

PRECEDING PAGE BLANK NOT FILMED

N 93 - 24924

Global Temperature and Salinity Pilot Project

Ben Searle

1 Introduction

I am very happy to have been asked to give a presentation on the Global Temperature and Salinity Pilot Project to this group of scientists and data managers since you are the people that will ensure the success of this important project. I have personally been involved with GTSPP since the idea for this project was first proposed in 1988 and I am the Australian member of the GTSPP Steering Committee.

Data exchange and data management programs have been evolving over many years. Within the international community there are two main programs to support the exchange, management and processing of real time and delayed mode data. The Intergovernmental Oceanographic Commission (IOC) operate the International Oceanographic Data and Information Exchange (IODE) program which coordinates the exchange of delayed mode data between national oceanographic data centres, World Data Centres and the user community. The Integrated Global Ocean Services System is a joint IOC/World Meteorological Organisation (WMO) program for the exchange and management of real-time data. These two programs are complemented by mechanisms that have been established within scientific programs to exchange and manage project data sets. In particular TOGA and WOCE have identified a data management requirement and established the appropriate infrastructure to achieve this. Where does GTSPP fit into this existing frame work.

For a number of years individual data centres and scientists have been thinking about the benefits of developing centralised continuously updated oceanographic data bases covering the global ocean. In the past there have been a number of barriers stopping the implementation of this dream. In particular the state of computer technology and the relative immaturity of communications systems have long prohibited such a concept. However, today these technologies have reached a level of sophistication and stability that allows us to address these problems.

Another important advance that has brought us closer to the realisation of a truly global database is the increased level of cooperation that has evolved out of necessity for global scale data collection and research programs. With an increasing emphasis on climate change and assisted by improvements in technology these large scale research projects are becoming common and the development of global ocean databases is finally feasible. The Global Temperature and Salinity

C-2

4

Pilot Project has evolved from the need for an improved data delivery system to support the global research effort.

2 History of GTSP

At a Meeting of Experts on IODE/IGOSS Data Flow held in Ottawa in January, 1988 data management experts from several countries were looking for ways to promote and improve data exchange programs. A suggestion was made that a global scale data base of ocean thermal data that was freely available and of a high quality would provide the most effective means of promoting data exchange. The development of this database and its dissemination to the scientific community would demonstrate the advantages of participating in data management programs.

As a result of the Ottawa meeting and through the endorsement of the participants at the February 1988 Wormely meeting of Experts on RNODC's and Climate Change and through other meetings held in the United States it was realised that the global data set concept should be developed into a pilot project.

In November, 1988 a proposal on a global pilot project was presented to the IGOSS V Committee by Canada and the USA. This proposal differed only slightly from the plan developed earlier that year in Ottawa with the inclusion of salinity as a second parameter for the global data set. The addition of salinity resulted from discussions held at various meetings between the scientific user community and the data management agencies including the NODC/ERL Workshop on Ocean Data Files in the US during June, 1988. As a result of the Canadian and US proposal IGOSS adopted Recommendation 4 which strongly supported GTSP.

To formalise the project and start planning and development two ad-hoc meetings were held in 1989. The first was in Washington during January, and provided the forum to discuss many of the technical issues and examine potential problem areas. The main area of contention, which seems to occur at most data related meetings, involved the quality control component of the project. This issue was discussed in detail but complete agreement on this complex problem could not be reached. However, the meeting did define the goals and the major elements of GTSP.

A second ad hoc meeting was held in Ottawa in July, and further progressed the project. Participants from existing research programs such as TOGA and WOCE provided their thoughts and experiences on oceanographic data management. The relationship between GTSP and other existing research and data exchange programs was also discussed at great length. At this meeting it was stressed that GTSP would complement and support existing data flow mechanisms such as IGOSS and IODE. It was also agreed that the most significant user group would constitute scientists working within elements of the World Climate Research Program (WCRP) such as TOGA and WOCE. The need for close dialogue

with all related scientific areas during the initial development and throughout the operation of the project was stressed.

Prior to IODE 13 in New York in January of 1990 a workshop was held to gain input from a wide range of marine scientists and data managers to further refine the projects development. At this meeting the terms of reference and composition of a Project Steering Committee were developed. The draft Project and Implementation Plans were reviewed. The concept for a pilot project was presented to the IODE Committee and was well supported.

As a result of these various meetings and with considerable liaison with the scientific community GTSP was born. To date there have been two meetings of the Steering Committee. The first was held in Brest, France during September, 1990 and the second was held in Obninsk, USSR during July, 1991. By this time a considerable effort had already gone into the planning and aspects of the implementation of the project.

During the initial planning of GTSP we were able to call on the experience of the US NODC who had recently established the Joint Environmental Data Analysis (JEDA) project with interested scientists from the Scripps Institute of Oceanography. JEDA developed from the need to involve science in the data management cycle, a weakness that was becoming more apparent as the demand for high quality data increased. JEDA revolves around the exchange of data between NODC and Scripps, with Scripps undertaking scientific evaluation of the data sets to determine their validity. A number of techniques are employed to check the data and as a result various products are developed. JEDA has proven to be a very successful concept.

3 Aims of GTSP

GTSP is aimed at addressing some of the problems that exist in present data exchange and management structures so that data delivery systems suitable to meet the demands of the 1990's can be provided. This pilot project is using a number of previously untried techniques and is taking full advantage of developing technology in data processing systems and communications. It will prove or perhaps disprove technologies and procedures necessary to effectively manage oceanographic data on a global scale.

GTSP will build on the existing data management structures such as IGOSS and IODE and provide a testing ground for proving new methodologies and procedures. GTSP is an exciting project that has already proven to be a successful venture of cooperation between data management and scientific agencies from a number of countries. As it continues to develop GTSP will provide the global marine community with a data exchange and management capability that will be the envy of other physical sciences. We have one advantage over other disciplines. GTSP is based on the collective experience of a number of data managers and

scientists but is starting from a fresh perspective with the advantage of modern technology which has only become relatively stable and robust in the last two or three years.

4 Objectives and Goals

GTSP has a number of specific objectives which include :

1. To create a timely and complete data and information base of ocean temperature and salinity data of known quality in support of the World Climate Research Program (WCRP) and of national requirements.
2. To improve the performance of the IOC's IODE and the joint IOC/WMO 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.
3. To disseminate, through a widely distributed monitoring report, produced on a regular basis, information on the performance of the IODE and IGOSS systems.
4. To improve the state of historical databases of oceanographic temperature and salinity data by developing and applying improved quality control systems to these databases.
5. To improve the completeness of these historical databases by the digitisation of historical data presently in analogue or manuscript form and by including digital data not presently at a World Data Center (WDC).
6. To distribute copies of portions of the database and selected analyses to interested users and researchers.

Since GTSP became operational an aspect of each of these objectives has been realised. As the pilot project continues and broadens to cover other identified activities it will move closer to fully meeting its objectives.

5 Structure of GTSP

GTSP consists of three main elements: 1) the data management component, 2) quality control, and 3) data flow.

5.1 Data Management

The data management component consists of the two main centres which are MEDS and the US NODC. MEDS is the focal point for real time data and undertakes the GTSP Quality Control checks on data received via GTS sources on a daily basis. NODC is responsible for the management of the main GTSP database which is a Continuously Managed Database and incorporates both real time and delayed mode data streams. The real time data stream comes to NODC on a weekly schedule after initial quality control at MEDS and delayed mode data comes via existing IODE sources.

5.2 Quality Control

On a monthly basis NODC produce data sets for the three ocean basins and these are made available to three scientific centres. Each month data for the relevant ocean region is checked by Scripps Institute of Oceanography, the Atlantic Oceanographic and Meteorological Laboratory and Australian Bureau of Meteorology Research Centre. Flags relating to data quality are returned to NODC for incorporation into the Continuously Managed Database

5.3 Data Flow

The communication of data and issue of products is an important part of GTSP. The capture of data is generally through the existing data exchange programs particularly IODE and IGOSS. However, data is also acquired directly from the various scientific programs through their data exchange mechanisms.

Real time data is transmitted by the WMO's GTS which is supplemented by NASA's SPAN network between the US and Canada. SPAN is also used to transmit the monthly data sets to the scientific centres responsible for quality evaluation. Delayed mode data will be acquire through the existing IODE channels and from the relevant WOCE and TOGA Data Centres. Other components of GTSP such as time series data will be implemented as resources allow.

The initial priority of GTSP was to establish the data flow for real time data to support operational activities. This has now been effectively implemented and efforts are proceeding well for the commencement of the delayed mode data stream. This stream will involve the WOCE Upper Ocean Thermal Data Assembly Centres who will be responsible for the detailed scientific evaluation of this data set. It will be the task of NODC to integrate both data sets to form the complete database.

The inclusion of historical temperature and salinity data that has not already been incorporated in exchange programs is vital for the study of long term trends and changes in the ocean. A number of initiatives are underway to capture data that is generally in an analogue form. Many agencies including the Japanese Oceanographic Data Centre, ICES, MEDS, US NODC and USSR NODC are aiming to actively track down and digitise manuscript or analogue data sets. Australia is investigating means of acquiring data presently outside the normal exchange mechanisms both from within the country and from around the South East Asian region. The capture of these data sets by GTSP will result in adding tens of thousands of observations to the global data base.

6 Benefits of GTSP

GTSP has an obvious large scale objective, but it is also intended that participants will benefit at local, national and regional levels. GTSP will provide

benefits to the user community through the improvement of data management in a number of general areas. The project is attempting to:

1. Increase quantities of temperature and salinity data;
2. improve timeliness in distributing data to the User Community;
3. standardise the quality control procedures;
4. produce regional and global data products;
5. standardise data formats;
6. introduce new methods and technology in communications, quality control and data management and the progressive transfer of this to all areas of the marine community; and
7. provide a framework for future data delivery systems covering other data parameters.

6.1 GTSP Will Increase the Quantities of Data Available

It is common knowledge that not all oceanographic data collected at sea finds its way into national or international archives. Data is lost at many points in the cycle including at sea, in the coding of radio messages, in scientists filing cabinets, in communication systems and in data management activities. A significant result area of GTSP is the reduction of data loss in all components of the data cycle, particularly communications. The improvement in communication techniques will increase the data flow and providing more data to the end user. The TOGA Subsurface Data Center in Brest provides a good example to illustrate the benefits of improving communications. TOGA is continually refining its data flow mechanisms and this has shown excellent results.

As a general trend the transmission of data to the TOGA Center has increased from less than 10,000 XBT observations per year in 1985 to over 17,000 per year in 1987. Also, the time taken for the higher resolution delayed mode data to replace the near real-time data is steadily being reduced.

GTSP has already been able to increase the quantity of real-time data available to the world's research community through the acquisition of GTS data from three GTS nodes. By using the access to GTS links from MEDS, the US National Weather Service and the US Navy's Fleet Numerical Oceanography Center, the project is able to capture a greater quantity of data than available to each center individually. The data from each of the three GTS nodes is acquired electronically by MEDS where all duplicates are removed.

By using this approach MEDS has been able to capture data that previously may have been lost. A future aim of GTSP is to incorporate other GTS nodes from Europe, Asia and Australia. As these centres begin to provide data a larger number of real-time observations will become available for operational users. Given the relatively low quantities of data presently available in some areas such as the Indian Ocean, Southern Ocean and South Pacific it is crucial that every effort is made to ensure that each observation collected becomes available for global research. GTSP has been able to demonstrate that data losses through

communication difficulties can and have been avoided. Having proven this, GTSPPP will attempt to implement these procedures on a wider scale to further reduce data losses.

An additional benefit of this approach to data capture is the ability of GTSPPP to more accurately monitor data flow through existing channels. This monitoring system is used to identify gaps in the global dissemination of data via GTS. Monitoring within GTSPPP has already identified some major discrepancies in the quantities of data received by GTS Centers throughout the world and as a result efforts can be made to rectify the transmission faults. Improvement in the existing systems has already occurred.

The monitoring program has three main objectives.

1. to assist with the development of duplicate algorithms
2. to identify areas in the communication system that need improvement
3. to acquire the most complete data set

6.2 GTSPPP Will Improve the Timeliness of Data Flow

One of the aims of GTSPPP is to improve the timeliness with which data is acquired and made available to secondary users. The first stage of achieving this improvement has concentrated on the real-time data flow which has more operational relevance than delayed mode data.

Increasing the rate at which data is made available will be of considerable benefit particularly to operational users of oceanographic data. As greater quantities and more timely data becomes available for near real time analysis, the reliability and accuracy of operational products will increase. Real-time data available from GTSPPP is of high quality and quantity than data sets available from any other sources.

6.3 GTSPPP Will Assist in Standardisation of Quality Control Procedures

The introduction of standardised quality control techniques that have been accepted by the scientific community will be of major benefit to the users of GTSPPP data. This data set will be of a known standard and quality control indicators will be attached to each observation. Users will be able to examine documentation describing the exact processes used to qualify all data supplied from the GTSPPP database. Countries without national oceanographic data centers will be able to request data sets from GTSPPP and pass it to their own research communities with the confidence that the data is of a known quality. The quality control policy adopted by GTSPPP is that no observational data will be altered. Quality flags are allocated to describe particular tests that the data has been subjected to and the result of those tests is also given to show acceptance or rejection of the feature. However, gross errors occurring in the header details such as incorrect position time or date will be changed if the correct value can be determined with a high level of confidence. For example, a cruise at mid latitudes in the southern

hemisphere may have one observation with a northern hemisphere quadrant identifier. This is obviously a mistake and can be corrected. In this instance a flag is provided to show that a change has been made and the original value is retained in the observation's history. Experience has shown that if only one character can be changed to make a sensible record then the change is most likely accurate. The changing of more than one character is potentially dangerous.

One of the most important aspects of GTSP and a key reason for its success is the introduction of the scientific community into the quality control cycle. GTSP data passes through two main stages of quality checking. The first consists of the tests described in the GTSP Quality Control Manual and the second phase involves scientific agencies with considerable expertise in the water masses of the three major ocean areas. Atlantic data is checked by scientists from the Atlantic Oceanographic and Meteorological Laboratories, Pacific data is checked by Scripps Institute of Oceanography and Indian Ocean data is checked by both the Australian Bureau of Meteorology Research Centre and the CSIRO Division of Oceanography. Why was this elaborate procedure established?

6.4 GTSP Will Incorporate Science into the Project

During the early discussions on GTSP it was recognised that an area of concern with the existing data exchange mechanisms was the lack of active participation of the scientific community. This situation has in the past lead to a level of suspicion about the validity of data archived in oceanographic data centres.

The success of JEDA has provided the impetus for the other GTSP participants to involve the scientists from their own countries and JEDA confirmed that the development of effective data management systems for the next century required the active participation of the scientific community.

As result of this concept a symbiotic relationship is developing between research organisations and the data management agencies resulting in the development of high quality data sets necessary to effectively conduct research into major global and regional problems. During the 1980s the international data management community realised it could not operate in isolation from research and began the process of becoming more closely linked with the science component of the marine community. For example this Ocean Climate Data Workshop is another in a series of meetings and initiatives that have occurred with greater frequently in the last few years.

Today both the scientists and data managers are realising that they each have an important role to play in the development of high quality and comprehensive global data sets. Each element has specific expertise to contribute but individually each group has neither the resources or the desire to undertake such a large and complex task. Scientists and data managers are now supporting each other within the context of GTSP.

The key element provided by science is the accuracy of the data. Accuracy and reliability are becoming increasingly significant given the delicate nature of major world climatic events such as El Niño and the need to accurately monitor the variables that cause such events.

No previous or existing program has undertaken this integration exercise to the extent being accomplished in GTSP. There have been some difficulties associated with the development of this relationship and these relate primarily to the operational nature of GTSP. Scientific agencies are finding it difficult to justify the allocation of scarce resources to what initially appears to be a monitoring or data management activity. This apparent conflict between science and an operational program has been resolved successfully by the scientific agency integrating the need for GTSP data into their research programs. For example, in Australia the Bureau of Meteorology Research Centre (BMRC) have been developing upper ocean models as part of their contribution to WCRP. Access to GTSP data has provided them with a greater quantity of data of a higher quality to run their models. This improves both resolution and accuracy of the results. An outcome of the BMRC model is a mapping product which in turn provides scientific evaluation of GTSP data. This relationship has developed by integrating aspects of the operational GTSP data into BMRCs research activities.

Research agencies should not be hesitant in becoming involved in GTSP. There is a justifiable role for the research component in the operational activities of this project. If we as a group involved in oceanography are to provide answers to the ocean component of global climate research we must work closely together to produce a useable data set that can be used for research. The scientific community need not change its methods in order to contribute to operational activities such as GTSP but rather they need to be creative when looking to integrate their science activities within the project. GTSP will continue to prove that operational programs and data management activities can successfully be integrated with existing research objectives.

Research funding sources must be shown that the development of global scale, high quality data bases are essential for future large scale research efforts and that scientific participation in the data management cycle can in fact increase the value of research efforts.

6.5 GTSP Will Produce A Number Of Products and Services

To meet one of the more important objectives of GTSP a number of products and services are being developed or are presently available from the projects participants. For example Australia, Canada, France, USA and the USSR are already producing products and services based on GTSP data. A typical outcome of scientific evaluation of data is a mapping product. This type of 'value added' information provides an important and high profile output from the project.

One significant product that is now being planned is the publication of CD ROMs of GTSP data. This will be a major contribution to global and regional programs and will provide countries with limited resources access to high quality global scale data sets. CD ROM technology is rapidly developing into an ideal means of distributing data products and as prices continue to come down this technology will allow many countries' to participate in major international programs. GTSP is planning to release two CD-ROMs. The first will contain temperature and salinity profile data and the second will consist of time series data.

6.6 GTSP will Assist in Standardising Formats

GTSP is presently utilising digital technologies for the transmission and management of oceanographic data. Communications technologies are now allowing the rapid transfer of large quantities of GTSP data between centres. When GTSP was being planned it was anticipated that the WMO digital code BUFR would be used between centres. However, BUFR has not yet been finalised and the standardisation of exchange formats within GTSP is still undecided. However, the difficulty with formats is becoming less of a problem today given the power of computer systems and software to translate one format to another.

6.7 GTSP Will Introduce New Technologies

GTSP is relying on a number of new technologies for the implementation and operation of the project. Within the quality control area there is a greater reliance being placed on powerful computer workstations for the semi-automated checking procedures. Similar systems are used for the mapping analysis and workstations provide the user interface for interpreting, editing and flagging data. The operation of the Continuously Managed Database (CMD) is one area where technology is providing considerable assistance. NODC is in the process of developing a system using a database machine to optimise the performance of the GTSP database. This development project is called Poseidon and was discussed in detail yesterday. Poseidon will provide a system that has very large performance advantages over the more traditional software approach to databases. Communications is another part of the project that has benefited from improved or new technology. The exchange of data between MEDS and NODC is via a high speed data link available under SPAN and with the emergence of global data networks, data centres or scientists from other countries can now relatively cheaply gain rapid access to GTSP data. GTSP will examine new technologies as they are introduced to ensure that the capture and exchange of data and the dissemination of information and products is carried out by the most effective means. GTSP has a high level of flexibility that will allow the introduction of new technologies without requiring major alterations to existing components within the project. The Project is also aimed at transferring new technology to IOC Member States. Implementing GTSP has involved the introduction of 'state of the art' computer applications in the areas of automated and semi-automated quality control, artificial intelligence, graphical data manipulation and presentation, numerical modelling, and data

and information products. The experience gained, expertise developed and standard software packages produced during this project will become available to other countries to assist them in improving their national data management capabilities.

The IOC is developing a concept known as OCEAN-PC which will provide common software for the management and analysis of oceanographic data. MEDS on behalf of GTSP will be coordinating with the OCEAN-PC Project Leader to maximise the potential of the system to deliver data to GTSP and the user. Some of the technology that has developed as a result of GTSP could be incorporated into OCEAN-PC and therefore become available to the wider community.

6.8 GTSP Will Assist In Developing a Frame Work for Future Data Delivery Systems

GTSP has already demonstrated improvements in a number of areas, particularly with the capture of a more complete set of real time data. Developments in the areas of delayed mode data and time series data indicate that they will also prove successful. GTSP has gained a high level of confidence within TOGA and WOCE and this level of acceptance should widen even further. GTSP will still take time to overcome some of the outstanding problem areas and completely implement all aspects of the project. Once this has occurred the project will have been operational for some time and I am confident that it will live up to our expectations. It appears that throughout the 1990s there will be numerous monitoring and data collection programs operating. Each of these will require the support of data managers and scientists to ensure the full potential of the program is realised and no data is lost. The mechanisms developed for GTSP will continue to be enhanced and should be capable of meeting the future requirements for data delivery systems.

7 How Can the Scientific Community Assist GTSP

For its success GTSP relies heavily on the participation of the scientific community at both agency and individual level. Scientists can contribute by ensuring that their data is made available quickly, at least via real time mechanisms. Providing data has caused some concern in the scientific world, because of the fear that the data could be used by other scientists for research purposes. However, real time data transmitted over GTS is at a low resolution and generally only suitable for large scale activities such as mapping or for operational purposes such as supporting fisheries activities. By providing data in real time a scientist is ensuring that at least a low resolution copy of an observation is permanently archived. If the original, high resolution observation gets lost at least some useable information will remain. Data collected by WOCE is being provided in real time although the high resolution observation is not being exchanged for a period of about two years so the scientist can undertake the analysis without the fear of being copied.

The scientific community can further assist GTSP by continually reviewing the quality control procedures and providing feedback on problems. Since the research area will be a major user of GTSP data it is best placed to provide constructive criticism about data quality

8 Conclusion

I believe that GTSP will form the basis of the global data management and exchange systems required in the years to come. It is built on the experience of a number of data managers and scientists and for the first time in international data exchange programs incorporates the scientific community as a significant component of the system. GTSP incorporates these two elements which are essential in ensuring the success of this massive task.

For too many years data managers have worked away with little regard for the needs of the research community who require high quality data sets and scientists have frequently scorned the efforts of the data manager. GTSP is successfully demonstrating that both groups can work together effectively to develop high quality, global scale databases which are needed for the research task ahead.

It is the scientific element of this project that makes GTSP different from other attempts at managing data on a global scale. This is not reducing the role of the data manager who must also bring his experience and knowledge to bear on this problem. We as data managers have a significant role to play in the study of world climate and investigations into the changes that are occurring as a result of natural cycles and human activities. The developing cooperation between science and operationa programs must continue and I am confident that GTSP will be the vehicle for this.

Question Period

Q. What is "Delayed Mode" data? Will it include the types of thermal data sets being produced by WOCE?

A. Delayed Mode is used in this paper as it has been defined by IODE. It is meant to include the high quality data sets that may take quite a bit longer to process. The project does go beyond real time data.