GCOS/GOOS/GTOS JOINT DATA AND INFORMATION

MANAGEMENT PLAN

Prepared by the Joint Data and Information Management Panel (JDIMP)

VERSION 1.0

May 2000

GCOS - 60
GOOS - 70
(WMO/TD No. 1004)

TABLE OF CONTENTS

ANNEXES

GCOS/GOOS/GTOS JOINT DATA AND INFORMATION MANAGEMENT PLAN

1. INTRODUCTION

1.1. Background

Three global observing systems are currently in the planning and development stage. These are the Global Climate Observing System (GCOS), the Global Ocean Observing System (GOOS), and the Global Terrestrial Observing System (GTOS). Collectively, these systems are called the G3OS.

Because of commonalities in the basic needs for collecting and managing both the observations (the data collected by the systems), and information about these observations (such as what instruments were used, observing practices and the quality of the data), the G3OS have agreed to cooperate with a single committee concerned with data and information management. The committee is known as the Joint Data and Information Management Panel (JDIMP).

This collaborative approach under the auspices of JDIMP will serve the needs of a wide variety of users involved with questions regarding quantitative representation of the current state of the environment, detection of changes in this land/ocean/atmosphere system, and attribution of identified changes to specific causes.

The expectation is that JDIMP will provide high-level guidance to the scientists engaged in the collection of scientific data and will serve as an agent of the broad, multi-disciplinary environmental information user community. It will accomplish this by encouraging end-to-end procedures that foster acceptable levels of data and information integrity, standardization, continuity, and accessibility.

The following plan presents the JDIMP basis for the G3OS Data Management Activities, and complements the program plans for each of the three observing systems. The respective program plans present detailed information on the system-specific requirements and procedures for each of the three observing systems.

1.2. JDIMP Terms of Reference

Recognizing the need for a comprehensive approach to formulate, implement, and oversee data and information management of the global observing systems, the Steering Committees of GCOS, GOOS, and GTOS have established a Joint Data and Information Management Panel (JDIMP).

The data and information management system for the three global observing systems, G3OS, should be developed, to the degree possible, to accommodate data and products from the various components of the global observing systems. To guide this development, the JDIMP should consist of a core group of members representing the various global observing communities, as well as representatives from contributing disciplines, programs, and agencies. JDIMP membership should encompass a broad range of expertise and include research scientists, who use and understand global data sets, and data and information management experts responsible for significant components of existing operational and research oriented global information management systems.

The JDIMP should be a highly focused "problem solving" group, concentrating on resolving crucial issues affecting the quality and maintenance of global observing system data sets, and

routes of access to them. Particular agenda items may require additional experts be invited to participate on an *ad hoc* basis. In discharging its responsibilities, the JDIMP will:

- Periodically review the G3OS data and information management plan(s) and principles of monitoring;
- As required, commission studies needed by specific observing system components;
- Review adherence of G3OS to cross-cutting principles of data and information management;
- Advise on implementation of data and information management as requested by the senior science committees and sub-panels of the G3OS, to ensure that, for example, data and products are provided as required and archiving activities are adequate;
- Act as a G3OS focus in relation to policy issues, e.g., proposals and actions threatening the availability of environmental data;
- Consider studies commissioned by specific observing system components and the implications for G3OS data and information management.

1.3. System Functional Diagram

The figure below is a model for G3OS Data and Information Management. Each box represents a function. The functions in red are outside the responsibilities of the JDIMP. The functions in blue are within the oversight of the JDIMP.

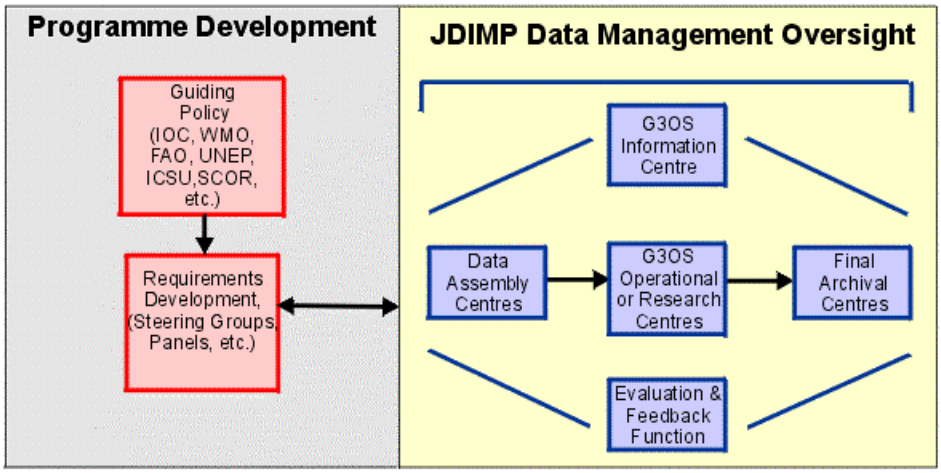

Figure 1. Data and Information Management Model for the G3OS.

1.4. Overview

JDIMP will have an end-to-end monitoring function for data and information management activities within GCOS, GOOS, and GTOS. Primary JDIMP oversight responsibilities will be concentrated in the areas of data archiving, general metadata requirements and products, and the access to the data and products by the users.

To the fullest extent possible, the JDIMP will rely upon existing national and international programs. Systems operated by these programs, such as the International Council for Science (ICSU) World Data Centers, World Weather Watch (WWW), Intergovernmental Oceanographic

Commission (IOC), World Hydrological Cycle Observing System (WHYCOS), and many others, will be used as appropriate to meet G3OS requirements. JDIMP will act as an advocate to ensure data and information are collected, validated, processed, archived, and made accessible with common standards.

JDIMP will also strive to promote the conformance of G3OS observational activities to the principles of environmental monitoring, developed and accepted by the JDIMP. These principles are:

- Assess the impact of new systems or changes to existing systems prior to implementation;
- Require a suitable period of overlap for new and old observing systems;
- Results of calibration, validation, algorithm changes, and data homogeneity assessments should be treated with the same care as the data;
- Ensure a capability to routinely assess quality and homogeneity, including high resolution data and related descriptive information for extreme events;
- Environmental monitoring products and assessments, like the Intergovernmental Panel on Climate Change (IPCC), should be well-integrated into global observing priorities;
- Maintain long uninterrupted stations and observing systems;
- Data-poor regions and regions sensitive to change should be given high priority for additional observations;
- Network operators, designers, and instrument engineers must be provided long-term requirements at the outset of new system design and implementation;
- Promote the conversion of research observing systems to long-term operations in a carefully planned manner;
- Data management systems that facilitate access, use, and interpretation are essential.

JDIMP will support capacity building projects and participation in JDIMP activities by the developing nations. However, this will not include the development of projects on data management that are not firmly rooted in the observational requirements of one of the G3OS component systems.

2. STRATEGIC GUIDANCE

2.1. Data Policy

The G3OS Data and Information Management Plan will be guided by the following principles:

- Full and open sharing and exchange of G3OS-relevant data and products for all G3OS users is a fundamental objective. They should be provided at the lowest possible cost to G3OS users;
- Preservation of all data needed by G3OS is required. Suitable archive facilities should be ensured for all G3OS data. As part of the end-to-end information management framework, all G3OS data sets should have a designated data custodian;

- Procedures and criteria for setting priorities for data acquisition, retention, and purging should be developed and implemented by participating nations and international bodies;

- An international clearinghouse process should be established to prevent the purging and loss of important data;

- To the maximum extent possible, data archives must include easily accessible information about the data holdings, including long-term quality assessments, supporting ancillary information, and guidance and aids for locating and obtaining the data;

- Internationally-agreed standards should be used to the greatest extent possible for the acquisition, processing, archiving, and distribution of global data;

- As appropriate, all G3OS data should be processed to a level which is generally useful to researchers without a detailed knowledge of the observing instrument (i.e. physical rather than engineering measurements). However, for data from those instruments that utilize sophisticated algorithms for this processing, archives must be adequately maintained to permit recalculation of the geophysical data as improved information becomes available;

- All data and products pertinent to G3OS should be described in internationally accessible, on-line computerized directories which conform to agreed-upon standards. Data centers or institutions unable to provide this service should provide descriptions of their G3OS data to centers that can;

- G3OS data sets should be harmonized to the extent possible to allow integration of national and regional data sets into a usable global information resource.

2.2. Guiding Policy

The policy functions for purposes of the G3OS are at the international level and are the responsibilities of bodies such as IOC, World Meteorological Organization (WMO), ICSU, and the Food and Agriculture Organization of the United Nations (FAO). These bodies identify the need for and sponsor cooperative international scientific and operational programs such as WWW, World Climate Research Programme (WCRP) and International Geosphere-Biosphere Program (IGBP). Bodies such as these have also specified the need for and have sponsored data collection programs to meet the needs of present and future international scientific and operational programs.

2.3. Requirements Development

The development of the requirements for the international research and operational programs and supporting global observing systems are accomplished through the vehicle of scientific steering groups (SSGs), steering committees (SCs), and scientific and technical panels. These various groups are composed of nominated experts in the fields relevant to the program being developed or managed.

The scientific and technical advisory panels of the G3OS systems 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 metadata to be collected and maintained with the data. In some instances, special requirements may apply to the handling of collected data. For example, for certain monitoring activities of the G3OS, such as short-term climate monitoring, the timely (real-time/quasi real-time) provision of data is essential; the science panels should identify such special requirements for data exchange.

The scientific and technical panels are also responsible for specifying general requirements 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 metadata. G3OS arrangements, including data assembly, processing, quality control, and distribution should promote the accommodation of both general and special requirements.

There are several such panels in existence at the moment including:

- Atmospheric Observation Panel for Climate (AOPC)
- Ocean Observations Panel for Climate (OOPC)
- Terrestrial Observation Panel for Climate (TOPC)
- GOOS Living Marine Resources Panel
- GOOS Coastal Module Panel
- GOOS Health of the Ocean Module Panel.

2.4. International Agreements

International conventions and other political drivers (e.g. Convention on Biodiversity, United Nations Framework Convention on Climate Change, Convention on Desertification) raise issues and lay the framework for actions to resolve these issues. The political drivers are the responsibility of the Sponsoring Organizations in cooperation with the Steering Committees, and the responsibility for responding to these issues lies with the respective G3OS scientific and technical panels and steering committees.

3. DATA PROCESSING, MANAGEMENT AND ARCHIVING

3.1. Data Assembly Centers

Data for international programs are generally collected by national activities. As the initial step in the JDIMP process, there is a need for the accomplishment of a data assembly function, also generally as a national activity. The data assembly center (or function) collects together specific data and metadata for one or more data sets identified by the panels as required for G3OS purposes. This is the initial area of JDIMP responsibility, to insure that data from all G3OS observing systems are being adequately assembled.

National standards and observing practices will not necessarily conform to the principles of environmental monitoring or meet the specific needs of GOOS, GCOS, or GTOS. For example, for detection of climate change it is necessary to have a period of simultaneous operation for calibration purposes between old and new sensors when a new version of a sensor is being installed. Additional metadata may also be required. Thus for the G3OS, the data assembly center and the national data collection activity may be asked to undertake special processing or handling for G3OS data. In any case, the data assembly center will have certain responsibilities for assembling national data and metadata, for processing the data in a certain way, and for forwarding the data to one or more international centers for further assembly, processing, distribution, and archiving according to the details of the end-to-end data management plan for the variables in question.

Some countries will probably have several data assembly centers 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 data assembly centers. It would also make sense for groups of developing or developed countries with common interests to share data assembly centers according to their capabilities and expertise.

3.2. G3OS Operational or Research Centers

The next required function in G3OS data management is provided by the G3OS Operational or Research Centers (or similar functional activity), which take 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 centers will be identified by the Scientific and Technical Panels and/or JDIMP 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 centers carrying out different steps in the processing. For example, it is usually more effective to move the data to the place where the necessary expertise and capacity exist 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 one center.

Once the need for a center is identified, then a national center with appropriate expertise and commitment can volunteer for the task; alternatively, it may be necessary for G3OS to solicit a center that is known to have the expertise to do the job. Above all, it will be necessary to look to centers that 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 center, than to develop new software systems and handle data that have not been seen before. For fledgling centers, joint work with established centers would be recommended initially.

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

3.3. Metadata Guidelines

Metadata, the information required to understand data set content and to optimize the usefulness of the data contained in a data set, are of the highest importance, and it is essential that G3OS data collection activities meet the JDIMP requirements discussed in this plan.

In discussions of metadata, JDIMP distinguishes between “directory level” metadata and “archive level” metadata. The directory level metadata, which is a subset of the archive level metadata, provides the general descriptive information needed by a prospective user to identify the data set in a high-level catalogue or listing of G3OS data sets. This would include information such as the specific variables and geographic area observed, the time and frequency of observations, and the duration of the observation periods. It should also provide data set locations, contacts, and access procedures/constraints.

Data collection programs are responsible for specification of archive-level metadata, with assistance from JDIMP if requested. The archive level metadata provides the detailed, technical material needed to understand the precise characteristics of the data and to assess their adequacy for use in a particular scientific research project. This level of metadata answers questions such as: What are the observation station names and types? What is the current station location and has there been a relocation during the period of observations? What is the site description? What are/were the periods of observations? What instrument is/was in use? What are/were the observing practices? What conversion algorithms are/were used? Is it an original (raw) data set? Is it quality controlled? Is it homogenized? What calibration practices and standards were employed? What calibration records are included with the data? What code/format was used? What code/format is being used now? What is availability/access of data?

Although the requirement for adequate metadata cannot be overstated, there will be situations in which metadata do not exist for potentially useful historical information. In these cases, the data should be flagged but not necessarily rejected.

3.4. Final Archives

Final archives will be identified for all G3OS data sets in establishing the supporting end-to-end data management systems. The purpose of the final archives is to preserve the data for the long term for future users, and in this role to guarantee indefinitely the data integrity and long-term survival. Each G3OS data set should be regularly forwarded to and maintained by at least one designated archive facility.

The World Data Center (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 a government center that is prepared to accept the responsibilities and commit to long-term stewardship.

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

- To accept the data and all available supporting metadata;
- To ensure that back-up copies of critical data are made and that metadata are preserved with the data;
- To store data either in original form or in a form from which all the original data and metadata can be recovered;
- To refresh or update the medium on which the data and metadata are stored so that both are readable in the future;
- To provide the data and all supporting metadata to users on request, free of charge, or at a cost no more than the cost of reproduction or transmission.

There are two additional minimum requirements for serving as an archive of G3OS data sets:

- To store the data so that they 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;
- To be able to retrieve data sets and their associated metadata as received at the archive.

4. USER INFORMATION AND FEEDBACK

4.1. G3OS Information Center

The Information Center under development to support the G3OS is highly distributed and deals with a large number of specific data that have been collected and processed according to agreed standards. This Center is known as the Global Observing Systems Information Center (GOSIC) (web address: http://oceanic.cms.udel.edu/gos/). The GOSIC has the responsibility to provide on-line information about the collection, flow, and availability of G3OS data and single-point access to the data sets, products, and services. This service 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. GOSIC systems use directory-level metadata to aid users in locating and obtaining data sets.

In addition to data being collected at present and in the future, the G3OS Science Panels seek out and include information on historical data sets that are judged to be valuable for the establishment of baselines and long-term studies of change. These databases are widely distributed over a large number of existing centers with various data management systems and access procedures. Users will be faced with a broad range of technologies and access procedures. To make it easier for users to cope with the wide range of technologies and procedures involved in accessing the data, the locations, contacts, and procedures will be documented in the GOSIC Data Set Information Base.

4.2. Standards

Where possible, accepted international standards will be used for the acquisition, processing and distribution of G3OS information. Where additional standards are required, they will be developed in cooperation with appropriate national and international programs. In the short term, emphasis will be on ensuring that data and products are available to users in standard formats, on low-cost media and through standard communications protocols. In the longer term, standards for automatic transmission of queries to and replies from databases are desirable.

4.3. Evaluation and Feedback

As a program committed to consistently providing high-quality data, information and services, G3OS must ensure that the entire information system is performing to specifications. This will necessitate the identification of individual(s) who will accept the responsibility for continuous monitoring of the systems that move data from collectors to processing centers, and from processing centers to users. Errors in data collection or processing procedures can result in erroneous scientific conclusions and incorrect advice to policy makers. In many cases, these errors are not detected for many months or even years.

To ensure that G3OS data are of suitable accuracy, it is essential that procedures to monitor the quality of the data be built into all aspects of operations. Additionally, the utility and efficiency of the information access and retrieval system must be monitored regularly.

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. There are many aspects to the evaluation of a data and information system for G3OS. As with the information center, the evaluation and feedback requirement applies to all aspects of the data collection and data flow right up to the final archive. Evaluation and feedback are an essential part of an end-to-end data management system. Data and information flow monitoring will endeavour to ensure that tasks are completed and the delivery of data accomplished on schedule. Information about bottlenecks or delays will be fed back to managers and the centers involved for corrective action.

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. JDIMP design panels will organize and implement the evaluation and feedback mechanisms necessary to ensure "the system is

working". In many cases, research and operational centers will be asked to assist with evaluation and feedback functions.

4.4. Services to Clients

Although not shown in Figure 1, services to clients will be provided from all the boxes on the diagram, both red and blue. The policy function will provide guidance and coordination to national and international activities seeking to address the problems created by global and regional change. The panels define and coordinate the data collection activities to satisfy the requirements for data and derived information and identify the relevant "best" historical data sets for users.

All the blue boxes will provide data and information as appropriate at various time scales to a variety of clients. The clients will include national and international researchers, decision-makers, politicians, operational agencies, engineers, economists, those studying social impacts, etc. The blue boxes will not provide interpretations, assessments, or predictions but rather the data and information from which others can do so.

ANNEX I

G3OS COMPONENT OBSERVING SYSTEMS

GTOS

Mission statement:

GTOS mission is to provide policy makers, resource managers and researchers with the data they need to detect, quantify, locate and understand changes (especially reductions) in the capacity of terrestrial ecosystems to support sustainable development.

GTOS focuses on five issues of global concern:

1. Changes in land quality;
2. Freshwater resources availability;
3. Pollution and toxicity;
4. Loss of biodiversity;
5. Climate change.

GTOS promotes:

- Integrated analysis of bio-physical & socio-economic data sets;
- Interaction between monitoring networks, research programs and policy makers;
- Data exchange and application;
- Quality assurance and protocols to harmonize measurements methods;
- Provides guidance in data analysis.

GTOS is NOT:

- A source of funding (but does assist others in seeking funds for scientifically sound work that is consistent with its objectives);
- A source of primary data;
- A research program.

GTOS Web Address: (http://www.fao.org/GTOS/Home.htm)

GOOS

The Vision: A unified global network to systematically acquire, integrate and distribute oceanic observations, and to generate analyses, forecasts and other useful products is needed to supply the information needed by governments, industry, science and the general public to deal with marine related issues, including the effects of the ocean upon climate.

The Mission: To use long-term, multi-disciplinary, operational oceanographic monitoring of seas and oceans as the basis for (i) enabling the use of ocean data in creating and disseminating reliable assessments and predictions of the present and future state of these environments in support of their health and sustainable use and (ii) contributing to prediction of climate change and variability, for the benefit of a wide range of users. This will (iii) guide the directions of scientific and technical research, development and training in the various disciplines of oceanography that will in turn facilitate the development and management of the system.

The Goals: (i) To serve the marine data and information needs of humanity for the efficient, safe, rational and responsible use and protection of the marine environment, and for climate prediction and coastal management, especially in matters requiring information beyond that which individual national observation systems can efficiently provide, and which enable smaller and less-developed nations to participate and gain benefit; (ii) To establish an international system to provide the required coordination and sharing of data and products that otherwise would not be possible.

The Sponsors: The Global Ocean Observing System (GOOS) is an ambitious undertaking by the Intergovernmental Oceanographic Commission (IOC) of the United Nations Educational, Scientific and Cultural Organization (UNESCO), the World Meteorological Organization (WMO), and the United Nations Environment Programme (UNEP), together with scientific guidance from the International Council for Science (ICSU).

The International Context: The United Nations Conference on Environment and Development (UNCED) called in 1992 for the creation of a global system of ocean observations to enable effective and sustainable management and development of seas and oceans, and prediction of future change. The establishment of such a system was also urged by the Second World Climate Conference in 1990 to provide the oceanographic data needed by the Global Climate Observing System (GCOS).

The Mode of Operation: GOOS is envisioned to resemble the global meteorological observation and prediction network, supported by national governments and implemented through the contributions of national agencies, organizations and industries, with the assistance of national and international data management and distribution bodies.

GOOS Web Address: (http://www.unesco.org/goos/

GCOS

OBJECTIVES

The GCOS is intended to meet the needs for:

- Climate system monitoring, climate change detection and monitoring the impacts of and the response to climate change, especially in terrestrial ecosystems and mean sea-level;
- Data for application to national economic development;
- Research toward improved understanding, modelling and prediction of the climate system.

IMPLEMENTATION STRATEGY

The GCOS Implementation Strategy is to clarify its role vis-a-vis the other observing systems for climate and the areas of action in which it would place its priority. In the future, GCOS will focus its resources in 4 strategic areas and will:

- Build on existing operational and scientific observing, data-management, and distribution systems, but using standards based on established GCOS principles. This involves working with partner observing systems and research programs and activities more closely;
- Build ownership by national governments and international agencies and organizations in implementing a multidisciplinary global observing system for climate. This will require extensive communication and involvement at the intergovernmental and national levels;
- Address deficiencies in the current Initial Operational System (IOS) at the regional level, where the impacts of climate variability and change and many of the known deficiencies in observing systems occur;
- Keep the IOS relevant to the user communities. Climate is a multidisciplinary problem in which user needs are continually changing and new understanding from research and technology are developing new ways to address requirements across domains and with new mixes of sensors.

Implementing GCOS will require a substantially different approach than that followed by its partner, domain-based observing systems. This will require building a broad level of interest and ownership that incorporates the multiplicity of issues, activities, organizations, and systems. There will be a continuing need to explain and rationalize the complexities of the many facets of climate at the scientific, programmatic, managerial, political and public levels within international and national societies.

GCOS Web Address: (http://www.wmo.ch/web/gcos/gcoshome.html)

ANNEX II

LIST OF ACRONYMS

AOPC	Atmospheric Observation Panel for Climate
FAO	Food and Agriculture Organization of the United Nations
GCOS	Global Climate Observing System
GOOS	Global Ocean Observing System
GOSIC	Global Observing Systems Information Center
GTOS	Global Terrestrial Observing System
G3OS	GCOS, GOOS, and GTOS, collectively
JDIMP	Joint Data and Information Management Panel
ICSU	International Council for Science
IGBP	International Geosphere-Biosphere Program
IOC	Intergovernmental Oceanographic Commission
IOS	Initial Operational System
IPCC	Intergovernmental Panel on Climate Change
OOPC	Ocean Observations Panel for Climate
SC	Steering Committee
SSG	Scientific Steering Group
TOPC	Terrestrial Observation Panel for Climate
UNCED	United Nations Conference on Environment and Development
UNEP	United Nations Environment Programme
UNESCO	United Nations Educational, Scientific and Cultural Organization
WCRP	World Climate Research Programme
WDC	World Data Center
WHYCOS	World Hydrological Cycle Observing System
WMO	World Meteorological Organization
WWW	World Weather Watch