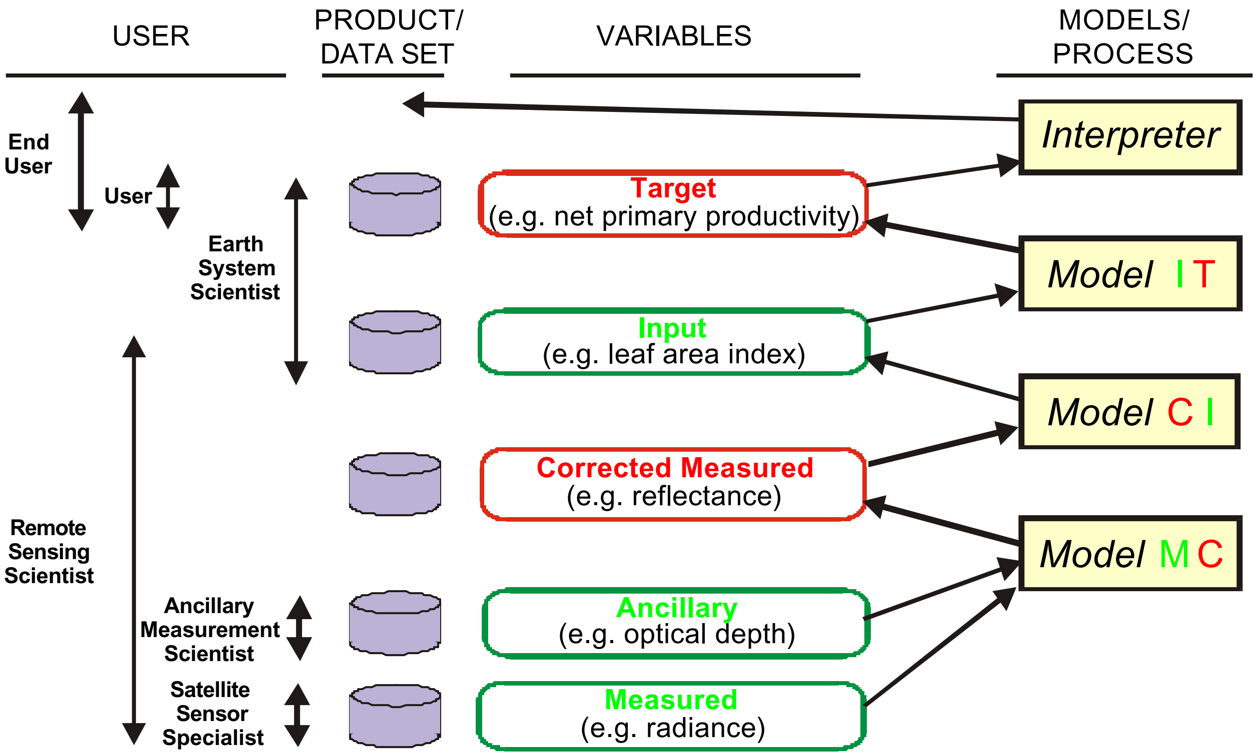


Figure 1 - Management of Satellite Data from Observations to End Users

The figure represents an end-to-end system in that it covers the management of the data from collection to its presentation in a product that was a stated goal of the observing system. In this case the final product or "target" is a presentation of global net primary productivity.

The first column identifies the users of the data and data products at the various stages of the life cycle of the data. Until the data reaches the stage of an environmental variable (in this case leaf area index), the users are the remote sensing specialists who know how to turn the original measured radiance into the value of an environmental variable at a point in space and time. At the "input" stage there will be "users" who use the environmental variable leaf area index to address local questions and the "end users" who use leaf area index as one variable in an analysis to address one of the larger problems for which the global observing system has been established, e.g. a global analysis of net primary productivity.

The second column simply shows that both data sets and products are needed at all steps in the management of the data.

The third column shows the type of parameters that are being managed at each step in the development of the data sets.

The fourth column represents the analyses that must be done in moving the data from one step to the next. The "model MC" is an analysis that accepts the measured radiances from the satellite along with any necessary ancillary data and produces data that are corrected for atmospheric effects, etc. The "model CI" takes the corrected sensor values and applies an algorithm to convert the corrected data to environmental values. In one of the steps the correct space and time coordinates will be inserted taking into account such things as orbital corrections.

These first two models will generally be operated within the remote sensing community of operational data collection centres and processing centres like the NASA Distributed Active Archive Centers (DAACs). The DAACs have been established to do the preliminary processing of the satellite observations and produce data sets and products that contain the environmental variables for users. DAACs do not exist for products from all sensors.

The "model IT" takes the input environmental variables and performs an analysis to produce the target product in question, in this case global primary productivity. This model will be developed and operated by "earth systems scientists" who will be environmental experts or process experts generally operating outside of the remote sensing community.

One feature of an end-to-end data management system that has not been discussed in the above figure is the requirement to archive the data at each step for future uses. These future uses will very likely include reanalysis projects as algorithms are improved for the models used in the various steps. It will also be necessary to ensure that the necessary meta data are assembled, made available, and archived for each of the modeling steps.

3.2.13 Management of Numerical Model Data

Generally speaking most of the answers we are seeking in regard to aspects of climate and global change will come through models that assimilate data and produce fields interpolated in space and time, in either nowcast or predictive modes. Thus the management of model data will be a fundamental element of GOOS.

Management of model data is not a new experience for meteorological and oceanographic agencies. Meteorological centres have been exchanging model data for decades on the GTS and some oceanographic centres have been dealing with gridded fields for a number of years as well.

For GOOS numerical model data will eventually have to flow in an operational mode from modeling centres to various users and other modeling/analysis centres. Some modeling data will be appropriate for long term archival for future users. Some archived model results will from time to time be made redundant by reanalysis projects. Decisions will have to be made as to whether the old version in fact has been replaced by the new or if both need to be retained. Formats will have to be designed and meta data standards established to ensure the model results that are archived will be adequately documented so as to be usable in the future.

Specific requirements for management of model data should begin to clarify as GODAE develops. It is suggested that the initial principles, techniques, and standards for management of numerical model data be planned and implemented as part of the GODAE data flows.

3.2.14 The Concept and Implementation of End-to-End Data Management Systems

GOOS data will be managed in a series of end-to-end data and information systems. Each system will consist of a variety of scientific, operational, and data management centres. The data management centres will in many cases already be supporting programmes of IOC, WMO, and perhaps FAO for the living resources component of GOOS. Information on the data flows, the participating centres and their responsibilities, and on access to the data and information will be available on line through the information centre (GOSIC). The systems will be described in terms of the generic JDIMP model for ease of understanding. (see Annexes A and B). Annex E describes the GTSPP, which is a specific example of an end-to-end data management system.

Participating centres will have to perform such tasks as analyses of data flows, scientific analyses of the data, generation of fields for models, quality analyses, etc. Some of these analyses will be required not to provide the data and data products required by the observing system, but to ensure that the system is working and that data flows from collection, to modeling centres, to final products in the hands of users. In terms of the JDIMP generic model these other analyses will be supporting the evaluation and feedback function and will be identified as such in the responsibilities negotiated with the centre.

Participation by centres in an end-to-end data management system will be on a volunteer basis. However centres that volunteer will be expected to take on the responsibility for the medium to long term. GOOS should consider whether centres might be asked to sign a memorandum of understanding (as some centres did for WOCE) making a commitment to certain activities for a certain period of time.

Responsibilities of centres providing data and information services on behalf of GOOS will include:

- acquiring data;

- analyzing it and performing quality control;
- preparing products for users, other centres, and for the data coordination/feedback function;
- passing on the data and/or products according to a schedule;
- providing data to long term archive centres;
- maintaining web sites with certain data or data products, and meta data,
- maintaining data and information files on-line to facilitate access by users,
- maintaining web pages linked to the data, data products and services provided on behalf of GOOS, etc. for access through GOSIC, and
- producing operational data flow analyses to demonstrate the performance of the data flow in achieving the spatial and temporal coverage and producing good data. This is to be available to the data coordinator (see section 5.3) for use in the role of monitoring the various data flows.

The end-to-end systems will be identified and described on-line through GOSIC.

4. The GOOS Initial Observing System (IOS)

This section describes the Initial Observing Systems for in situ and satellite data separately. However the implementation of GOOS systems will in many cases involve data flows that include both satellite and in situ data. These requirements for merged flows will be identified and specified as the science and data management panels develop the specific systems.

4.1 The IOS for In Situ Data

The GOOS Initial Observing System (IOS) for in situ data is built from contributions of existing programmes of existing operational bodies. Many of these programmes are not exclusively contributions to GOOS although for many of them GOOS will become the major client. Many of these programmes will have evolved for other purposes but also address, are compatible with, and satisfy GOOS requirements. In principle these programmes can provide contributions to GOOS as well as continuing to serve the original groups of clients. What is important is to make effective use of expertise and resources to serve all the clients of IOC, WMO and the other international sponsors.

The IOS is described in terms of "Level 1" contributions, "Level 2" contributions, and the Pilot Projects. Level 1 contributions are those of existing bodies that have agreed that at least some of their data can be regarded as a contribution to GOOS. This of course also implies that these bodies agree to meet the standards defined by GOOS for timeliness, processing, quality control, and archiving of the portion of the data that they handle that is "GOOS data".

Level 2 contributions are data that are needed by GOOS but for which specifics have not yet been negotiated or for which the observational requirements have not yet been defined for the relevant GOOS module.

GOOS also has a number of planned or operating pilot projects that are collecting or will collect data in the process of developing strategies and methodologies for the implementation of the global GOOS.

4.1.1 Level 1 Contributions

- The global Ship of Opportunity (SOO) network (formally under JCOMM).
- The WMO's Voluntary Observing Ship (VOS) network (formally under JCOMM) whose Implementation Plan will seek to address GOOS/GCOS requirements.
- The Global Sea-Level Observing System (GLOSS) (formally under JCOMM).
- The WMO's Global Telecommunications System (GTS) is acknowledged to serve operational oceanographic requirements.
- The Global Coral Reef Monitoring Network (GCRMN).
- The Global Temperature and Salinity Profile Programme (GTSPP) (formally under JCOMM and IODE).
- The ocean data buoys coordinated by the Data Buoy Cooperation Panel (DBCP) (formally under JCOMM).

- The ENSO observing system including the Tropical Atmosphere Ocean (TAO) and TRITON array of moored buoys in the equatorial Pacific (formally under JCOMM).
- The GOOS Data Center of the Atlantic Oceanographic and Meteorological Laboratory (AOML) of the U.S. NOAA.
- The Continuous Plankton Recorder (CPR) Survey.
- The International Bottom Trawl Survey (IBTS) of ICES in the North Sea and adjacent areas.
- Time Series Station 'S' off Bermuda.
- Time Series Station BRAVO in the Labrador Sea.
- The Electronic JCOMM Products Bulletin.
- The Global Observing Systems Information Centre (GOSIC).
- California Cooperative Oceanic Fisheries Investigations (CalCOFI)
- The ocean observing satellites of NOAA and other agencies (see table 1 below)

Some of these elements should also be listed as part of the GCOS Initial Operational System

Table 1. Remote Sensing Satellite Missions Contributing to the GOOS Initial Observing System. Note: (i) SST is covered by AVHRR on NOAA satellites (already in the GOOS-IOS), or is to be determined; (ii) the trend is for research missions to graduate into operational missions.

Status	Ocean Surface Topography	Ocean Vector Winds	Ocean Colour
In orbit	Topex-Poseidon (1992)	Seawinds on QuikSCAT (1999); operational data stream	SeaWIFS (1997); MODIS on TERRA (1999)
Awaiting launch	Jason-1 (summer 2001); operational data stream	Seawinds on ADEOS-II (end 2002); ASCAT on METOP (2003); operational	MODIS on AQUA (end 2001)
Planned follow-on	Jason-2 (summer 2001); operational partners (NOAA/EUMETSAT)	ALPHA.SCAT on GCOM-B1 (2006)	VIIRS on NPP (2005)
Goal	Jason-3 (c.2009); operational	Decision to be determined: active or passive microwave on NPOESS	VIIRS on NPOESS (c. 2009); operational

4.1.2 Level 2 Contributions

- Selected ocean observing satellite missions
- Appropriate parts of JCOMM (to be negotiated)
- Appropriate parts of IODE (to be negotiated)
- Appropriate national observing systems (to be negotiated)
- Appropriate commercial observing systems (including oil platforms)
- The international Mussel Watch programme
- Appropriate parts of IOC's harmful algal bloom programme

4.1.3 Pilot Projects

In addition to the observations being made through the GOOS-IOS, data are also being collected through a number of GOOS pilot projects. Those that are currently active include

- Baltic GOOS (BOOS) (a EuroGOOS regional project)
- MFSPP (Mediterranean Forecasting System Pilot Project)
- PIRATA
- GODAE
- Argo
- RAMP (Rapid Assessment of Marine Pollution)

4.1.4 Regional Programmes

In addition, GOOS is being developed by a number of regional groups that will produce data, information and products. These include (A = active; U= under development):

- NEAR-GOOS (north-east Asia Region) (A)
- EuroGOOS (Europe)(A)
- WIOMAP (Western Indian Ocean Marine Applications Project) (U)
- MedGOOS (Mediterranean)(U)
- IOCARIBE-GOOS (Caribbean and Gulf of Mexico)(U)
- PacificGOOS (Pacific islands)(U)
- SEACAMP (S.E.Asia Centre for Atmospheric and Marine Prediction)(U)
- Black Sea GOOS (U)

There will not be a one to one correspondence between these sources of data and the end-to-end data management systems that process and deliver the data. Nor will there be a one to one correspondence between the data management systems and the data collection systems. For example the GTSPP processes sub-surface profile data from the SOO network, from PALACE floats, from the TAO and PIRATA networks, and from drifting buoys.

4.2 The IOS for Satellite Data

An integrated programme of in situ and remote sensing data collection is necessary to the success of the GOOS. The requirements for satellite data were originally defined by the Ocean Observing System Development Panel and updated by the Ocean Observing Panel for Climate. These requirements for the GOOS Climate Module are included in Annex C.

The partners for an Integrated Global Observing Strategy (IGOS-partners) have specified in the IGOS Oceans Theme document (Appendix 1) what remote sensing elements they consider as commitments to global observations. These elements comprise a mixture of research and operational missions, not all of which would be considered integral parts of the GOOS IOS, which is a matter for the GOOS Steering Committee to examine.

The IOS for satellite data will consist of those satellite missions that produce or will produce the data specified in the requirements in the table in Annex C. Because of the long lead-time in the development and launch of satellite missions the IOS for the next few years already exists or has been planned. If there are changes in the planned missions these changes are likely to be launch delays or cancellations. Therefore it should be possible to document the relevant missions and assemble sufficient information to plan the requirements for satellite data management systems in GOOS. For purposes of this suggested task it would be useful to have continued access to the CEOS/WMO database of requirements and capabilities of environmental satellite systems. It is important that this database be adequately maintained and kept up-to-date.

As discussed in section 6.6 it is proposed that initial ideas for the management of satellite data in GOOS be developed based on the requirements and experience of GODAE. Since the table in Annex C has been designed with GODAE in mind, the task suggested in the previous paragraph should not conflict with GODAE.

There may also be a need to develop a catalogue of historical satellite data sets and products useful to GOOS.

5.Access to Data and Information, Dataflow Monitoring

This section on providing access to data and information and dataflow monitoring incorporates three elements of importance to GOOS data and information management. Section 5.1 discusses the Global Observing Systems Information Centre (GOSIC). GOSIC is designed to provide a central point of access to information and data of the programmes of the three global observing systems.

Section 5.2 describes the GOOS policies in regard to data sharing and release and Section 5.3 describes the functions of the data coordinators who will facilitate the tracking of data and information to ensure the data and products are available according to planned schedules.

5.1 Single Point of Access to Data and Products through GOSIC

GOOS, as well as the other two global observing programmes has already identified centres that will serve as a core for implementation of the G3OS data and information systems. Others will be identified in the future. Data, data products, and information for the G3OS will be managed using a highly distributed system involving dozens of organizations, hundreds of databases and data and information products, and thousands of users.

Having information about the participating centres, their holdings, and their services easily available will stimulate their use, facilitate access to G3OS data and information, and will encourage other data centres to join the system.

A single entry point for users for this data and information best serves the distributed nature of the proposed GOOS DIMS. The Global Observing Systems Information Centre (GOSIC) has been designed to serve as such a centralized point of entry for information on the G3OS programmes, and to provide access to G3OS data.

GOSIC provides basic user services, explains the G3OS data system, and provides an overview of the data and information that are available. In addition, the Information Centre offers a search capability, optimized for G3OS data and information and facilitates access to the worldwide set of observations and derived products.

GOSIC includes the following four major elements.

A Data Access Module

GOSIC provides a detailed view of and on-line access to the end-to-end data and information management systems and the databases that are the basic elements of the three observing systems. Examples are the GCOS Surface Network (GSN), the Global Temperature and Salinity Profile Project (GTSPP, of GOOS) and the GTOS Thematic Network - Ecology (GTN-E, of GTOS). These systems are generally operated by a number of agencies that contribute various data and information management services according to their capability and expertise. The sum of the contributions add up to a carefully designed end-to-end system that provides all the necessary steps from collection of data through to data and information product generation, access, distribution, and archival. In fact these end-to-end systems will in some cases intersect as a copy of data from one system at one point in the processing may be passed to another system for its use.

A Meta Data Directory

GOSIC provides a searchable on-line meta data directory for databases that are part of the observing systems. Specifically the database holds directory interchange format (DIF) records

- for historical databases that are judged relevant to G3OS requirements as identified by the science panels.
- for databases that hold and provide data being collected by the operational observing systems.

The meta data directory is compatible and will be maintained compatible with the NASA Global Change Master Directory (GCMD) and will be hosted by NASA on behalf of GOSIC. This will provide access to the broader community of both GCMD and G3OS users. NASA will provide a GOSIC "view" by limiting GOSIC queries to G3OS entries.

To ensure up-to-date information in the directory, procedures are under consideration to have centres holding the databases maintain the DIF records on their sites for periodic harvesting and automatic processing. Hyperlinks in the DIF record will provide access to data for browsing, ordering, or downloading; and for access to associated products.

The development of the GOSIC meta data directory is being coordinated with development of the IODE Marine Environmental Data Inventory (MEDI) directory that is also based on the NASA GCMD format and standards. MEDI is being developed as an offline input tool that can be used by the data centres to prepare and maintain the meta data records on their site. Upgrading of the MEDI software to be Web compatible is also under consideration so that it can serve as a general purpose application for on-line filtering and mapping of data for GOSIC and IODE.

Data Flow Diagrams for the Operational Observing Systems

For each end-to-end data management system the data flow diagram

- shows the data flow from initial assembly to final archival
- identifies the participating centres
- links to a web page maintained in the participating centre that;
    a) describes the G3OS responsibilities of the centre,
    b) provides complete contact information,
    c) provides, in a standard template, hyperlinks to download data or submit an order, view or download data products, view or download inventory information, and
    d) provides access to data flow information for evaluation of system performance.

Programme Information on the Observing Systems, Observing Requirements, and Development of the Observing Systems.

GOSIC also contains information on

- G3OS programme planning
- Data management planning
- Observational requirements as defined by the science panels
- Specifications for quality control and analysis of the observations

GOSIC is designed to:

- Provide for searches for data and information across all participating G3OS data centres using the Internet.
- Return results regardless of the data format, or where the data are located.
- Provide results back in a standard easy-to-read, easy-to-understand format.
- Allow users to determine the type, quality, volume, and distribution of the data by linking to a web page provided by the participating data centres.
- Allow users to order or download data sets.

The Information Centre is not a repository for data. Rather, it maintains a database of meta data on the data sets that are available in the three programmes. Hyperlinks point to the data centres where the data and information can be obtained. GOSIC does not create or modify the presentation of data. To the extent that it is possible, the programme that is the source of the data and information is identified.

GOSIC serves the three observing systems commonly referred to as the G3OS. This includes the Global Climate Observing System (GCOS), the Global Ocean Observing System (GOOS), and the Global Terrestrial Observing System (GTOS). The development of GOSIC is overseen by a review group that has been selected by the Secretariats of the three observing systems. This group looks at progress in the development of GOSIC on a semi-annual basis and provides guidance to the Secretariats.

5.2 Data Sharing, Access and Release

One of the primary GOOS principles (D7) deals with managing, processing, and distribution of data. Specifically,

> "In concert with the policies of IODE, IGOSS (now JCOMM) and GCOS, and following the data management plan for the World Weather Watch of the WMO, commitment is required by GOOS participants to establishing, maintaining, validating, making accessible, and distributing high quality, long term data meeting internationally agreed standards. Preservation of GOOS data is required in suitable archives following appropriate procedures and criteria for data acquisition and retention, and should include information about data holdings. Data should be processed to a level that is generally suitable for the generation of operational products and for research, and described in internationally accessible on-line computerized directories that can also be made available by other means. GOOS contributors are responsible for full, open and timely sharing and exchange of GOOS-relevant data and products for non-commercial activities. Exchange implies that donation by individual nations gains access to data from others as well as to products derived using all available data, such that the benefit of cooperation exceeds the cost."

GOOS has also established guiding principles for data sharing.

a) The data obtained by GOOS will be the most useful to the most people if there are no periods of proprietary holding, nor any restrictions on to whom or when the data are disseminated.
b) The second principle is that the quality must be assured, which may mean some delays for some of the data.
c) The third principle is that national and personal interests and needs are part of the overall equation, so it may not be possible to apply the first principle totally in all cases.

It may not always be possible to hold to the first principle. Some research data sets will not be part of GOOS, if the investigators involved are not able to release the data in a timely way for any reason. GOOS does not plan to "publish scientific papers" based on the GOOS data, but rather to see that end users such as coastal zone managers and ship routers can get the data and products they need, when they need them. GOOS needs for data, and a scientist's need for the data, may be quite compatible. GOOS may be able to obtain access to scientific data for compilation into products provided the data are not released in original form for a specified time.

Both scientific and operational agencies are becoming more aware of the need to make data available quickly even if it must be released in a preliminary form. One of the objectives of the CLIVAR data management system is that CLIVAR data should be assembled and distributed swiftly and that all data, model output, data assimilations, and re-analyses should be made accessible to the entire climate community (free and open exchange).

CEOS has adopted principles on satellite data exchange in support of operational environmental use for the public benefit, and for support of global change research. These principles recognize that environmental data should be made available to users who need them within time scales compatible with the requirements.

Operational environmental use for the public benefit is defined as

- Use of data to provide a regular environmental service for the public benefit;
- Carried out by public national or international Earth observation agencies, or other entities designated by governments or public authorities, to support public benefit mandate;
- Examples include use of data to carry out a mandate of environmental observation and prediction or missions relating to environmental management or regulation.

The last point is of particular interest because it applies the term operational to activities with much longer time scales than are usual. Environmental management and regulation is a process that takes place over months to years rather than hours to months.

5.3 The Data Coordinator

The JDIMP data and information management model (see Annex A) identified two overarching functions. The first was the information centre (GOSIC). GOSIC provides services to the G3OS user community. GOSIC also assists the observing systems in managing all aspects of the data collection by providing information to support the second overarching function, the evaluation and feedback function. The mechanism for effecting the evaluation and feedback function is the data coordinator.

Experience with the distributed data systems has already demonstrated the need for a "data coordinator" to ensure the data systems work as designed. The University of Delaware operated the "Data Information Unit" (DIU) which had the responsibility of implementing a Web based information system to report the status of data collection for WOCE. The persons who reviewed reports on progress in data collection and contacted the principal investigators (PIs) by various means to obtain information on the status of the data collections programmes and then fed information to the databases in the information centre were termed the data coordinators. The WOCE International Project Office also assigned resources to work in cooperation with the DIU on data coordination.

The GCOS JSTC (reference the 1998 meeting) noted the WOCE experience and agreed that GCOS, GOOS, and GTOS were likely in due course to need data coordinators (referred to in the meeting report as "International Data and Product Facilitators. Note also that JSTC has recognized there will be a requirement for more than one.).

The Data Coordinator is the key to the evaluation and feedback function for the GOOS data and information management system. For WOCE the function has been carried out mostly by reading reports and manually developing tables of data status for installation on the Web server. This effort involved several people part time for WOCE. Data coordination for GOOS will include much more in situ data, satellite data, data products, and model data. Thus the volume of work makes it necessary to automate the data coordination function of GOOS. Fortunately the requirements are such that automation of many of them should be straight forward provided the GOOS operational or research centres accept the responsibility of generating on-line files of the necessary information.

The data coordination operation will have the greatest chance of success if the coordinators are located in an operational centre that has access to the networks carrying the data flow and is involved in the capture and processing of at least some of the data on behalf of GOOS.

The duties of the coordinator might include:

- Monitoring, analyzing, and reporting on the data flows including modeling and satellite data with the help of automated systems and data flow information in digital form provided by participating centres.
- Specifying data flow monitoring products that are necessary to automate the data flow monitoring function and working with the participating centres to effect their implementation.
- Facilitating the work of the science panels in developing a comprehensive set of products by identifying potential data products that are already being made available by centres and making recommendations to the appropriate panel for their consideration as GOOS products.
- Providing information documents to be made available through GOSIC. The documents would be related to statistics of data flows and analyses of observing system performances, products, and the production of products.

The activities of the data coordinator in relation to the development of automated data flow analyses and to the proactive role in the development of products are extensions to the duties of the present WOCE data coordinators.

6. DIMS Planning, Cooperation with Other Programmes

Data and information management planning for GOOS is done at two levels and will have to be coordinated internationally with a number of other organizations and programmes. The two planning levels are

- the GOOS science panels that identify scientific aspects and products for the data management systems, and
- the IOC and WMO data management programmes operated by IODE and JCOMM.

The requirements for coordination with GCOS, GTOS, and with the global oceanographic and meteorological experiments are discussed in section 6.3 below.

GOOS data management activities need to be coordinated with those of the existing bodies in IOC and WMO, primarily JCOMM and IODE. These bodies have responsibilities for ocean and marine meteorological data management. In fact several programmes included in the IOS for GOOS for in situ data are operated by these existing bodies. This cooperation is discussed in section 6.2.

The Global Observing Systems Space Panel (GOSSP) and the Committee on Earth Observation Satellites (CEOS) are discussed in section 6.4 below. GOSSP had a major role in coordinating the general requirements of the G3OS with the space agencies but has completed its work and is disbanded. This responsibility is now dealt with directly between the G3OS and CEOS. GOOS has chosen to appoint a rapporteur to facilitate this coordination.

6.1 The Science Panels

The responsibility for specifying and developing the detailed aspects of GOOS data management rests with the science panels. This includes the elements of the end-to-end systems that must be developed for each data flow, the specifications for the analysis and quality control of the data, specifications for meta data, the identification of final archives, and the products that are needed for users and managers of the systems. The expertise for this task rests in the science panels.

The responsibility of the science panels to specify both useful and essential data products that need to be developed to meet the needs of end users and operators of the observing systems is of particular importance. This requirement must be addressed from the beginning. It is what defines the resulting systems as end-to-end. It provides the rationale so that as systems develop and evolve, all the necessary elements are put in place for product generation. GOOS can not be successful unless products are prepared and delivered according to specifications and according to a predefined schedule.

The science panels can also be considered to have a role to moderate between the users and the observing systems for specifying data products. This will ensure the relevance of the products. For this responsibility the science panels need to consider adequate participation of users on each panel.

The science panels, of course, also have the responsibility to identify the scientific questions that must be answered, to design an observation plan to collect the necessary data, and to identify suitable instruments and technologies to make the measurements.

Implementation of the specified systems will be accomplished by partnering with existing organizations or through other international mechanisms as discussed in section 6.5 below.

6.2 Planning and Coordination with JCOMM and IODE

The Joint Commission for Oceanography and Marine Meteorology (JCOMM) was created by merging the programmes of the Joint IOC/WMO Committee on IGOSS and the WMO Commission on Marine Meteorology (CMM).

It was recommended that this new joint body be responsible for the development, implementation, and maintenance of operational data collection and dissemination systems to meet the needs of traditional clients, the needs of GOOS and GCOS, and the needs of global science experiments such as the World Ocean Circulation Experiment (WOCE), and CLIVAR. Concentration was to be on systems that deliver parameters for which a defined need has been established for global or regional programmes, and for which some international standards have been agreed for both coverage and accuracy. There will be a heavy work programme for the new body in responding to GOOS and GCOS and it will have to concentrate its work on the priority areas in order to be effective.

IODE is a long-standing programme of IOC that deals with data on a historical time scale rather than operational. It serves a complex multidisciplinary community with interests that are research based. Much of the data are collected for local and regional purposes, and have limited geographical or temporal coverage in terms of GOOS applications. However IODE and its data centres have broad expertise in working in all areas of geophysical, biological, chemical, and physical oceanography and in standards for exchange and archival of these other data types. There will be cooperation between IODE and JCOMM in building and implementing end-to-end data management systems and cooperation between GOOS and IODE will also be beneficial.

Both IODE and JCOMM have regular sessions that discuss needs for international data exchange and management, and services to users. Both expect national centres to volunteer to undertake data management activities at national expense on behalf of the international community. There is a requirement for GOOS and GCOS, and the international programmes such as WOCE and CLIVAR to coordinate and plan activities with IODE and JCOMM to the benefit of all.

6.3 Coordination of GOOS Data Management Planning with the G3OS and CLIVAR

Data management planning for the G3OS began in GCOS. The Data and Information Management Panel (DIMP) was established in 1995 to formulate and develop the GCOS Data and Information Management Plan, monitor the overall implementation of the plan, and to make reports and recommendations to the GCOS Joint Scientific and Technical Committee. Later the responsibilities of the GCOS DIMP were expanded to include data and information matters for GOOS and GTOS and it was renamed JDIMP (Joint Data and Information Management Panel).

JDIMP was specifically not charged with the development and implementation of the detailed data and information systems for the three observing systems. JDIMP had an overarching responsibility to provide general guidance to the scientific panels engaged in design and implementation. However JDIMP also served as an agent of the broad, multi-disciplinary environmental information user community by facilitating the development of end-to-end procedures that foster acceptable levels of data and information integrity, standardization, continuity, and accessibility.

The JDIMP data management plan addresses the higher level aspects of data management for the G3OS. It is available on line. (http://www.gos.udel.edu/publications/jdimp_plan.htm)

JDIMP also developed a generic data management model that is included here in Annexes A and B.

GCOS, GOOS, and GTOS will each have a more detailed plan of data management aspects of their observing system based on the work of the science panels. This document is the detailed plan for GOOS. In addition programmes such as CLIVAR will require the development of its own data management plan as its data collection requirements will not completely overlap those of GOOS and GCOS. Of course CLIVAR data management can not be developed separately from the G3OS plans because of the significant overlaps in data requirements and the necessity to share the data management burden.

6.4 The Committee on Earth Observation Satellites (CEOS)

All three of the observing systems will be heavily dependent on observations from satellites as well as in situ measurements. The science panels specify the requirements for observations including both satellite and in situ data. The Global Observing System Space Panel (GOSSP) was established to monitor evolving requirements for and capabilities of space observations and to coordinate these requirements with the space agencies, but has now been disbanded.

Since GOSSP is no longer providing this interface with the space agencies, it is essential that GOOS and the other observing systems develop appropriate links with the Committee on Earth Observation Satellites (CEOS). Satellite data is essential for the success of GOOS and will without doubt provide a large volume of data for all three components of the G3OS. The CEOS has led the development of an Integrated Global Observing Strategy (IGOS). This strategy is being fleshed out through an international partnership of several international organizations, including the space agencies, funding agencies, the scientific community, and international research programmes tasked with planning and coordinating global environmental observations from space and in situ. The goals of the IGOS Partners include avoiding unnecessary overlaps, addressing gaps in observations, linking activities with complementary observation programmes, and above all implementing effective data management and the provision of pertinent and useful data products on a regular basis.

The IGOS Partners have decided to test and demonstrate how the IGOS approach should work by focusing on selected themes. The Partners have established a number of theme teams whose task is to develop strategies integrating space-based and in-situ observing programmes so as to produce products meeting users' needs. The first such theme is the Oceans Theme, which was published by NASA in January 2001. It provides a means for the managers of major ocean observing systems to commit to working together towards a common goal. The IGOS Partners have also prepared a report on Data and Information Systems and Services, which has led to the development of a set of DISS Principles regarding matters related to the end-to-end process from collecting data to providing products and archives. These thematic strategies and principles are all extremely relevant to GOOS, and have been developed in collaboration with the GOOS community. It is important that GOOS and those leading the further development of GOOS data management systems develop effective links with CEOS and the IGOS Partnership process.

6.5 Implementation of GOOS D&IM through Partnering

As discussed above the initial implementation of GOOS data and information management activities is being accomplished by connecting a number of existing data management systems such as GTSPP and GLOSS. These existing data management systems now include more than data centres and organizations that have taken on voluntary responsibilities on behalf of IOC and WMO. Several of these centres have been funded under research budgets as part of WOCE and TOGA and some will transition to CLIVAR. If the work of these organizations is not continued at the end of WOCE or is not taken on by another organization, then some of the end-to-end systems in the existing GOOS IOS will have gaps in their services. GOOS will have to address means by which to continue the work of centres and bodies that may cease operations at the end of WOCE, or find other centres to take on the activities in order to maintain the IOS.

GOOS will also require a mix of scientific organizations, operational agencies, and IOC and WMO centres to develop and implement the new data systems. This mix of partners is needed because the scientific and data management expertise must be brought to bear in concert if a successful result is to be achieved. In many cases the scientific centres or the data management centres are doing most of the required work anyway and it only remains to coordinate activities to implement the larger system.

GOOS may be thought of as an operational programme but it will need a lot of scientific input for some considerable time into the future. This is also true of the existing IOC and WMO data management programmes because of the scientific basis of existing services. DBCP and GLOSS have their scientific experts now to guide the development and implementation of their data programmes and will continue to need them.

It is in the interests of research programmes such as WOCE and CLIVAR to partner with GOOS in the implementation of end-to-end data management systems. Such participation may involve them in additional work in evaluating the quality of the data, producing products for users on a regular basis, or providing meta data to the system. However they will gain by having data centres relieve them of assembling data sets from diverse sources, formatting the data to a common agreed format, and applying agreed standardized quality control and duplicates management procedures. The result should be more effective data management with better data sets delivered sooner, and better services and products for research clients and for all other clients.

GOOS partnerships for the initial operating system include working with IODE and JCOMM. JCOMM includes programmes such as DBCP, GLOSS, etc. Not all aspects of these programmes are relevant to GOOS nor will GOOS programmes completely meet the needs of these other programmes. Each programme will benefit from having access to additional data relevant to their needs, and all will benefit from not having to develop and implement parallel systems.

Of course national data centres or other centres with other international responsibilities will continue to produce analyses and products from their own data, including GOOS data, for their national or existing international clients. These products would not carry a GOOS endorsement or label. However it is hoped that data centres using GOOS data or products for national and other international requirements will acknowledge the source of the data or the product.

As discussed above it is essential that as GOOS data management develops a partnership with CEOS and IGOS is actively sought and developed to facilitate joint planning, coordination and effective cooperation.

6.6 Implementation and Testing of GOOS D&IM through the Pilot Projects

GODAE is a major pilot project of GOOS It is intended to provide a practical demonstration of real-time, global ocean data assimilation in order to provide increasingly precise descriptions of the ocean physical state at high temporal and spatial resolution consistent with a suite of space and direct measurements and appropriate dynamical and physical constraints. The utility of the resulting products to the user community will be a major focus.

The main scientific objectives are:

- The application of state-of-the art ocean models and assimilation methods for short-range open-ocean forecasts, for boundary conditions to extend the predictability of coastal and regional subsystems, and for initial conditions of climate forecast models.
- To provide global ocean analyses and re-analyses for developing improved understanding of the oceans, improved assessments of the predictability of ocean systems, and as a basis for improving the design and effectiveness of the global ocean observing system.

GODAE is founded on the belief that such a demonstration is vital to the realization of a permanent, global ocean observing network and prediction system, with all components functional and operating on a global domain and delivering useful products in a timely manner. GODAE is first and foremost dedicated to the development of global systems. It is not a research programme, but a practical test of our ability to deliver useful products in a timely manner using state-of-the-art scientific methods. The focus is on the ocean and operational oceanography, not on the coupled ocean-atmosphere or land-ocean systems; but GODAE places great emphasis on links (coastal-open ocean coupled ocean-atmosphere for climate; physical-biological) as applications as the basis for GODAE products. GODAE provides a basis for these extensions.

Planning for GODAE is already underway. The experiment will run primarily in the year 2003 to 2005 time frame. Data management for GODAE will involve acquiring and merging satellite data and in situ data. It will also involve the generation and distribution of data sets from numerical models.

Because of the broad range of technologies, data, and data applications GODAE provides an excellent opportunity to further develop the requirements for management of satellite and numerical model data. Section 4.1.3 contains a list of other pilot projects of GOOS. The development of the pilots should be used to specify and further develop data management requirements particularly in the case of new systems for coastal seas. This is an efficient method of development because actual needs, data characteristics, and volumes will become apparent during development of the pilot project. There will be no guesswork involved.

7. Archiving Practices and Standards

This section is intended to provide some guidelines and information about recommended archiving practices for long-term storage of GOOS data, information, and meta data. The network of World, Regional, and National Oceanographic Data Centres will provide a major component of the long-term archives. Programmes lacking long-term archiving capability will have to take full advantage of these well established systems.

The purpose of the final archives is to preserve the data for the long term for future users. The World Data Centre (WDC) System is an example of such a final archive, and much of the GOOS data will probably be archived there. In cases where a WDC does not exist or is not appropriate for a particular data set, then an alternate permanent archive must be found. Usually this will be a government centre that is prepared to accept the long term responsibilities.

There will be some specific requirements placed on centres for long term archival of GOOS data for future users. For example GOOS data will in general have been collected and analyzed to a higher standard than other data. Therefore it should not be merged with other data if its GOOS identity is to be retained.

The minimum responsibilities for a permanent archive are generally that the archive agrees to:

- accept the data and all available supporting meta data,
- store the data either in their original form or in a form from which all the original data and meta data can be recovered

- refresh or update the medium on which the data and meta data are stored so that both are readable in the future, and
- provide the data and all supporting meta data to users on request, free of charge, or at the cost of reproduction.

In addition, there is an additional minimum requirement for GOOS data sets. This requirement is to

- store the data so that they can be distinguished and recovered separately from similar data that are not GOOS data (and therefore have not necessarily had the same care taken in collection, quality control, and processing).

Final archives need to be identified for all GOOS data sets in establishing the supporting end-to-end data management systems. The scientific panels should identify the meta data that is to be collected and stored with the data in the archives.

Based on the experience of the international science programmes, much of the data, special analyses, and the more important data products will periodically be issued on CD-ROM. The CD-ROMs will usually contain software for viewing and extracting the data, analyses, and products. This will simplify the work of the archival centres in that for many users they only need provide a copy of a CD-ROM set. The archive centre may not have to deal with requests for data on a regular basis either. Usually the centres participating in the programme will be maintaining databases on a long-term basis as data sets are built up. These centres will service requests and data subscriptions for users. Data will be stored in permanent archival centres to guard against loss and to service users after the original centre has terminated operations.

8. Capacity Building Programmes

A paper on "Data Management And Exchange Using GOOS As a Basic Example" was presented at the joint IOC/IGU seminar "The Role of Ocean Science and Geography in Facing Ocean Management For the 21st Century" (Sagres, Portugal, September 3-5). In relation to data management and capacity building the following points were made.

> "*Given that the goal is evolution of a comprehensive, and effective, high quality global data and information management system operating in all coastal countries, developed and developing alike, a considerable amount of effort will have to be devoted to building the capacity of the developing states so that they can contribute to and benefit from GOOS.*
>
> *Capacity building for improved data and information management requires the following main elements:*
>
> *(a) the building of underpinning institutions, such as National GOOS Coordinating Committees to promote GOOS, build awareness, make plans and attract funds;*
> *(b) the building and resourcing of underpinning infrastructure elements, such as National Oceanographic Data Centres;*
> *(c) developing mechanisms to ensure access to all appropriate data and information from in-situ and space-based systems;*
> *(d) building a data rescue programme to ensure that best use is made of what has already been collected.*
> *(e) recruitment and training of skilled manpower to manage data effectively and to use it to create information products valuable to the user community;*
> *(f) the building of comprehensive linkages to the user community to ensure its needs are met.*"

Most of the actions suggested in this text have significant data management aspects. The following paragraphs suggest a context for developing the data management capacity that is suggested.

Capacity building for data management capabilities should more often be managed within an end-to-end system. In other words perhaps a capacity building programme for GOOS data management might better be built around an identified requirement that is important to a group of developing countries rather than giving the data management training in isolation from an application. The data management aspect would be one element of implementing an end-to-end system with a defined and needed product, and the data collection, processing, and analysis to produce it.

The data management training could be given to more project personnel so that if some employees move on the knowledge would remain. The GLOSS programme offers this sort of capacity building. IODE also has a successful programme in capacity building. The IODE programme could more often operate as part of an end-to-end system for direct application to a product provided the opportunities are identified and the response is jointly planned.

GTOS has an interesting approach to the problem of helping developing countries to benefit from the international systems and to become contributors of data to the system. The idea is to start supplying them right now with data (CD-ROM sets and satellite data sets) from a system they are interested in so they can begin to experience the benefits of participation.

A workshop is needed on capacity building related to data management within end-to-end systems. Participation should be by invitation to international organizations that have data programmes in developing countries. (e.g. IODE, GLOSS, UNEP, IGBP, World Bank, etc.) The best ideas and experience of everyone could be captured for GOOS.

9. The Way Forward

The "Way Forward" for GOOS data and information management consists of a two-pronged approach. Firstly, further development and implementation will focus on the IOS and the pilot projects as discussed earlier. This will proceed in cooperation with the existing data management systems of IOC and WMO, primarily the subsidiary bodies IODE, JCOMM, and CBS, and with various participating national activities. Secondly, an ocean data and information technology project has been initiated by the GOOS Steering Committee to play a lead role in developing and implementing improved data management technologies for future GOOS requirements.

9.1 Use of the Existing Systems

The point has been made many times that the existing data management systems of IOC and WMO will be used for GOOS and GCOS. The options were to build a complete new system from scratch, to use the existing systems as they were, or to use the existing systems augmented with increased participation from other national organizations and with additional cooperative programmes with other international organizations. The option of a new system was not possible or reasonable because of unnecessary duplication of effort and it would take a decade of meetings and work at least to get operational. Using the existing systems as they are was not an option because there is not sufficient capacity and expertise to take on the additional complex and large workload required by GOOS and GCOS. The workable option was to build on the existing systems by soliciting increased national participation to increase capacity and expertise, and to share the workload by seeking cooperative programmes with other international programmes that manage ocean data.

The existing systems of IOC and WMO that are mainly concerned with ocean data management are managed by IODE and JCOMM. JCOMM has been created recently by combining the Joint IOC-WMO Committee for IGOSS and the WMO Commission on Marine Meteorology. One of the more important reasons for forming JCOMM and one of its main responsibilities was to work with GOOS and IODE to design and implement the operational data management systems for GOOS and GCOS.

In developing the sub-structure for JCOMM, it was proposed that the IGOSS-IODE Steering Group on End-to-End Data Management become an IODE-JCOMM group with more or less the same responsibilities. IODE and JCOMM will be having a meeting of experts on this subject to define the scope of its responsibilities and in the light of this to review and revise as necessary the terms of reference. Depending on the results, then both IODE and JCOMM may have to revise their sub-structures further to deal with their GOOS and GCOS responsibilities.

It is recommended that both IODE and JCOMM be encouraged to proceed with these developments and to implement strong and effective sub-structures that can take and deliver on their responsibilities for supporting the implementation of the data management aspects of the GOOS and GCOS observing systems and pilot projects.

To accomplish this mission these sub-groups will then work with the individual scientific development groups guiding the development of the components of the IOS and the pilot projects to;

- review the existing components of the IOS to bring them up to the standards of the GOOS D&IM Plan,

- develop the data management systems for the pilot projects and manage their conversion to permanent programmes when and if appropriate,
- identify and utilize existing capabilities and capacities in agencies that can take on more work of a similar nature with little or no additional systems development or operational costs,
- identify and recruit additional national agencies with the capacity and expertise to carry out identified responsibilities in effecting the various data flows,
- identify opportunities for mutually beneficial cooperative programmes with other international organizations and negotiate joint programmes,
- keep the evolving GOOS data management system under review to ensure it continues to function as designed.

It should be noted that the lead for defining the data management requirements including assembly, QC, analysis, product generation, archival; and all time scales lies with the scientific developers. The responsibility of the data managers is to implement the scientific designs. The scientists and data managers would cooperate on identifying and recruiting other national agencies and seeking cooperation with other international programmes.

The responsibilities of the proposed IODE-JCOMM sub-group could be viewed as what might have been expected of a body such as a GOOS data management committee tasked at an operational level and not a policy level. In the present situation it is expected that having both the IODE-JCOMM group and a GOOS DMC would be less effective than having just the IODE-JCOMM group because of overlap, cost of two sets of meetings, and difficulty of coordination.

9.2 A GOOS Initiative for Development of Improved Data and Information Management Practices

The GOOS SC has examined and approved a proposal for an ocean data and information technology project. The purpose of the project is to address the shortcomings of existing ocean data management systems in a range of technical areas such as telemetry capacity, data assembly practices, and uptake of modern information technology, as well as a lack of sufficient intellectual engagement and commitment of resources to build the necessary highly efficient and effective systems.

The project will address these data management issues on a GOOS-wide front by conducting a suite of studies of existing systems and 3rd party systems to evaluate capability, properly scope the total requirements for all aspects of GOOS D&IM, and evaluate available and emerging technologies for applicability to GOOS requirements. The studies will include intermediate workshops and will lead to a working conference to seek consensus on an action/implementation plan for improved GOOS data and information management.

Annex A

The JDIMP Data Management Model

The JDIMP has developed a generic model for end-to-end data management systems for the G3OS. This model provides for all the functions necessary to build end-to-end data management systems that incorporate the features that have been discussed in the WOCE data management systems.

In figure 2 below, services are provided to users from all of the boxes in the diagram. The science boxes provide advice and guidance. The data management boxes provide data sets, data products and analyses, and supporting information.

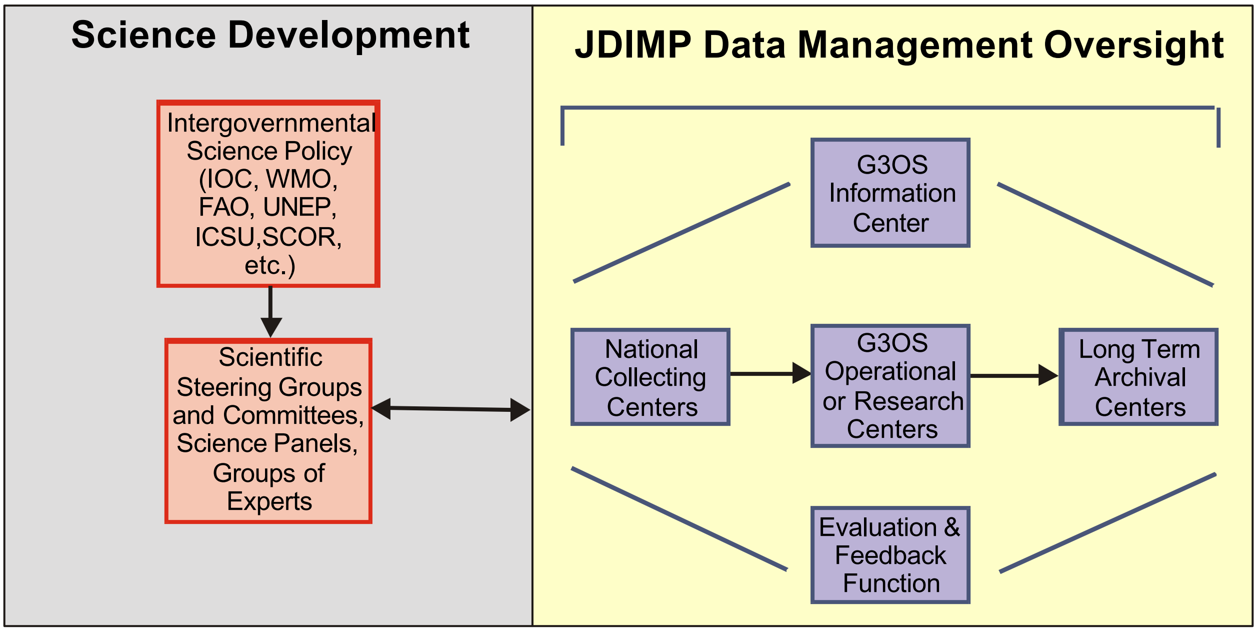


Figure 2. The JDIMP Generic Model for G3OS End-to-End Data Management Systems

In the figure the organizations under the heading "Science Development" are responsible for developing the science requirements and the data management requirements. These data management requirements specify such things as

- analysis and quality control techniques,
- meta data requirements for understanding and interpreting the data,
- time frames for delivering the data,
- data sets and data products required at the various stages of the data flow, and
- the location of the final archive for the data for future users.

The organizations under the heading "JDIMP Data Management Oversight" are the organizations that acquire and manage the data in the manner specified by the science groups.

- National Collecting Centres acquire the data from national sources, perform any necessary preliminary processing and forward the data to G3OS Operational or Research Centres.
- G3OS Operational or Research Centres perform an agreed function on the data and pass it on to another such centre or the final archives. For any one data flow there will usually be more than one of these centres and there could be several.

- The Final Archives is an organization that agrees to hold the data and make it available to future users. The archive centre will also have to accept certain responsibilities in its handling of G3OS data as described in a later section.
- The information centre (GOSIC) is one of two overarching functions of G3OS data management. This centre provides a comprehensive one-window view on all aspects of G3OS data collection so that users can find and understand the data they need and access it at the appropriate stage of its life cycle.
- The evaluation and feedback function is a distributed function that includes several types of feedback. The purpose of this function is to ensure that the end-to-end systems are working. Feedback here includes information on data flows and quality from national centres and Operational or Research Centres, the data coordination function which reviews overall performance and availability of data, ad hoc studies of the quality of the data, etc.

More detailed descriptions of the various boxes in the JDIMP model are included as Annex B.

Annex B

The JDIMP Model for End-to-End Data Management Systems

1. The Science Policy Function

The science policy function for purposes of the G3OS are at the intergovernmental level and are the responsibilities of bodies such as SCOR, IOC, WMO, ICSU, etc. These bodies identify the need for and sponsor cooperative international science programmes such as the WCRP and IGBP. Bodies such as these have also specified the need for and are sponsoring operational data collection programmes to meet the needs of present and future international scientific and operational programmes. Specifically this function has defined the need for GCOS, GOOS, and GTOS.

2. The Science Development Function

The development of the science aspects of the international science programmes and supporting global observing systems are accomplished through the vehicles of scientific steering groups (SSGs), steering committees (SCs), science panels, and groups of experts. These various groups are composed of nominated scientists in the fields relevant to the observation systems being developed or managed.

For G3OS data management the groups that are defining the requirements for data and data management are the scientific and technical advisory panels. There are several such panels at the moment.

- Atmospheric Observation Panel for Climate (AOPC)
- Ocean Observation Panel for Climate (OOPC)
- Terrestrial Observation Panel for Climate (TOPC)
- The Coastal Ocean Observations Panel (COOP)

The science panels of the G3OS are responsible for defining the observations required in the future for the observing systems to achieve their goals, for identifying the historical data sets that are relevant and important, and for specifying the meta data to be collected and maintained with both the historical and new data. There are two types of meta data. The meta data referred to here is that required for a user to understand the limitations and circumstances of the data in order to use it properly. For example this type of meta data would include instrument type, method of analysis, and quality flags for the data points. The other type of meta data is the data set meta data which would include such information as formats, holders of the data, and the media on which it can be made available.

The science panels are also responsible for specifying at least the necessary elements of the end-to-end data management systems to be developed to process, quality control, deliver the data to users, and provide for final archival of the data and meta data.

3. National Collection Centres

Data for international programmes are collected by national activities. The collection centre function for the G3OS is generally a national activity. The data assembly centre collects together specific data and meta data from national activities for one or more data sets identified by the scientific and technical panels as required for G3OS purposes. Generally speaking, a nation will collect data to a standard that meets its national needs. This national standard will not necessarily meet the needs of a programme such as GOOS, GCOS, or GTOS. For example, for detection of climate change it may be necessary to have a period of simultaneous operation for calibration purposes between old and new sensors at a station. Additional meta data may also be required. Thus for the G3OS, the national collection centre may be asked to undertake special processing or handling for G3OS data.

In any case, the collection centre will have certain responsibilities for assembling national data and meta data, for processing the data in a certain way, and for forwarding the data to one or more international centres for further assembly, processing, distribution, and archiving according to the details of the end-to-end data management plan for the variables in question.

One country may have several collection centres for the various components of the G3OS, and for the different variables being collected. However, it is not necessary for each country to have its own collection centres. For example it would make sense for groups of developing or developed countries with common interests to share collection centres according to their capabilities and expertise.

4. G3OS Operational or Research Centres

The next required function in G3OS data management is the G3OS Operational or Research Centre. This is an organization that takes on responsibilities for specific activities in support of the G3OS or G3OS users. The activity may be a special processing function, acquisition and quality control of certain data, generation of specific products, monitoring of a data flow, or compilation and distribution of specific data sets or data products to selected users.

The need for the various Operational or Research Centres will be identified by the Science Panels and will be based on the processing, quality control, analyses, etc, that are required to prepare the data sets, generate the data products, and deliver them to the users. Data will often be moved by electronic network between a number of Operational or Research Centres carrying out different steps in the processing. It is more effective to move the data to the place where the necessary expertise and capacity exists to perform a function on a data set rather than to try and develop parallel capabilities at the site where the data are located. It is also a way of sharing the workload and not over-burdening a few centres.

Once the need for such a centre is identified in developing the end-to-end data management plan, then a centre can volunteer for the task or it may be necessary for the data management planners to solicit one that is known to have the expertise to do the job. Above all, it will be necessary to share the workload and look to centres, which have expertise and experience with the data, in question. It is much easier to deal with an increased volume in data that are already being handled at a centre, than to develop new software systems and handle data that have not been seen before.

Operational or Research Centres will for G3OS purposes be expected to commit to a task for the medium to long term. If a centre is carrying out an important step in the end-to-end management of a data flow, the whole flow will collapse if the centre withdraws its service. Thus there would need to be adequate notice of termination of a service to allow time for an alternate centre to be established.

5. Long Term Archives

The purpose of the long-term archives is to preserve the data for the long term for future research and engineering users. The World Data Centre (WDC) System is an example of such a final archive and much of the G3OS data will be archived in the WDCs. In cases where a WDC does not exist or is not appropriate for a particular data set, then an alternate permanent archive must be found, usually in a government centre that is prepared to accept the responsibilities.

The minimum responsibilities for a permanent archive are generally that the archive agrees to

- accept the data and all available supporting meta data. ,
- to store it either in its original form or in a form from which all the original data and meta data can be recovered
- to refresh or update the medium on which the data and meta data are stored so that both are readable in the future, and
- to provide the data and all supporting meta data to users on request, free of charge or at the cost of reproduction.

In addition, there is an additional minimum requirement for G3OS data sets.

- To store the data so that it can be distinguished and recovered separately from similar data that are not G3OS data (and therefore have not necessarily had the same care taken in collection, quality control, and processing).

Final Archives will be identified for all G3OS data sets in establishing the end-to-end data management systems.

6. Information Centre

The Global Observing Systems Information Center (GOSIC) is located in the College of Marine Studies of the University of Delaware in Lewes, DE. This centre has had previous experience with management of global marine and some atmospheric data sets for the World Ocean Circulation Experiment and the TOGA Coupled Ocean-Atmosphere Response Experiment. This experience is being applied to providing an information centre for the G3OS.

The Information Centre has the responsibility to provide on-line information about the collection, flow, and availability of G3OS data. This information is intended for use by G3OS data collectors and data managers, by those evaluating the effectiveness and performance of the systems, and most importantly, by those who need the data for operational or research and engineering studies.

7. Evaluation and Feedback (Data Coordination)

As with the information centre, this function applies to all aspects of the data collection and data flow right up to final archival. Evaluation and feedback are an essential part of an end-to-end data management system. There are many aspects to evaluation of a system such as the G3OS. Each element of the system is required to perform one or more functions on the data and information and pass one or more deliverables on to the next step according to some schedule. Data and information flow monitoring will endeavour to ensure the tasks are completed and the flows are on schedule. Information about bottlenecks or delays will be fed back to managers and the centres involved for action to correct.

Data will be collected according to a "best practice" and evaluation of the quality of the data will occur at several points in the life cycle of a data set. There will be a requirement to evaluate the performance of the quality control activities to ensure they are achieving the established goals. Information on problems, particularly recurring problems, will be fed back to facilitate resolution of the problems.

Included in the evaluation and feedback function is the monitoring of user satisfaction. In this context "user" includes the science bodies defining the requirements, the managers of the data collection and management systems, and the scientific, technical, and operational users of the data sets and products.

There will be a requirement to ask and answer the question "is the system working"? Evaluation and feedback are the means to answer that question. The scientific and technical advisory panels will define the evaluation and feedback mechanisms necessary to ensure "the system is working". In many cases centres will be asked to assist with some of the evaluation and feedback mechanisms. These centres will act as "Operational or Research Centres" for the evaluation and feedback function.

Annex C

Requirements for Satellite Data for GOOS

The table below shows the space-based data requirements for GOOS. These initial requirements have been compiled with particular reference to the needs of the Global Ocean Data Assimilation Experiment (GODAE). GODAE is a major numerical modelling experiment that will assimilate ocean data from in situ observations and satellites. GODAE is likely, in general, to be more demanding in terms of spatial and temporal resolution, but with decreased emphasis on the deep ocean and perhaps slightly weaker accuracy requirements.

Global Observations of Ocean Circulation - Space-Based Data Requirements

Details			Optimized requirements				Threshold requirements			
Code	Application	Variable	Horizontal Resolution (km)	Cycle	Time	Accuracy	Horizontal Resolution(km)	Cycle	Time	Accuracy
ALTIMETRY										
A	Mesoscale variability	sea surface topography	25	7 d	2 d	2 cm	100	30 d	15 d	10 cm
B	Large scale variability	sea surface topography	100	10 d	2 d	2 cm	300	10 d	10 d	2 cm
C	Mean SL variations	sea surface topography	200	> 10 yr	10 d	1 mm/yr	1000	> 10 yr	10 d	5 mm/yr
D	Circulation, heat transport	sea surface topography	100	NA	NA	1 cm	500	NA	NA	5-10 cm
REMOTE SALINITY										
E	Circulation water transport	surface salinity	200	10 d	10 d	0.1 PSU	500	10 d	10 d	1 PSU
SCATTEROMETRY										
F	Wind-forced Circulation	surface wind field	25	1 d	1 d	1-2 m/s 20	100	7 d	7 d	2 m/s 30
SEA SURFACE TEMPERATURE										
G	NWP; climate, mesoscale models	Sea surface temperatur e	10	6 h	6 h	0.1 K (relative)	300	30 d	30 d	1 K
SEA ICE										
H	Ocean-ice coupling warnings	sea ice extent, concentrat on	10	1 d	3 h	2%	100	1 d	10 d	10%
OCEAN COLOR										
I	Biogeochen istry, transparency	ocean color signal	25	1 d	1 d	2%	100	1 d	1 d	10%

Footnotes:
A requires wave height + wind (EM bias correction) measured from altimeter, water vapor content measured from on board radiometer, and ionospheric content / measured from 2 frequency altimeter.

B requires precise positioning system with an accuracy of 1-2 cm for a spatial resolution of 100 km. C requires precise monitoring of transit time in the radar altimeter. A, B and C require repeat track at ± 1 km to filter out unknowns on geoid. A requires adequate sampling which implies at least 2, and preferably 3, satellites simultaneously. A, B and C require long lifetime, continuity, cross calibration. D requires absolute calibration. F: The requirements on the wind field for sea state determination normally exceed sampling requirements for wind forcing G: High resolution SST from new geostationnary satellite + combination with low satellite

Annex D

Explanation of the Headings in the Categorization of the User Applications

- *Character of the products, product delivery and the product→service chain*. What is the style of information required by the users? Is it a multi-dimensional array or, at the other extreme, indices. What are the user delivery requirements?
- *Character of the inputs.* Are the input requirements complex (e.g., many different data types and data forms) or are they more specific (e.g., a small number of types and dimensions)? Are the data in conventional digital form or in some less conventional form (e.g., photographic images)? Are the metadata requirements complex? Etc.
- *Input delivery requirements.* This includes telemetry from platform to assembly centre, and assembly centre to processing. In some cases the processor (e.g. a model or a coastal manager) may have confined windows for data input. In other cases there may be a priority among possible inputs. There may also be requirements in terms of the format and standards, e.g. it must have some level of QC and be in WMO format.
- *Quality control/assembly requirements.* This is an assessment of the level of QC that is possible given the constraints on input delivery, and the QC needed because of the application of the product (the needs of the user). In some cases the QC may be automated, in others it may require high levels of scientific involvement. Before the data is made available to "processors", be that a model or a human interpretation, it usually has to be assembled in a form that removes the dependencies on the originator (e.g., platform or telemetry specific formats) and enables the "processor" to efficiently ingest the information.
- *Character of the processing* with particular reference to the use of models and data assimilation. "Processors" will vary from application to application. In NWP the processors are complex model assimilation systems that are capable of ingesting raw data and of producing complex arrays for the dependent fields. In other cases, e.g. environmental indices, there may be a more qualitative, subjective component.
- *The level of feedback.* The design of GOOS (e.g., see GOOS 98) emphasizes the need to have information flow from the processors and users back to the data originators and assemblers. The observing system cannot evolve without this feedback. For each form/style/mode of D&IM, we need to characterize the existing levels of feedback and evaluation and the needed levels.
- *State of permanent archives for data and products.* Requirements for improving the state of the permanent archives for the data sets for this application should be identified here.
- *Strengths and building blocks.* This is an extremely critical aspect since it is this information that will dictate how this area is treated in the D&IM Plan. It only makes sense to discuss detail if there is a reasonably robust approach that is ready for sustained application. If none exists then perhaps a reference to research and/or pilot projects that might be delivering capacity is appropriate.
- *Weaknesses and needed capacity*. Even for existing systems there may be aspects that are missing or not working well. These should be identified here.
- *Opportunities?* Are their opportunities that have already been identified that could be used to develop this form of D&IM? In particular, there may be national initiatives that could be adapted for GOOS.

Annex E

The GTSPP Data Management System

1. Status

The GTSPP is a joint IOC/WMO project that knits together both real-time (typically JCOMM) and delayed mode (typically IODE) data collections of global ocean temperature and salinity observations into a single programme. Participants are governmental and scientific organizations in various countries who support their contributions to GTSPP through their own budgets. It was initiated jointly by the IOC Committees for IGOSS (sponsored by IOC and WMO) and IODE in 1989 as a pilot project, and converted to a permanent programme in 1996.

Tasks in the GTSPP are shared amongst the participants. Real-time data processing services are provided by the Marine Environmental Data Service (MEDS) of Canada. The U.S. NODC provides data processing services for delayed mode data and maintenance of the Continuously Managed Database (CMD). The Atlantic Oceanographic and Meteorological Laboratory (AOML), the Commonwealth Scientific and Industrial Research Organization (CSIRO) and the Scripps Institute of Oceanography (SIO) provide scientific advice, quality assessment and flagging of the data, and analyses and presentations of the data handled by the project. Through cooperation with WOCE, the WOCE Subsurface Data Centre in Brest has also contributed data and expertise. Other data centres in IODE provide data to the project as they are processed. Cooperation with the GODAR Project also brings data into the GTSPP.

2. Responsibilities

Observations: The GTSPP concerns itself with temperature and salinity profiles collected from the world's oceans. Other observations made in association with the T and S profiles, such as other profiles or surface marine observations, are also carried with the data. All available data from XBTs, XCTDs, research CTDs, moored and drifting buoys, and from profiling floats are included in the GTSPP data flow.

Services: One of the goals of the GTSPP is to provide data of the highest possible quality as quickly as possible to users. The foundation of this goal is the CMD. This database holds both real-time and delayed mode data. Where both the real-time and delayed mode data exist from a particular location and time, the delayed mode is retained in the CMD because it represents the highest resolution and highest quality data. The contents of the CMD are available upon request from the U.S. NODC.

Products: Products include filtered data sets on request or on a regular basis by subscription, data products including presentations of the distributions of temperature and salinity in space and time for science and fishing operations, and data flow analyses for managing the data flow.

3. Publications

In cooperation with other partners this includes:

- The GTSPP Project Plan
- IOC Manuals and Guides #22
- CSIRO Quality Control 'Cookbook'
- AOML Quality control manual
- All meeting reports
- GTSPP CD ROM

4. Structure

The GTSPP Steering Group has become the Joint IODE-JCOMM Steering Group on End-to-End Data Management Systems. This group will meet as needed to assist in the implementation of new joint systems. GTSPP will be an agenda item at these meetings.

The GTSPP functions by actions undertaken by participants to achieve common goals agreed to at the meetings. Since it is a collection of volunteer organizations, adjustments are always needed to accommodate changes in levels of participation of members. These adjustments are made by current members taking on new roles or by recruiting new members.

5. Observing Network

The GTSPP takes advantage of a number of services and infrastructures available at both the international and national levels. Internationally, the WMO provides the use of the GTS for the transmission of oceanographic messages through the JCOMM programme. GTSPP uses this service to acquire the data exchanged this way.

Some nations have developed an extensive infrastructure to provide and service ships of opportunity in the collection of temperature (and some salinity) profiles around the world. These programmes have become a key component of the SOOP, and GTSPP provides the data management component.

Many nations undertake both monitoring and research data collection programmes at sea. These may be through autonomous instruments, such as floats, or from ships. Data collected are provided to their National Oceanographic Data Centres or to NOCs of the IODE system. From them, T and S data are provided to the GTSPP for inclusion in the databases.

6. Data Exchange and Management

Real-time data are managed by MEDS. The data are received and processed through quality assessment and duplicates resolution software three times each week. At the same schedule, the data are transferred to the CMD held in the U.S. Users who require fast availability to these data can contact MEDS for this service. MEDS provides response to one-time requests or routine downloads of the data received. At present there are both Canadian and international users of the service.

Monitoring of the exchange of real-time data takes place primarily at MEDS. Each month MEDS reviews data from ships that show a more than 10% failure rate on profile data. Systematic problems are noted and ships operators are notified by email. Those ships that have had problems consistently over time are specially noted in the report.

GTSPP also monitors the data received by 4 different centres acquiring GTS data around the world. These include a German, a Japanese and a U.S. site. Each month a report is prepared comparing the data received from all of the sites. Discrepancies noted by these reports are used to track down problems with data routing on the GTS.

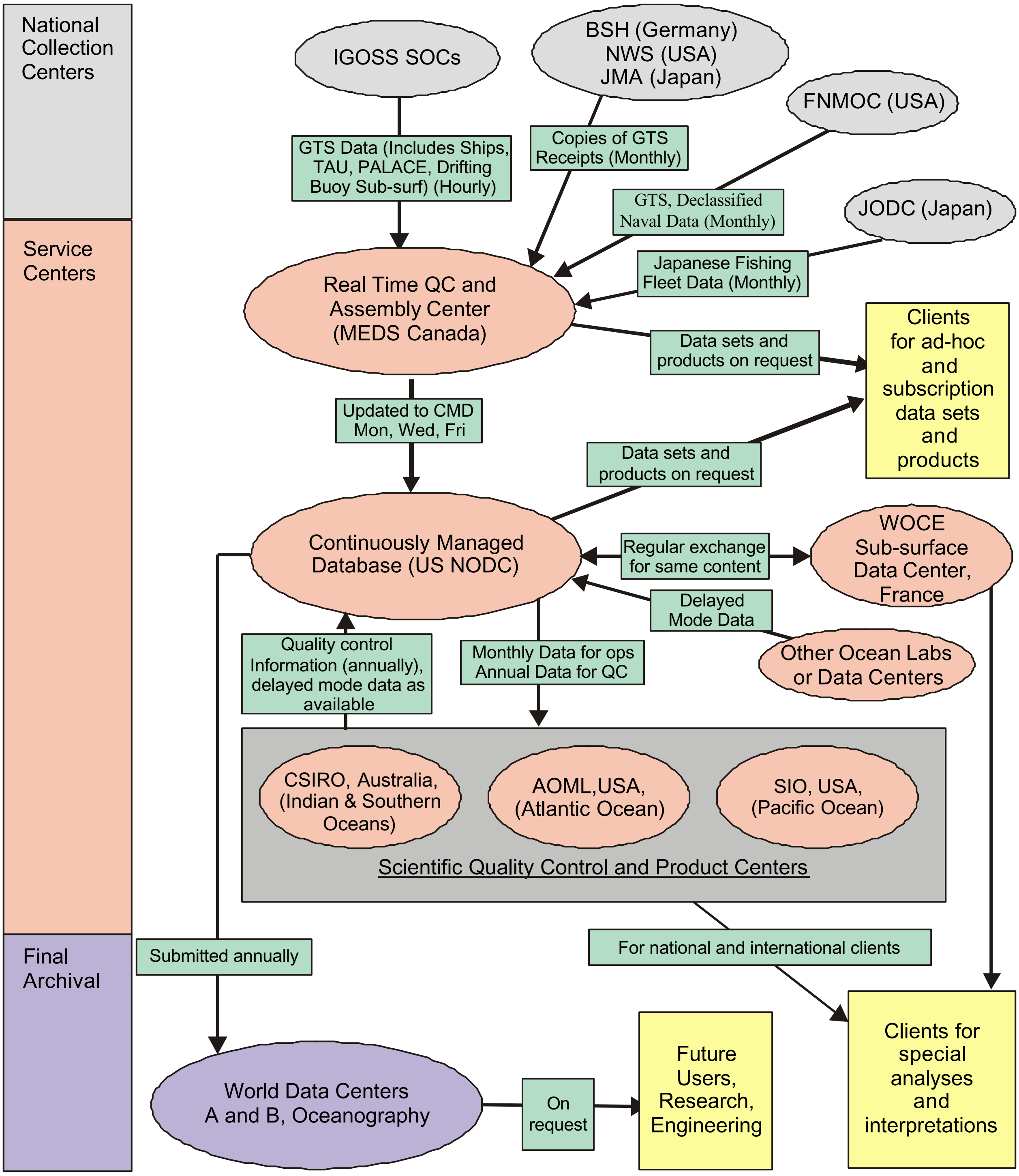
GTSPP has also participated in special and routine monitoring projects of the GTS run by WMO.

Delayed mode data are managed by the U.S. NODC. They accept the real-time data from MEDS and update the CMD when data are received 3 times weekly. Delayed mode data are acquired either from other NODCs, or, by arrangement, from projects such as WOCE, GODAR and SOOP. Users are supported in a similar manner as at MEDS. The data are also available on the Internet.

Through comparisons with the holdings of the WOCE Subsurface Data Centre in Brest and the U.S. NODC, discrepancies in content have been corrected. Scientific data quality assessment has been provided on yearly files by the scientific institutions noted above. Not only does this provide another level of assessment, but also it promotes the collaboration and exchange of expertise between scientific and data management organizations.

7. Detailed Data Flow

The following dataflow diagram describes the GTSPP flow in detail. See the Global Observing Systems Information Centre (GOSIC) web site for the current version of this diagram. Access to specific GTSPP data and data products is available through GOSIC. The on-line diagram in GOSIC contains links to five Operational or Research Centres. These links will connect to web pages that list the data and data products prepared by GTSPP in support of GOOS. Provisions will be made in the service centre web page for obtaining the data and data products on-line. For the rest of the boxes there are links to text that provides information on the centre, data flow, or user.


National Collection Centers
Service Centers
Final Archival
IGOSS SOCs
BSH (Germany) NWS (USA) JMA (Japan)
FNMOC (USA)
JODC (Japan)
GTS Data (Includes Ships, TAU, PALACE, Drifting Buoy Sub-surf) (Hourly)
Copies of GTS Receipts (Monthly)
GTS, Declassified Naval Data (Monthly)
Japanese Fishing Fleet Data (Monthly)
Real Time QC and Assembly Center (MEDS Canada)
Data sets and products on request
Clients for ad-hoc and subscription data sets and products
Updated to CMD Mon, Wed, Fri
Data sets and products on request
Continuously Managed Database (US NODC)
Regular exchange for same content
WOCE Sub-surface Data Center, France
Delayed Mode Data
Other Ocean Labs or Data Centers
Quality control Information (annually), delayed mode data as available
Monthly Data for ops Annual Data for QC
CSIRO, Australia, (Indian & Southern Oceans)
AOML,USA, (Atlantic Ocean)
SIO, USA, (Pacific Ocean)
Scientific Quality Control and Product Centers
For national and international clients
Submitted annually
World Data Centers A and B, Oceanography
On request
Future Users, Research, Engineering
Clients for special analyses and interpretations