GTSPP Project Plan (Sep, 2002)

1. Introduction

1.1 Purpose of Document

The Global Temperature Salinity Profile Programme (GTSPP) is a joint World Meteorological Organization (WMO), and Intergovernmental Oceanographic Commission (IOC) programme. Functionally, GTSPP reports to the Joint Commission on Oceanography and Marine Meteorology (JCOMM), a body jointly sponsored by WMO and IOC and to the IOC Intergovernmental Oceanographic Data and information Exchange, IODE.

Development of the GTSPP (then called the Global Temperature-Salinity Pilot Project) began in 1989. The short-term goal of the GTSPP was to respond to the needs of the Tropical Ocean and Global Atmosphere (TOGA) Experiment and the World Ocean Circulation Experiment (WOCE) for temperature and salinity data. The longer term goal was to develop and implement an end to end data management system for temperature and salinity data, which could serve as a model for future oceanographic data management systems.

GTSPP went into operation in November 1990. The first version of the GTSPP Project Plan was published the same year. Since that time, there have been many developments and some changes in direction including a decision by IOC/WMO to end the pilot phase and implement GTSPP as a permanent programme. Member states of IOC and WMO decided it was time to publish an updated version of the programme plan. This was done in 1998.

Once more, the evolution of the global observing systems, and needs of participants necessitate a redrafting of the project plan. GTSPP is now a part of the Global Ocean Observing System (GOOS) and needs to reorient its operations to fit. Likewise, participants have expressed the desire of having a project plan that was better suited to garnering support from their national funding agencies.

1.2 Rationale for the Programme

GTSPP was developed as a response to changing needs in the ocean science and marine operational communities. Scientists conducting research into oceanic and atmospheric processes are addressing global issues and need multi-disciplinary data sets for their experiments. The operational community needs more data in operational time frames to conduct safe operations at sea and to support modeling. Both the scientific and operational communities are applying the knowledge of ocean processes gained in many decades of research. Various National Weather Prediction (NWP) models and the Global Ocean Data Assimilation Experiment (GODAE) are the latest thrusts to assimilate atmospheric and oceanic data into coupled ocean-atmosphere models to address weather prediction, sustainable development, climate change, and human and environmental safety.

As a result of these needs, data collection programmes are more complex and produce significantly larger volumes of data. The quality control problem is more difficult because of the increased volume and complexity of the data. Improved means of using high-speed data communications and the World Wide Web in international data management are being developed. Data and improved products need to be available more quickly than in the past. In short a new paradigm for international oceanographic data management is required. GTSPP has and will continue to address the development of that paradigm.

Part of the development of this new paradigm is the adoption of standards in all aspects of the end to end data system. GTSPP has been in the forefront of developing standards. However, it must also look wider a field to standards developed by other organizations but which can be adopted by the marine community so that a wider range of data may be easily used,

2. Objectives

The original objectives formulated for the GTSPP appear in Annex A. These were reformulated in the previous project plan and this reformulation remains appropriate. The objectives are as follows.

(i) To continue to provide a timely and complete data and information base of ocean temperature and salinity profile data of known and documented quality in support of global and local research programmes, national and international operational oceanography, and of other national requirements.

(ii) To continue to improve data capture, data analysis, and exchange systems for temperature and salinity profile data by encouraging more participation by member states in GTSPP, by locating new sources of data from existing and new instruments and implementing the systems to capture and deliver the data, by taking full advantage of new computer and communications technologies, and by developing new services and products to enhance the usefulness of GTSPP to clients and member states.

(iii) To continue to develop and implement data flow monitoring systems to improve the capture and timeliness of GTSPP real time and delayed mode data, and to distribute information on the timeliness and completeness of GTSPP data bases so that bottlenecks in the data flow can be identified and addressed.

(iv) To improve the state of databases of oceanographic temperature and salinity profile data by developing and applying improved quality control systems, by implementing new data centre tests for QC as appropriate for new instrumentation; by working with the scientific partners of GTSPP to train data

centre staff and transfer scientific QC methods to the centre, and by feeding information on recurring errors to data collectors and submitters so that problems can be corrected at the source

(v) To facilitate the development and provision of a wide variety of useful data analyses, data and information products, and data sets to the GTSPP community of research, engineering, and operational clients.

The pilot project had another objective to improve the completeness of the historical databases by digitizing historical data existing only in analogue or manuscript form and by including digital data not presently at a World Data Centre (WDC). This responsibility was passed to the IODE Global Data Archaeology and Rescue Project (GODAR).

3. Benefits to Member States

GTSPP has an obvious international objective, but it is also intended that member states benefit at the national and regional level. In particular, member states benefit from GTSPP because:

- Member states can receive the most complete and timely data set available to produce the most reliable and accurate operational products for human safety and environmental protection for national, regional or international use. GTSPP is always seeking to add data flows from new instruments (e.g. P-ALACE floats) and new data sources to the GTSPP data flows.
- Member states benefit from global standardization of quality control procedures and the results of the various tests are carried out on the data. There is also standardization of data formats and processing methodologies to simplify data exchange nationally and internationally. The GTSPP will continue to develop improved QC and standardization practices in cooperation with its supporting science centres.
- Member states are able to receive regional and global data flow information and inventories enabling them to obtain the data of interest to them from the international data exchange organizations. GTSPP will continue to improve and expand these sources of information on the availability of data.
- New state of the art methods and technologies in communications, software applications, quality control, and data management are under constant development and are available for transfer to member states.

The following are some examples of benefits of participation in the GTSPP by some IOC and WMO member states.

- GTSPP data management and quality control systems for temperature and salinity data have been written in the national data centre and have been distributed and implemented in oceanographic laboratories in the member states. This has saved development costs by sharing the workload of the software development and results in fully processed data of a much higher quality arriving at the national data centre much sooner than in the past.
- Data flow monitoring reports that are prepared by comparing the GTS data flows in Europe, Asia, and North America have enabled many member states to improve their understanding of what temperature and salinity data are available from the GTS and to take the necessary steps to ensure they are acquiring the most complete data set available.
- Because of the improved availability of ocean temperature and salinity data in operational time frames oceanographic labs in some member states have decided to begin regular preparation and distribution of temperature and salinity distribution maps for use by fishermen and by aquaculture interests. Because of the interest in the products these labs have in some cases actively sought out other sources of temperature and salinity data. They then convinced these other sources to submit their data to GTSPP in operational time frames so it would be available for their products. In the process they have increased the coverage of the GTSPP data set for all users.
- Feedback of information on problems in the data to the operators of some ships of opportunity have enabled these operators to correct the problems. The number of problems in the monthly reports has noticeably decreased since this reporting was initiated. This results in an improved data set for all users.

- There has been a significant improvement in the quality control carried out on their data by data centres participating in GTSPP. Before GTSPP procedures were implemented, data sets had many errors in position, date-time, and variable fields. The data sets also had a large number of duplicate observations as data sets from various sources were combined. This required users to develop and maintain their own systems for quality control and duplicates removal. Many users now use the data sets directly, saving the workload of maintaining QC and duplicates removal systems and the workload of operating them.

It is expected that as GTSPP moves forward and member state participation increases, more and varied benefits of this nature will be realized.

4. Governance

4.1 Parent Bodies

The parent bodies for GTSPP are the Intergovernmental Oceanographic Commission and the World Meteorological Organization At the Seventh Session of the Joint IOC-WMO Committee on IGOSS (intergovernmental Global Ocean Services System) it was recommended that GTSPP become a permanent programme of IGOSS and the intergovernmental Ocean Data and Information Exchange (IODE, a subprogram of IOC). The Fifteenth Session of the Committee on IODE concurred and GTSPP is now a permanent programme rather than a pilot project.

The permanent programme was been renamed the Global Temperature-Salinity Profile Programme and retained the acronym GTSPP.

JCOMM, the successor to IGOSS, held its first meeting in 2001. At this meeting, JCOMM formally adopted sponsorship of GTSPP. The chair of GTSPP reports to the Expert Team on Data Management Practices, a subgroup of the Data Management Programme Area. GTSPP also reports directly to the IODE

4.2 Steering Group

The terms of reference and the composition of the Steering Group on the GTSPP were accepted the Thirteenth Session of the Committee on International Oceanographic Data and Information Exchange (New York USA, 17-24 January 1990).

The Steering Group has been responsible for the continuing development and implementation of the Programme. The Group oversees and directs the development of the Programme and promotes and encourages participation in the GTSPP. The Group includes members from the scientific community and provides scientific and data management guidance to participants. The Group specifies, maintains, and distributes relevant documentation to JCOMM and IODE contacts and representatives of scientific programmes, and reports to the IODE and JCOMM Committees. It also invites scientific experts at the discretion of the Chair.

The complete terms of reference and composition for the existing Steering Group are contained in Annex B. These terms of reference will be reviewed and updated as appropriate at meetings of parent bodies of GTSPP and at which it reports its activities.

5. Participation

There are 14 organizations in 8 member states of IOC and WMO participating actively in the GTSPP. Active participation means that the member state or organization is doing something for GTSPP on a regular basis.

It is a goal of GTSPP is to increase this participation. The types of contributions that are being made now are described here.

- data centres provide data assembly, data management, data system products

- science centres provide scientific QC, science oversight, analysis products
- data contributors encourage data submission, assist data assembly
- meeting national objectives furthers the international data system

Ideas on new contributions are also included. Member states that do not yet do so are encouraged to examine their data related activities and where appropriate to actively support the GTSPP. Expressions of interest should be directed to the chair or members of the GTSPP Steering Group.

The existing contributions are varied. Some organizations supply data. Others, such as data centres provide data assembly, quality control, duplicates management and data system performance reports. Some agencies, such as those running national programmes, or data centers, acquire data from national programmes and contribute these either to the real-time distribution system or to GTSPP centres. Scientific organizations provide a vital role carrying out scientifically based quality control to data sets for their region of expertise and providing advice to how the data system should work. Other organizations are providing computer software for use in GTSPP data centres.

The best progress will be realized in the future if contributions come from more member states so that a heavy workload is not imposed on a small number of active participants. If the workload is spread:

- the work tasks can be completed in a timely manner,
- a large number of agencies gain a sense of involvement by actively contributing to GTSPP, and
- the workload of individual agencies is reduced to a level where it can be undertaken without compromising other activities.

As indicated above there are many areas where member states can make valuable contributions. Within their own marine science communities Member states can:

- encourage an increase in the number of temperature and salinity observations transmitted in near real time,
- undertake quality control checks according to the GTSPP Quality Control Manual on data collected by national programmes and attach GTSPP metadata and flags,
- improve mechanisms allowing the more timely submission of both near real time and delayed mode data
- improve mechanisms allowing the submission of greater vertical resolution in near real time data
- encourage national research agencies to develop data and information products as part of the scientific quality assurance process and as a service to national and international users,
- actively acquire historical temperature and salinity data that has not previously been exchanged,
- provide advice and assistance to the programme in the areas of data management, quality control, communications, and product development, and
- provide software that can be used by centres managing the GTSPP data and can be distributed to centres in member states.

For example one area that needs improvement is the submission of near real time salinity observations, surface and sub-surface. Globally there are very little salinity data transmitted over the GTS. Member states should make every effort to encourage their research vessels to submit real-time data from whatever instruments can provide this type of data.

All member states automatically play a part in GTSPP through their normal JCOMM/IODE activities. To participate more actively in the GTSPP, member states can write to the IOC Secretariat, the WMO Secretariat, or to the Chairman of the GTSPP Steering Group outlining the areas in which they wish to contribute.

6. Functions

This section deals with the functions that are required for the GTSPP. It begins with a brief overview of the operations of the GTSPP. This is followed by a more detailed review of the functions that are required to support these operations. Agencies that support these functions have the responsibility to do at least the tasks noted. Of course, as technology and knowledge evolve, better ways to accomplish the goals will

also change. Each agency is expected to look for ways to change their operations to improve the performance of the GTSPP, to raise this with the Steering Group and to implement the agreed actions.

6.1 GTSPP Operations

Figure 1 presents the data flows of national and international programmes within which GTSPP is placed. The boxes in the Figure represent generic centres. A given international JCOMM or IODE centre may fit within several boxes in carrying out its national and international responsibilities. The following sections discuss this Figure in terms of essential elements of the GTSPP.

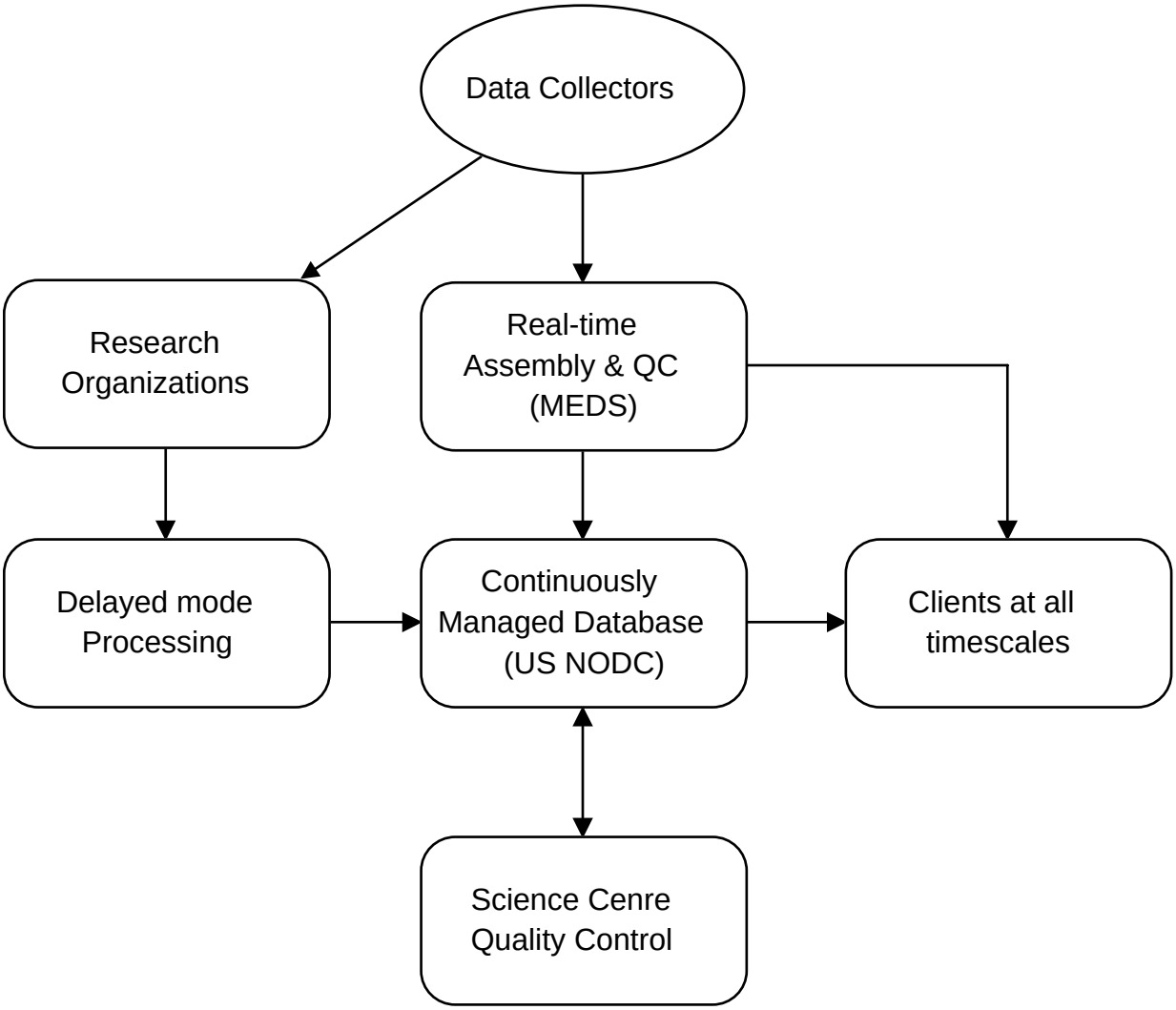

Figure 1: GTSPP data flow

6.1.1 Near Real Time and Operational Time Frame Data Acquisition

Near real time data acquisition within GTSPP depends on the GTS of the World Weather Watch of WMO and the telecommunications arrangements for BATHY and TESAC data established by JCOMM. Copies of other real time or operational time frame data sets are acquired from any other available sources via the Internet of other high-speed networks. The goal is to ensure that the most complete operational time frame data set is captured.

Figure 3 is a graphic representation of the GTSPP operational time frame data flow. The "data collectors" in the top boxes follow one of two procedures. In the first case the data are provided to GTS centres that place them on the GTS within minutes to days of its collection. In the second case the data are supplied to a national organization that forwards it to the real time centre in MEDS within a few days to a month of its collection.

The real time data that are circulated on the GTS are copied by both MEDS and the Specialized Oceanographic Centres (SOCs) of JCOMM and by users of real time data who have access to the GTS. These users include meteorological and oceanographic centres that issue forecasts and warnings, centres that provide ship routing services, and centres that prepare real time products for the fishing industry. Some of the JCOMM SOCs provide copies of the data they receive to MEDS for a monthly analysis of GTS data flows.

MEDS compiles the global data set from the various sources, applies the documented GTSPP QC and duplicates removal procedures, and forwards the data to the US NODC three times per week. At NODC the data are added to the continuously managed database, on the same schedule. There are also several clients that receive copies of the data sent from MEDS three times per week. These are clients who do not need the data within hours but rather within a few days. By getting the data from the GTSPP Centre in MEDS they save having to operate computer systems to do quality control and duplicate removal.

Note that the regular route for real-time data to the box marked "Operational Clients" in Figures 1 and 2 is not affected by GTSPP. This route provides for uninterrupted flow of data for weather and operational forecasting through the national weather services of member states. These centres need the data in hours rather than days.

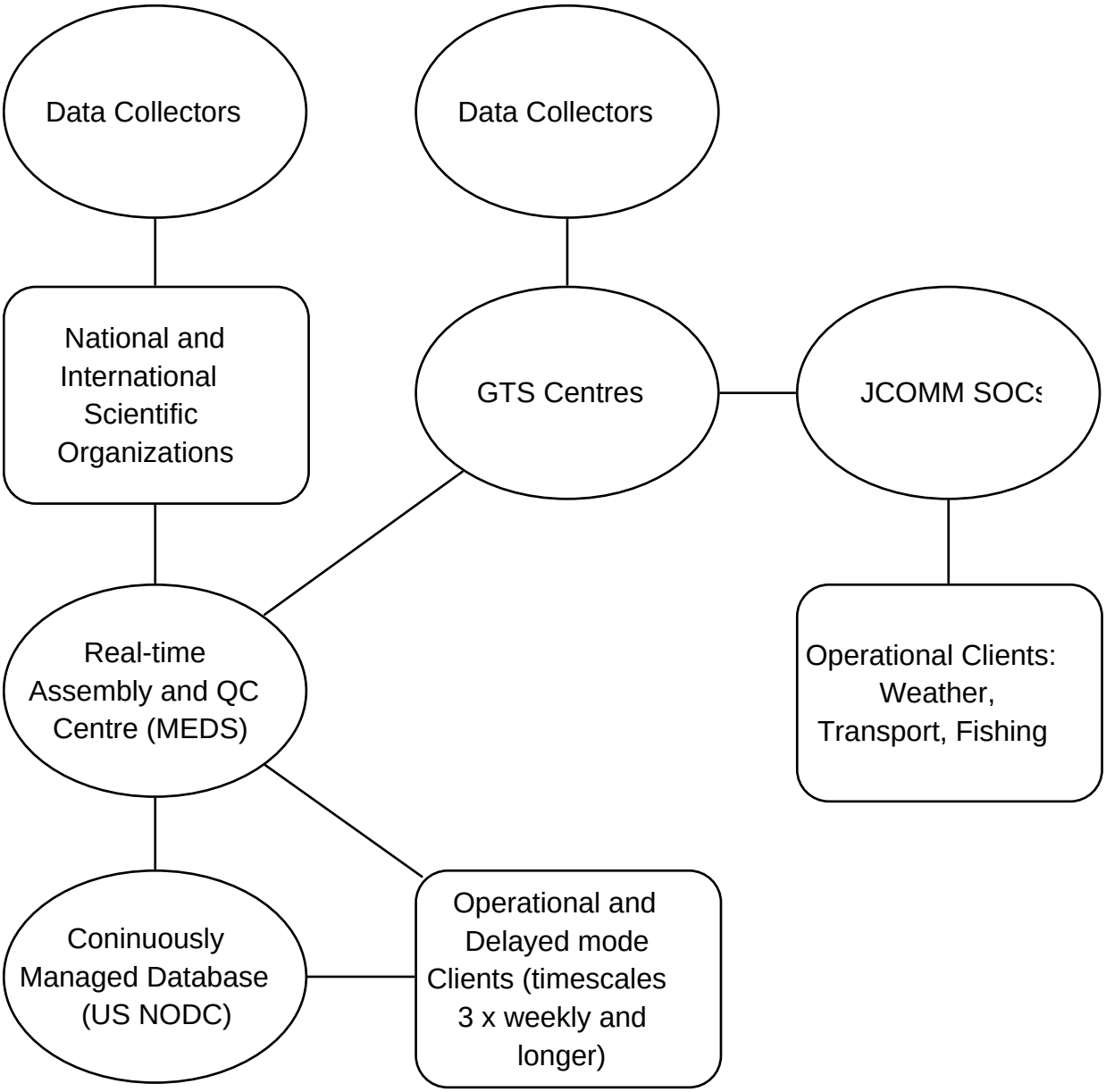
Data Collectors
Data Collectors
National and International Scientific Organizations
GTS Centres
JCOMM SOCs
Real-time Assembly and QC Centre (MEDS)
Operational Clients: Weather, Transport, Fishing
Coninuously Managed Database (US NODC)
Operational and Delayed mode Clients (timescales 3 x weekly and longer)

Figure 2: Real-time data flow

6.1.2 Delayed Mode Data Acquisition

GTSPP utilizes, to the extent possible, the existing IODE data network and processing system to acquire and process delayed mode data. The box entitled "Delayed Mode & Historical Data" in Figure 3 shows the delayed mode data flow in graphic form. The data flow into the continuously managed database in the US NODC is through a "Delayed Mode QC" process. This process is analogous to the QC carried out on the real-time data and conforms to the specifications of the GTSPP QC Manual. In some cases, where appropriate arrangements can be made this QC process could exist and be performed in another national oceanographic data centre on behalf of NODC.

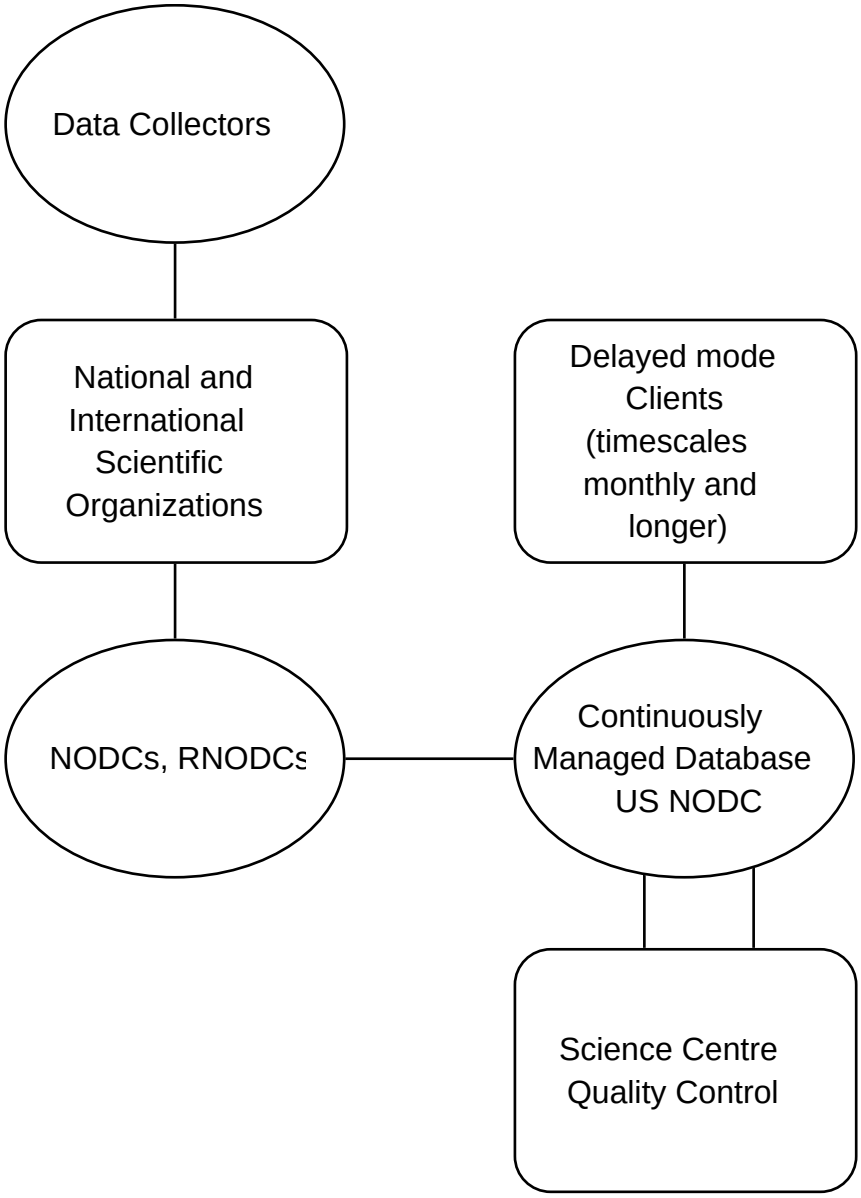

Figure 3: Delayed mode data flow

Having proceeded through the delayed mode QC process, the data then follow the same route as the real time data through the rest of the CMD process; however, on a different time schedule because of the

more irregular times of arrival. During the merging of the data into the CMD, any duplicates occurring between near real-time and delayed mode data sources are identified with the highest resolution copy being retained as the active CMD version.

Acquisition of delayed mode data from the Principal Investigators is a priority for the GTSPP. It is necessary for GTSPP participants to work diligently in a pro-active manner with both the Principle Investigators and the data managers of the global programmes that collect the temperature and salinity data. The goal is to get the delayed mode data into the CMD within one year or less of its collection. An excellent way for any national oceanographic data centre to support GTSPP actively is to obtain national data sets of temperature and salinity data, apply GTSPP QC procedures, and submit them to the US NODC for inclusion in the CMD.

6.1.3 Communications Infrastructure

The ability to acquire data and to ensure that it is provided to the GTSPP Continuously Managed Database in a timely manner is dependent on the communications infrastructure. Over a period of time, the objective is to move to faster and more automated means of communications for data acquisition and dissemination, and for the dissemination of data products and information.

The Internet and the World Wide Web have become very important in the dissemination of data, information products, and information concerning the programme. Some of the products have near real time uses, the applicability of which can be greatly enhanced by their timely distribution. Publication of such products on the World Wide Web has become the option of choice to provide the data quickly to clients that have access to the Internet. For users in member states that do not have access to the Internet, other options for telecommunicated products have to be implemented. For example GTS data flow products have been circulated to some member states by fax. Member states that cannot receive products and data available from GTSPP on the Internet should contact the IOC Secretariat with a request that some other arrangements be made.

6.1.4 Quality Control Procedures

The quality control element continues to be crucial to the successful operation of GTSPP. It is essential that all data maintained in the GTSPP Continuously Managed Database (CMD) is of a known quality and that the scientific community, including the major global research programmes, accept the validity of the quality control flags attached to the GTSPP data.

The philosophy of GTSPP is to create a complete database. This is achieved partly by retaining all data that meet the minimum level of quality. Once an observation has successfully passed the first level of checks (which consist of relatively simple data structure, format and range related tests), it will not be removed or deleted from the CMD except in the case of exact duplicates.

The GTSPP Quality Control Manual covers the checks that are undertaken by participating data centres. It describes the approach of using a small number of quality control flags by using a series of data quality levels to identify data validity.

The quality control levels range from a minimum series of checks up to the more sophisticated numerical assessments. All the data that have been checked and have passed the lowest quality level criteria are included in the CMD with the appropriate flag and it will then be available to the user.

A significant feature of the GTSPP quality control procedures is that they form the basis for a global quality control standard. The scientific community accepts this standard because they have been heavily involved in its development. Gaining this acceptance has been a major achievement for GTSPP. It is very important that the scientific community continue to be involved in suggesting, reviewing and agreeing to GTSPP quality control procedures.

It was decided at the Fourth Session of the Steering Group on the GTSPP that it is desirable and necessary to transfer the scientific QC procedures being done by the WOCE UOT Science Centres to the data centres. This "transfer of technology" to the data centres is a positive step in its own right. Upgrading data centres skills has been a goal of IODE and can only lead to better data centre performance and services.

On the other hand involvement of the scientific community in the development and maintenance of GTSPP QC procedures and the skills of the data centre persons performing the QC must be maintained at a significant level to ensure continued success. GTSPP through its participating centres must work to maintain effective working arrangements with the science programmes to ensure the adequacy of its QC activities. Only in this way can the continued validity and acceptability of the GTSPP data set be guaranteed.

6.1.5 Continuously Managed Database

All the activities of GTSPP, including data acquisition, communications, data flow analysis, data quality control, and scientific analysis, result in the creation of the global temperature and salinity data set. To manage this data set, a Continuously Managed Database (CMD) was implemented.

As data are acquired in both near real-time and delayed mode, they are added to the database. Delayed mode data have a higher resolution and are calibrated and quality controlled by the originator. Thus the delayed mode data represent a "better" version of the data for all purposes. This "better" version replaces the data obtained in near real-time. This version of the data arrives later than the real-time version. The Continuously Managed Database therefore holds the most current and highest quality data set available at any given time. Its contents will be refined continuously as additional quality checks are undertaken. The term replace here means replace as the active copy of the observation in the database. As stated earlier, observations that have passed quality control and entered the database are not removed. They are flagged to indicate that a higher quality version of the observation exists in the database.

6.1.6 GTSPP Data and Information Products

GTSPP has a role in the production of two types of data and information products. The first role is a direct one whereby GTSPP actually generates the product as part of the operation of the programme. The second role is in providing a higher quality and more complete data set to others who are producing products and information about ocean temperatures and salinities.

It is important that GTSPP compliments rather than competes with other national and international organizations in generating products. The GTSPP role is to provide a high quality data set that will assist these organizations in producing higher quality products. An example would be GTSPP providing the most complete and best quality in situ data set available for a numerical modelling study.

By supporting the work of the scientific experts in other organizations GTSPP can expand its expertise in uses of the temperature and salinity data as well as assisting the work of the experts. These experts are in many cases in a better position to produce scientific analyses and products because of special expertise and local knowledge.

The development and distribution of data and information products helps to promote and advertise the GTSPP. The products demonstrate additional uses for the data and show the effectiveness of GTSPP as a data management model and as an important contributor to the global research programmes.

6.1.7 Data Flow Monitoring

The implementation and operation of GTSPP involves the use of considerable resources, including manpower, computer systems, and communications networks. In order to maximize the effectiveness and the data returns for the investment, it is essential that the data flow of the system be carefully monitored.

The monitoring programme is intended to assist all participants, including data contributors and user groups, to have knowledge of the effectiveness of the programme and to modify aspects that are not performing efficiently. It also provides an indication of where improvements in the system are required as the programme develops.

6.2 Real-time data assembly

The development of the Global Ocean Observing System (GOOS), and components of the GODAE require data to be delivered in real-time. For many years the Global Telecommunications System (GTS), operated by member states of the WMO, was the only widespread communications network for exchanging data on a global basis. In recent years, the development of the Internet is challenging this exchange network. The GTS is still the dominant source of data for the GTSPP but the Internet as an alternate exchange network is becoming increasingly important.

Control of the data structures and transmission protocols for the GTS is vested in WMO committees. There is a strong formalism required to create or modify these standards. The rigidity of the change control process is both an asset and a liability. As yet, there are not the same standards used for exchanging data using the Internet. So, the system is much more flexible, but suffers from a lack of standards.

The real-time data assembly process for the GTSPP entails a number of functions. These include the following

- Gain access to the GTS and Internet. For the GTS, this means gaining either a direct access, or making arrangements with a national meteorological service to get access to the data exchanged on GTS. Gaining access to the Internet may be simple or not depending on national rules and accounting and the available telecommunications circuits coming into a country.
- Acquire the data. For data coming from the GTS, it is necessary to develop software to read the GTS messages (ASCII and binary forms) and convert the contents to a more usable form. Despite the formats being standardized on the GTS, there are always instances of miscoded messages. For data derived from the Internet, the main task is dealing with the multiple formats and the varied contents of data files. Such is the consequence of little to no standardization
- Manage duplications and data versions. The present communications protocols of the GTS make it prone to data losses and data duplications. One of the necessities of the GTSPP has been to acquire all of the BATHY and TESAC messages as seen at 4 nodes on the GTS. There is, of course, a very great degree of duplication, but it is not complete from one node to another. It is only by accessing these 4 nodes that the GTSPP has been able to be confident that all of the real-time data have been acquired. For the Internet, it will be necessary to check that multiple copies of the same data are detected as well as differing versions resulting from successive application of quality control procedures by different components of the Argo data system.
- Contribute to standards setting. It is important for the oceanographic community to interact with the WMO committees responsible for maintaining the code forms of the GTS. The GTSPP participants are well placed to fill this role because of their familiarity with the GTS and the code forms used.
- Upgrade the assembly process with changes in technology. In recent years, there is the drive to increase the vertical resolution of the temperature and salinity profiles reported in real-time. The real-time data assembly centre must work with the oceanographic collection system to be sure that the data system can accommodate the increased resolution.
- Provide access to the data. Some clients of the GTSPP need the data very quickly after collection and are willing to trade speed of access for reduced data quality. GTSPP participants handling the real-time data need to take these users into account in building and maintaining their data systems.

6.3 Delayed mode data assembly

Format and real-time transmission bandwidth limit the amount of data and information that can be exchanged in real-time. Often the delayed mode data stream has higher vertical resolution and almost certainly higher measurement precision. The increased information content means that the delayed mode data are the most desirable.

When assembling the delayed mode data it is important that the following functions be included.

- Acquire high resolution data. High vertical resolution and high measurement precision data should be sought out from all sources. This requires an active role in finding the agencies that make the data collections and developing a working relationship to acquire the data in a timely way. There are many avenues that can be used including cruise reports, monitoring real-time data submissions, and personal contacts.
- Accept and include other profile types. It is not unusual for temperature and salinity profile data to be collected in combination with other profile types. Although the GTSPP is not focused on these other types, when these other data have been collected with the T and S profiles, it is much better to keep all of these measurements together.
- Identify data sources. Data sources should be identified with the data because it assists in the process of identifying versions of data, and in knowing who to contact if questions arise about the data or information provided.
- Connect real-time to delayed mode data. All data submitted in real-time should eventually be replaced by their delayed mode versions. When the delayed mode data do arrive, it is important to maintain information that allows the data to be linked to the real-time data that were collected. Inevitably, other versions of the real-time data will exist, and keeping this linking information permits duplications to be avoided.
- The acquired data should be converted to GTSPP standards. This includes documentation of the software used in the conversion, in recording ancillary information about the collection (such as who, what project, what instruments, measurement precision, etc.), and preserving data even though there is some question about its quality.
- Continue to suggest improvements. As technology changes and needs change, there is a continual need to keep data management operations in tune to requirements. It is important to consider in an on-going way where improvements can be made.

6.4 Continuously Managed Database

A goal of the GTSPP is to provide the highest quality, highest resolution data to clients as soon as possible after the data are collected. Meeting this goal is embodied in the concept of the Continuously Managed Database (CMD). This is the archive where all of the real-time and delayed mode data come together, where version control is exercised and where clients can access the data.

The functions that are fulfilled by the CMD include the following

- Match delayed mode data to real-time versions. This function is complicated by the fact that virtually all information that can be used to make the match is also subject to revision by quality control processes that occur before the data arrive at the GTSPP centres. This matching process is greatly assisted if information about real-time data submitted accompanies the delayed mode data.
- Control versions. Data are continually being worked and reworked, and this results in problems being found and revisions sent to the CMD. Version control is effected by contributors proving a history of what was done.
- Implement standards. Where standards exist, such as the GTSPP format, it is important that data meet the standards. Where standards do not exist, efforts should be directed to define and then implement new standards,
- Interact with global archives. GTSPP manages only a portion of the ocean profile data. The high level of scrutiny characteristic of GTSPP data means that it constitutes the highest quality version available. It is important that the global archives are closely liked to the GTSPP archives.
- Continue to suggest improvements. As technology changes and needs change, there is a continual need to keep data management operations in tune to requirements. It is important to consider in an on-going way where improvements can be made.

6.5 Quality Control

A major task of the GTSPP is assuring a high level of data quality. This is accomplished through a staged approach where data centres handle initial checking, followed by scientific scrutiny carried out by science centres. This cooperation is one of the great strengths of the GTSPP and ensures that the very best quality data are available at any time.

The quality control functions include the following

- Evaluating the quality of real-time data. Real-time data are distributed in order to meet problems with needs for rapid data delivery. There is a tradeoff of timeliness against the degree of scrutiny that can be carried out. In the present operations of GTSPP, this level of quality control is carried out by data centres.
- Evaluating the quality of delayed mode data. These data constitute the scientific data resource. Normally, the time constraints for delivery of these data to clients is more relaxed than for real-time data. Dealyed mode data are the highest resolution data available and therefore deserve the highest level of scrutiny. This high level of scrutiny requires manual processes by people familiar with the oceanographic regions from which the data were collected. In the present operations of GTSPP, this level of quality control is carried out by science centres.
- Implement standards. It is important that clients can be confident that even though data quality assessment may be distributed, that the data assessment is being carried out in an equivalent manner everywhere. This requires the definition and implementation of standards at all centres carrying out quality control.
- Validation and improvements of present practices. In the course of using tools to assess data quality, it is important to be aware of the weaknesses of existing tools and to both suggest and implement changes in evaluation procedures. Experiences and knowledge need to be traded between participants to broaden the capabilities of the entire project.
- Documenting procedures. Quality control procedures need to be well documented in order that users of the data can be informed of what is done and have confidence. This applies equally to both real-time and delayed mode procedures.

6.6 Duplicates Handling

In the effort to acquire as much data as possible, the GTSPP has built some redundancies in data acquisition. This has proved to be a worthwhile thing to do, but has also ensured that identifying duplicates is an important function.

Duplicates identification includes the following.

- Eliminate multiple, identical copies of data. The real-time data system is built to guarantee multiple copies of identical data. Both real-time and delayed mode data streams are subject to the same data being received more than once. It is important to have tools that reliably identify exact duplicates and to select the best copy to retain in archives.
- Identify and manage data that are nearly identical. Transmission errors, corrections en route and other processes guarantee that data derived from the same original observations, appear with other than identical content. Tools to identify likely duplications are needed and to provide assistance in selecting the best copy to be retained.
- Identify real-time and delayed mode versions of the same data. Typically, data distributed in real-time has lower resolution (vertical or measurement), reduced numbers of variables, and uncalibrated measurements compared to what is received in delayed mode submissions. Tools are required to match the real-time and delayed mode versions of the data so that only the highest resolution version is delivered to archives and clients.
- Recommend improvements. New strategies for the tagging of data and identification of duplicates should be considered for their practicality and implementation in the GTSPP.
- Define, implement and document standards. It is important that GTSPP set and implement standards for determining the most desirable version of data (generally the version with the highest resolution and highest quality). It is equally important that clear documentation of the processes used in duplicates handling be well described and readily available to clients.

6.7 Data and Products

A primary purpose for the GTSPP is to deliver both data and products to users. Data can be provided by many participants in the data system depending on the requirements on speed of delivery, data quality and resolution. Products include notifications of data collected, as well as analyses that present ocean conditions.

Data and products include the following.

- Notification of data collections. Some users want to know what real-time data have been collected in given regions and in given time frames. Other users want reports of data available over all time. Both of these are provided within the GTSPP data system.
- Rapid access to data. Some users require the data collected by real-time distribution systems, but to which they have no access. GTSPP provides this function. Others wish to have data with some quality control applied, but in a fairly short time frame after data collection.
- Integrated data. Data collected by a variety of instruments often arrive at the GTSPP in mixed data formats. By assembling all of the data into a single structure, and applying standard procedures, the program can provide an internally consistent data collection to a user.
- Analyses of ocean properties. Access to collections of data provides the resource on which scientific analyses can be based. These analyses are typically done in the domain of the Science Centres of the program and support their research needs. Equally, it would be possible for a collaboration between science and data centres to provide certain scientific products. At the same time, these products can be of use to secondary users.

6.8 Monitoring and Feedback

The GTSPP data system manages data that are distributed within hours and other data that do not become public for months to years. In addition, when problems are noted in data collected, it is important to alert the collectors so that remedial actions can be taken. All of this requires a deliberate and routine monitoring of data system functions (collection, distribution, archiving) with appropriate feedbacks provided.

Monitoring and feedback functions include the following.

- Monitor the real-time data stream. A number of monitoring procedures are needed ranging from assisting WMO in evaluating GTS performance, reporting of data received from various sources, alerting collectors when data quality problems are found, assessing the quantity and variety of data distributed in real-time, and reviewing the success of international programs in meeting sampling objectives.
- Monitor the delayed mode data stream It is important to monitor if the higher resolution delayed mode data have been received to replace the real-time stream, to assess the quality of the data and interact with data providers, watch the volume and timeliness of data receipts and strive to improve the data submission process.
- Assess completeness of the data archive. Because data arrive at both short and long time scales, at low and high resolution and from multiple sources, there is an ongoing need to examine the completeness of archives. Where data are known to exist and either have not reached GTSPP archives at all, or only in lower resolution forms, it is important to initiate actions to seek out the data.
- Continue to suggest improvements. As technology changes and needs change, there is a continual need to keep data management operations in tune to requirements. Monitoring reports have the advantage of showing trends in processes monitored and help assess the success of a program to meet objectives. As changes take place, changes and improvements to the monitoring and feedback processes are needed.

7. Cooperation with Other Programmes

When the Global Temperature-Salinity Pilot Project began, the first elements of the World Climate Research Programme (WCRP) and the Tropical Ocean and Global Atmosphere (TOGA) Experiment were being implemented. The World Ocean Circulation Experiment (WOCE) was being developed.

In fact GTSPP was developed to respond to the needs of TOGA and WOCE and had a very successful working relationship with those programmes. Since that time TOGA has been completed, and WOCE has moved from the data collection phase of the programme to synthesis and analysis activities. Two new programmes have been established and GTSPP is turning its attention to serving those programmes as well as continuing to work with WOCE. The new programmes are the Global Ocean Observing System (GOOS) and CLIVAR (the Climate Variability Programme). The GTSPP principle continues to be to work in cooperation with existing programmes that either require or collect and manage temperature and salinity data. Both GOOS and CLIVAR will require the use of T&S data and be involved in specifying requirements for its collection. GTSPP will work with its parent committees of IGOSS and IODE to implement mutually beneficial programmes with GOOS and CLIVAR.

The following sections outline the areas of potential additional services to existing or planned programmes.

7.1 IODE – GODAR - WDCs

The Fourth Session of the Steering Group (Washington, DC, USA, 16-19 April 1996 dealt with cooperation with the Global Ocean Data Archaeology and Rescue (GODAR) Project, redefined the limits of the GTSPP responsibilities, and reviewed the roles of the science centres and the data centres in regard to quality control of the data. Decisions were made that GTSPP would deal only with profile data and that aspects of the GTSPP related to historical temperature and salinity data collected before 1990 would be left to the GODAR Project. In each case GTSPP had not yet developed the ability and systems to manage these other data. In each case another competent organization had offered to undertake the work as a contribution to international data management.

At present, the GTSPP archives are routinely passed to the GODAR archives and these in turn contribute to the archives held in WDCs, World Data Centres. The GTSPP data constitute an important component of the GODAR products and form a significant portion of the data from the 1990's onwards.

GTSPP is a joint WMO IOC project. As one parent body, the IODE committee requests a report from the GTSPP at each of its meetings. The GTSPP also takes direction from IODE.

7.2 JCOMM

The Joint Commission on Oceanography and Marine Meteorology (JCOMM) was formed from the joint IOC / WMO IGOSS programme and the WMO Commission on Marine Meteorology. JCOMM's responsibility is to foster improvements in the collection, processing and distribution of data collected at sea. The GTSPP, being a joint IOC/WMO programme now falls under the umbrella of JCOMM and thus reports to the Data Management Programme Area, DMPA.

In addition, the GTSPP forms the data management component of the Ship Of Opportunity Programme (SOOP) which reports to the Observations Programme Area of JCOMM. Supporting the SOOP was a natural outgrowth of the GTSPP cooperation with the WOCE UOT programme (described later). At the first DMPA Coordination Group Meeting, held in Paris in June, 2002, the GTSPP is being asked to participate with appropriate experts from the Aerological Sounding Acquisition Project, ASAP, and Volunteer Observing Ship, VOS, programmes to look for ways to cooperate in data management. The goal of the cooperation is to provide standardized techniques for handling data, and a unified data product.

7.3 GODAE - Argo

The Global Ocean Data Assimilation Experiment (GODAE) is an initiative to include ocean observations into weather and climate prediction models. The Argo programme is a sub-programme of GODAE concerned with providing a global collection of temperature and salinity profiles on time and space scales required by GODAE. The GTSPP is making both short and long term contributions to Argo.

In the short term, the real-time data stream acquired through the GTS by GTSPP centres has been made available to the Argo data system. This has allowed Argo to demonstrate early access to data which is critical in helping the new programme attract supporters.

In the longer term, GTSPP has contributed ideas in data handling that are reflected in the Argo data management system. Many of the contributors to GTSPP are important contributors to the Argo data system. Argo has been able to build on the experiences learned by GTSPP. The experience contributes to the management of real-time data, to formats and contents, to data system structures and the archive functions.

An important continuing contribution to Argo will be the integration of data from profiling floats with data collected by other profiling instruments. This is crucial to carrying out the intercomparison studies that are needed to examine differences in measurements caused by different instruments. In addition, as GTSPP contributes to the integration of profile with upper air and along track measurements through participation in JCOMM, this will be brought to support Argo data collection.

7.4 GOOS

The Global Ocean Observing System is responsible for the development of future ocean observing systems that will address issues including climate change and prediction, health of the oceans, marine living resources, the coastal zone, and ocean services. The Intergovernmental Committee for the Global Ocean Observing System (I-GOOS) co-ordinates member states efforts for establishing GOOS. The Joint Scientific and Technical Committee for GOOS (J-GOOS) is the body that supplies scientific and technical expertise to the development of the observing systems. The intention is that the data collection and management aspects of GOOS be built on the existing systems of JCOMM, IODE, GLOSS, DPCP, and others. Thus GOOS is a future client for GTSPP services.

As with CLIVAR, GTSPP was developed to meet the perceived needs of GOOS. The GTSPP model is regarded as a prototype for the data management systems that will be needed for many aspects of GOOS. In particular GTSPP has and will continue to put significant effort into the development of quality control systems, data flow monitoring, and the development of the continuously managed database concept. These developments have placed GTSPP in an excellent position to contribute significantly to the implementation of the data management aspects of GOOS.

One of the important initiatives of the GTSPP is focused on the provision and distribution of useful data products and information to the client community. The initial approach to this aspect of the programme is to develop a distributed set of World Wide Web sites among the UOT DAC sites with data products and information for clients.

7.5 CLIVAR

CLIVAR is a 15-year Climate Variability (CLIVAR) programme to study climate variability and predictability. It is the primary WCRP programme for the study of the role of the ocean in the coupled climate system. It can be regarded as a programme that builds on the knowledge gained through TOGA, WOCE, and the 1995 study of anthropogenic effects on climate change that was carried out by the Intergovernmental Panel on Climate Change.

CLIVAR anticipates that the IOC and WMO and their subsidiary bodies will work together to codify and monitor the standards for making, processing and archiving environmental observations for operational needs in the terrestrial, marine, and cryospheric environments. This will lead to the maintenance and

continued growth of a reliable global climate record. CLIVAR also anticipates that IOC and WMO will institute and maintain an international infrastructure for encouraging, assisting and coordinating individual nations contributing to the deployment and maintenance of operational observing systems.

To serve CLIVAR's needs these data must be provided in a timely fashion. It is also important to ensure that as much data as possible that are collected by observing systems deployed for research purposes are inserted expeditiously into the operational data processing and archiving streams.

GTSPP was designed with requirements for programmes such as CLIVAR in mind. GTSPP has been organized to provide data in the most timely manner possible and with early as possible examination and documentation of the data quality. GTSPP also has made significant progress in acquiring data collected by the oceanographic research organizations to make it available to national and international clients. GTSPP therefore expects to provide continuing support to climate research and prediction by supporting CLIVAR as it has supported TOGA and WOCE.

7.6 WOCE - UOT

The WOCE programme consisted of a number of experiments. One of the experiments was the Upper Ocean Thermal Programme (UOT). The main data requirements of the UOT were for heat and salt distributions of the upper ocean. The development of accurate distributions of these variables requires the most complete and accurate temperature and salinity databases. GTSPP helped with the management of these data during the data collection phase of the WOCE programme through participation in the WOCE UOT Data Assembly Centre (DAC).

Another important focus of the GTSPP programme was in working with the WOCE Data Products Committee (DPC) to build the UOT part of the WOCE Data Resource. The WOCE Data Resource is an active distributed archive of the WOCE data that will continue to make the data available to users after the completion of WOCE. In many cases the Data Resource will continue to grow where observation systems are put in place to continue important WOCE measurements.

Another area of WOCE - GTSPP cooperation is related to the TOGA/WOCE Subsurface Data Centre that had been established in Brest, France. This centre operates a database with sub-surface temperature and salinity data for a specific area of ocean and time frame of interest to TOGA and WOCE. There is a regular exchange of data between the Brest Centre and the GTSPP Centre in the US NODC in Washington DC. This exchange ensures that each centre has copies of all available data relevant to its mandate. GTSPP will continue to work with the Brest Centre to ensure that both Centres are working with the complete database.

Annex A: Initial Objectives of the GTSPP

The objectives at the beginning of the Pilot Project were as follows.

(i) To create a timely and complete data and information base of ocean temperature and salinity data of known quality in support of global research programmes and of national requirements.

(ii) To improve the performance of the Intergovernmental Oceanographic Commission (IOC)/ IODE and World Meteorological Organization (WMO)/IOC IGOSS data exchange systems by actively pursuing data sources; exercising the data inventory, data management, and data exchange mechanisms as they are intended to work; and recommending changes where necessary to meet national and international requirements.

(iii) To disseminate, through a widely distributed monitoring report produced on a regular basis, information on the performance of the IODE and IGOSS systems.

(iv) To improve the state of databases of oceanographic temperature and salinity data by developing and applying improved quality control systems to these databases.

(v) To improve the completeness of these historical databases by the digitization of historical data presently in analogue or manuscript form and by including digital data not presently at a World Data Centre (WDC). **Note that this objective has been undertaken by the GODAR project of IODE. GTSPP is concerning itself with data collected in 1990 and later.

(vi) To distribute copies of portions of the database and selected analyses to interested users and researchers.

Annex B: Terms of Reference and Composition for the Steering Group on the Global Temperature-Salinity Pilot Project

Terms of Reference

(i) Complete development of the GTSPP Project Plan and Implementation Schedule following the principles described in the current draft.

(ii) Complete development of the GTSPP Quality Control Manual.

Meet semi-annually at the expense of the participating countries to review the status of the implementation and further develop the GTSPP.

Actively promote the GTSPP and provide information to the users of GTSPP services, such as the planners of international science programmes.

Provide scientific and technical guidance to GTSPP participants in the implementation and further development of the scientific and data management aspects of the six GTSPP elements including:

a) near real time data acquisition
b) non real time data acquisition
c) communications infrastructures
d) quality control procedures
e) continuously managed database
f) GTSPP data and information products

Prepare, maintain and distribute documentation relevant to operation of the project.

Report after each meeting and as otherwise necessary to keep all IODE and IGOSS contacts, as well as the representatives of the science programmes, informed on the status of implementation of the GTSPP.

Submit status reports on the GTSPP to the sessions of the Committee on IODE and the JWC-IGOSS.

Prepare and submit to the sessions of the Committee on IODE and the JWC-IGOSS a report on the status of IGOSS-IODE data flow.

The group will select a Chairman at its first session and will review the Chairmanship biannually. The composition of the group is as proposed below. The composition of the group will be regularly reviewed in consultation between the Chairmen of IODE and IGOSS.

Composition

One representative from each of the participating countries (initially Australia, Canada, France, USA, Russian Federation) as chosen and funded by the countries, to provide expertise to the project. One or more experts may accompany these representatives;

(ii) Representatives from the WOCE IPO, and the WOCE UOT Programme to provide scientific guidance to the project;

Additional invited scientific advisors selected by the Chairman of the Steering Group on the GTSPP;

Representatives from the RNODCs-IGOSS are invited to participate in this Group.