Limited Distribution **IOC/IODE-XVI/32**
Paris, 17 October 2000
Original: English

INTERGOVERNMENTAL OCEANOGRAPHIC COMMISSION
(of UNESCO)

Sixteenth Session of the IOC Committee on International Oceanographic Data and Information Exchange (IODE), Lisbon, Portugal, 30 October – 9 November 2000

Project Proposal for IODE: The Surface Salinity Project

Submitted by R. Keeley, MEDS, Canada

Surface Salinity Project: Executive Summary

Justification

The Ocean Observations Panel for Climate and its predecessor examined the usefulness of surface salinity data in the context of climate change detection. They stated that "At high latitude, sea surface salinity is known to be critical for decadal and longer time scale variations associated with deep ocean over turning and the hydrological cycle. In the tropics, and in particular in the western Pacific, and Indonesian Seas, and in upwelling zones salinity is also believed to be important." They quote the benchmark sampling strategy to be one sample per 200 km square every 10 days and with an accuracy of 0.1 PSU. They also state that the tropical western Pacific and Indian Oceans and high latitudes should receive the highest priority.

Present Situation

The main instrument for in-situ collection is the thermosalinograph, TSG. At present a number of countries have national agencies which operate TSGs. Some of the data collected appear on the GTS in TRACKOB reports, the code form used for reporting data within 30 days of collection. France, Germany, Japan and the U.S.A. are the major contributors, although other countries, such as Australia and Canada, are either developing programs or indicating interest as well. The lead country has been France with the group in Noumea having the most experience and capabilities in collecting and processing surface salinity data.

A number of national agencies are known to collect data from TSGs but these data may or may not be sent to a national archive. In other cases, an international agency, e.g. ICES, archives these data for some of its member countries. There is no global cooperation in the management of these data, and so the data are collected and archived under different systems, with different data quality processing, different metadata and different data distribution facilities.

The Surface Salinity Project Goals

The objective is to build an effective international project to manage the surface salinity data that are now collected. This means working with data collectors and users to meet the benchmarks of spatial and temporal sampling and data accuracies set forth by OOPC. The project has the following goals.

- To build a comprehensive archive for surface salinity data. This encompasses data collected by any instrumentation at any time. It will contain sufficient metadata that users will have clear information about accuracies, instrumentation, sampling, etc.
- To add value to the archive by refining and standardizing existing quality assessment procedures carried out on data and documenting both what was done and the results.
- To provide data and information to users in a timely fashion. At any time after data collection, a user should be able to access the highest quality, and most recent data available. Users will be able to distinguish "levels" of quality in the archives. Users will be able to utilize the data and easily combine them with data from other sources.
- To work with data collectors to improve the data acquisition systems and to provide information to data collectors about the data they provide.
- To work with scientific organizations interested in surface salinity data to provide products to a broader community.

Liaisons

A surface salinity project cannot exist in isolation from other international structures and organizations. The project will need strong connections to the WMO since it is responsible for the operations of the GTS, and this is the main communications system presently for data exchange. The SOOPIP is the group that coordinates upper ocean sampling under J-COMM. Since many of the ships that are used in SOOP also potentially would have a TSG installed, it makes sense to work closely with SOOPIP.

The project does not need to manage all of the surface salinity data itself. There are other activities, where surface measurements are made and these are already well managed. Examples would be measurements made by surface drifters, by moored platforms, and by platforms (ships or profiling floats) measuring salinity profiles. In this case, the project can make the appropriate connections to the agencies managing the data to assemble the complete surface salinity to provide the required products.

Last updated: 16 Oct, 2000

Management of Surface Salinity Data

Justification and Guidance

The Ocean Observations Panel for Climate, OOPC, and its predecessors examined the usefulness of surface salinity data in the context of climate change detection. They state that "At high latitude, sea surface salinity is known to be critical for decadal and longer time scale variations associated with deep ocean over turning and the hydrological cycle. In the tropics, and in particular in the western Pacific, and Indonesian Seas, and in upwelling zones salinity is also believed to be important." They quote the benchmark sampling strategy to be one sample per 200 km square every 10 days and with an accuracy of 0.1 PSU. They also state that the tropical western Pacific and Indian Oceans and high latitudes should receive the highest priority.

CLIVAR planners have stated what they think are the attributes of a successful data and information management system. In the Proceedings of the International CLIVAR Conference held in Paris in 1998, one attendee states that the following points are likely to be prominent.

- Swift assembly and distribution of data, ready availability and free access.
- Data sets and products are comprehensive in terms of the variables
- Encourage exploration of historical and paleo data sets.
- Encourage processing methods that eliminate or minimize the production of spurious signals.
- Ensure that adequate and timely data are available for the initialization and validation of climate forecast systems.
- Maximize the utility of the sustained observing system for various process or intensive studies.
- Work with GOOS/GCOS, World Weather Watch and Global Atmosphere Watch in development and implementation of the strategy.

Finally, the latest draft data and information management plan under development for GOOS advises an end-to-end system for the various components contemplated. In particular, it suggests strategies that should be employed to develop the systems. This project should review the contents of the GOOS plan, extracting those ideas and suggestions that should be used to direct the development of a Surface Salinity project.

Present Situation

The main instrument for in-situ collection is the thermosalinograph, TSG. At present a number of countries have national agencies which operate TSGs, but the data may or may not be handed to a data centre for archiving. Some of the data collected appear on the GTS in TRACKOB reports, the code form used for reporting data within 30 days of collection. France, Germany, Japan and the U.S.A. are the major contributors, although other countries, such as Australia and Canada, are either developing programs or indicating interest as well. The lead country has been France with the group in Noumea having the most experience and capabilities in collecting and processing surface salinity data.

A number of other national agencies are known to collect data from TSGs but these data may or may not be sent to a national archive. In other cases, an international agency, e.g. ICES, archives these data for some of its member countries. There is no global cooperation in the management of these data, and so the data are collected and archived under different systems, with different data quality processing, different metadata and different data distribution facilities.

It is clear that at present surface salinity data are not well managed either internationally or in many cases at the national level. It is the object of this document to provide the framework to pull these disparate efforts together into a coherent global management scheme for surface salinity data. The objective is to provide the program that can support the programmes that make the measurements, to meet the scientific objectives, and manage the data to meet as many as possible of the desirable attributes stated earlier.

Canada and France have expressed interest in participating in a global data management strategy for surface salinity data. However, to be successful, it will be important to get the active participation of agencies in other countries.

The Goal of the Surface Salinity Project

The objective is to organize the surface salinity data that are now collected and to work with data collectors to improve what is presently collected to try to meet the benchmarks of spatial, temporal sampling and data accuracies set forth by OOPC. The resulting data management system should meet as many of the attributes of a successful system as stated by CLIVAR (as quoted above). More specifically, the project has the following goals.

- To build a comprehensive archive for surface salinity data. This encompasses data collected by any instrumentation at any time. It will contain sufficient metadata that users will have clear information about accuracies, instrumentation, sampling, etc.
- To add value to the archive by refining and standardizing existing quality assessment procedures carried out on data and documenting both what was done and the results.
- To provide data and information to users in a timely fashion. At any time after data collection, a user should be able to access the highest quality, and most recent data available. Users will be able to distinguish "levels" of quality in the archives. Users will be able to utilize the data and easily combine them with data from other sources.
- To work with data collectors to improve the data acquisition systems and to provide information to data collectors about the data they provide.
- To work with scientific organizations interested in surface salinity data to provide products to a broader community.

Administrative Considerations

Duplicate, or near duplicate data, is a long standing problem for both national and international programmes. This arises quite simply. When data are sent in real-time through the GTS, they often represent a lower resolution than the original. Sometimes, information about position, date and ship identifier is also incorrect. When the delayed mode data appear at the archive, some of the problems may have been corrected, calibrations applied to the data and other changes that make the data look sufficiently different from the real-time, that software alone cannot (sometimes not manual scrutiny either) recognize that the two submissions represent the same original observations.

It is also common for data centres to receive either the same data multiple times from an originator, or the same data from two different originators. These submissions may be many years apart, and the data may have gone through changes and "fixes" by the originator(s). Again, these near duplications may not be easily detected.

There is extensive international data exchange and so copies of the same data end up in many different archives. Each archive centre has their own procedures for managing data and although care is taken not to alter data, mistakes in format conversion, or not storing all of the metadata that accompanies the data can result in loss of information. A user getting data from two, or more, archives will see the same, or nearly the same, data and this leads to confusion and loss of confidence in the whole data archival/management process.

The ease of exchanging data over the Internet is compounding this problem. Now, users can go to archives or to data collector's sites, or other places and download data that has undergone changes that are often poorly documented. It is easy for a user to get many copies of the same data, and if they don't realize it or don't have a simple way to find these duplications, whatever results derived from the data may be suspect.

The Surface Salinity Project should take steps to solve this problem. There are two issues to be addressed. The first concerns unequivocal identification of variations of data that really come from the same source. The second is knowing the origins of data and being able to identify what handling the data have undergone.

The proposed solution to the first problem is to attach a unique tag to the data at source. This tag must accompany the data everywhere and never be altered. For the surface salinity program, we must decide what "unit of data" acquires the tag. Is it, for example, each observation, or can we group observations, such as by cruise, and attach a single tag to the group? The SOOPIP has asked for a proposal for such a tag, and a copy of that is attached as annex A.

The solution to the second problem is to record the agencies through which the data have passed. Just as for the unique tag, this information must always accompany the data. Every agency handling the data adds its own record to this history, and no agency may delete this. This history along with the unique tag will help a user identify unique data and also see where changes have been made. It can assist in the specification of what constitutes the "best copy" of data. The Surface Salinity Project needs to decide what information must be carried in this history. Experience with the history structure of the GTSPP format may provide some help.

Liaisons

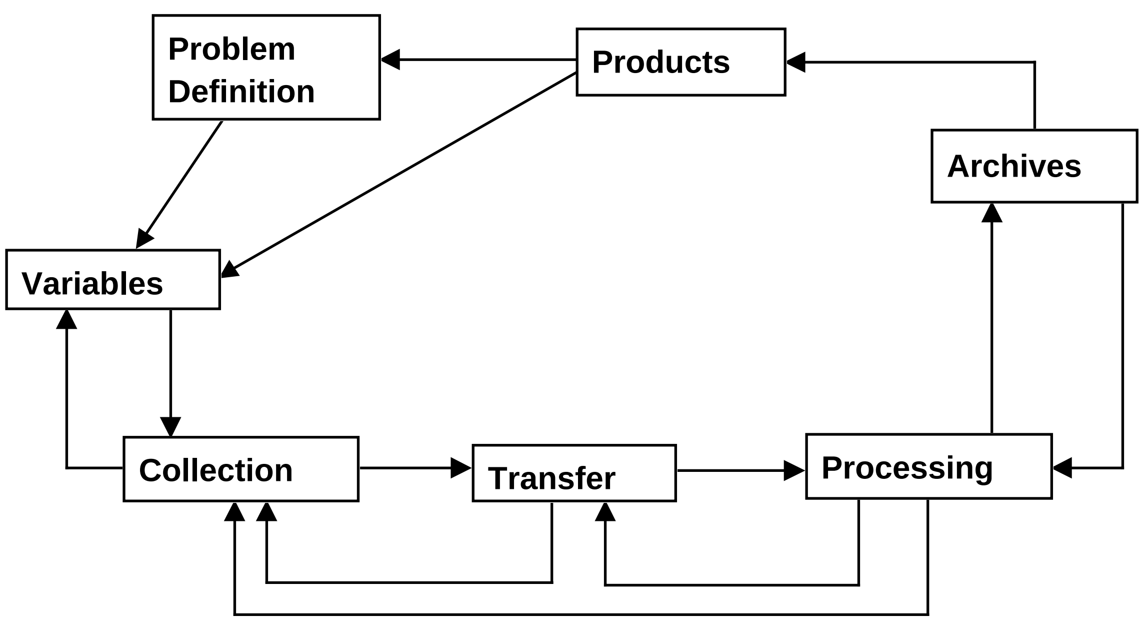
A surface salinity project cannot exist in isolation from other international structures and organizations. A surface salinity project will need strong connections to the WMO since it is responsible for the operations of the GTS, and this is the main communications system presently for data exchange. The SOOPIP is the group that coordinates upper ocean sampling under J-COMM. Since many of the ships that are used in SOOP also potentially would have a TSG installed, it makes sense to work closely with SOOPIP. It should be considered whether a surface salinity project should be a separate entity or whether it should be administered under a group such as J-COMM or SOOPIP.

The Surface Salinity Project does not need to archive all of the surface salinity data itself. There are other activities, where surface salinity measurements are made and these are already well managed. Examples would be measurements made by surface drifters, by moored platforms, and by platforms (ships or profiling floats) measuring salinity profiles. In this case, the Project can make the appropriate connections to the agencies managing the data to assemble the complete surface salinity to provide the required products. The section on Archives and Products later on deals with this in more detail.

Data Management Overview

The following diagram shows a schematic of the functions that bear on the management of scientific data. Each of these functions has a number of components. The arrows connecting the functions represent the transfers of information between functions. This document will provide details of the various functions as they apply to the management of surface salinity data.

It is important to keep in mind that other variables are also measured coincident with surface salinity. For example, surface meteorological variables are sampled and these data processed through their respective systems. In the design of a surface salinity system, it makes sense to remember that although surface salinity may require different processing than, for example, sea level pressure, in the end some users will want to re-assemble the complete set of coincident observations for their analyses. Whatever the system that is built, it should make this re-assembly process simple.



A Surface Salinity System

1. Problem Definition

This component encompasses the definition of the scientific problem to be addressed, and the national and international program development that may be required. For this, the Project should use the requirements set forth by GOOS/OOPC.

2. Variables

This component is driven by the scientific problem to be addressed. It is the scientific problem which dictates what are the variables, accuracies, resolution frequency and important metadata to collect. As stated by OOPC, surface salinity must be measured every 10 days in a 200 x 200 km area to an accuracy of 0.1 PSU.

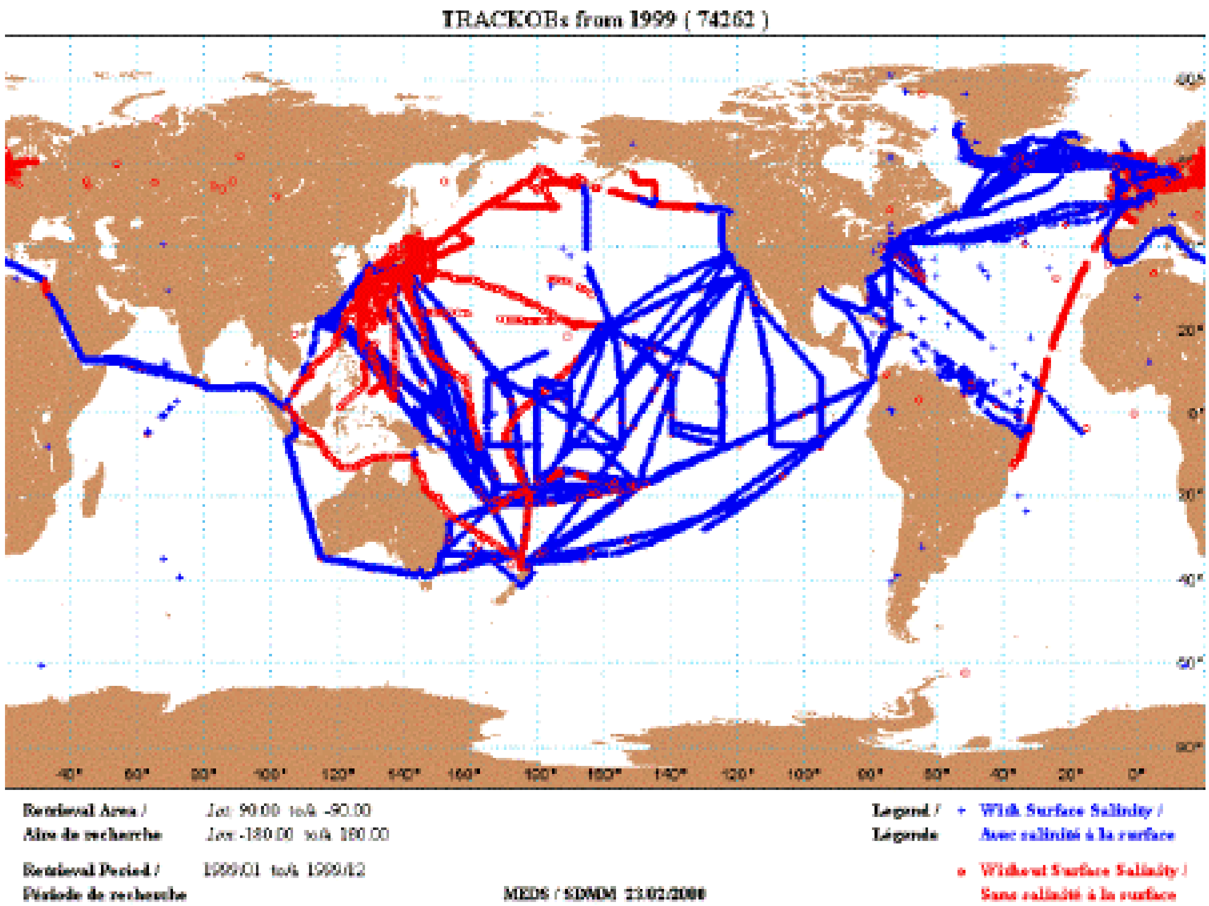
3. Collection

This function describes the data collection process including the mechanisms and instrumentation, the agencies involved and the timeliness constraints.

The primary instrumentation that can provide in-situ surface salinity data is the thermosalinograph. This instrument is capable of measuring temperature to accuracies of 0.01 degrees and salinities to 0.02 PSU. These are clearly better than the accuracies giving in the scientific justification. However, in order to achieve this level of accuracy, care must be taken in both the installation and maintenance of the TSG. A manual has been prepared by the French in Noumea that gives detailed instructions on the installation and maintenance of TSGs so that the highest possible quality of data are returned. This is available through the SOOP home page (http://www.ifremer.fr/ird/soopip/publications.html).

What is done now:

The only data collection that can be assembled readily to see what is the state of global surface salinity measurements is that from the GTS. The figure below shows the total number of TRACKOB reports distributed in 1999. Each observation is preented in red or blue, with those in blue having surface salinity observations and those in red without. In total there were about 75,000 observations.



It is clear that just as for other variables, sampling is highly variable with certain areas being much better sampled than others. The requirements for spatial and temporal coverage specified by OOPC is not met.

If all of the data collected by national agencies could be easily assembled, it is likely that the above figure would have some of the blank areas filled in. However, lacking any well-organized archiving system, at this point we cannot know.

What needs to be done:

a. The project needs to identify the agencies that collect surface salinity data, but that do little more to archive the data or have no routine mechanism for providing data or products to other users. These agencies need to be contacted to arrange the data to be sent to a surface salinity project processing facility.

b. The project needs to ensure that the accuracy of these historical data is at least as good as the stated requirements of the OOPC. Information about the instrumentation used and calibrations carried out should also be gathered and sent to a processing facility.

c. The project needs to encourage additional agencies to start to collect data using the protocols outlined in the TSG manual. This can be carried out simultaneously with contacting present data collectors. A demonstrably successful program will also help encourage others.

4. Transfer

The transfer component, in fact, spans many of the other functions in that this is the way data are moved in the international system. This function is concerned with data formats and file contents, the transfer mechanisms themselves and, monitoring the speed and reliability of data transfers.

What is done now

For real-time transfers of data, there is the existing TRACKOB code form used on the GTS. This was developed a number of years ago, but its use has only grown in the past few years. The TRACKOB code form provides for only a minimum of metadata to accompany the measurements. TRACKOB provides simply the averaging period used when reporting data, as well as the position and time information.

It is not known what are the delayed mode data formats. It is in this area that ICES is currently writing a document that describes what information should accompany surface observations. It seems likely that there are a number of different formats that have been developed by each agency and/or country for their own purposes. There is no internationally agreed standard.

What needs to be done

a. The project needs to set standards for the temporal and spatial scales of data that will be archived. TSGs sample surface values at frequent intervals (1 second) but these are averaged for real-time reporting. The appropriate temporal standard depends on the problem being addressed and the spatial scales on the ocean region being sampled. Guidance from scientific bodies is needed to resolve this issue.

b. The project needs to consider an exchange format for data coming from the collectors to the assembly centre(s). The project should examine the work sponsored by ICES in this regard, and whatever formats are currently being used. It is important to work with the data providers rather than impose a format on them.

c. The project needs to decide how data will be made available to participants in the project. For example, if data are to physically move from one site to another, a suitable exchange format will need to be defined. If data can be accessed and manipulated remotely by project participants, an exchange format may not be required.

d. One of the outputs of the project will be the data themselves. The project needs to define how the data will be provided to users, keeping in mind the attributes of a successful data management system stated by CLIVAR. A corollary to this requires the project to consider what other data complement surface salinity measurements and what are the formats used to provide these complementary data to users.

e. The project will need to address the content and format issue for international exchange, with particular attention to carrying the appropriate metadata concerning the measurements and processing. Based on experience from GTSPP, the following attributes should be included in a format.

- position, time, ship identifier
- line number assignment (compatible with SOOP lines) when appropriate.
- averaging information for surface observations.
- identification of QC tests applied and by whom, and tests failed.

- identification of quality of the observed values.
- time stamp the processes used in the system.
- associated observations, such as meteorological observations.
- associated metadata not initially considered.
- information about whether calibrations have been applied.

5. Processing

This function includes all of the data processing components needed to handle data. This includes descriptions of how duplicates will be identified, data quality assessment procedures, the timeliness of these functions as well as monitoring reports to provide back to data collectors to improve the collection process.

What is done now

MEDS acquires real-time data from the GTS, passes the data through simple range checking software and archives everything including time received and GTS bulletin information. These data are available to whomever wishes them, but on request.

Noumea carries out calibrations, quality control and archiving of data collected by their vessels. The archive is an Oracle database. They also have received some data from other countries and they carry out quality control and archiving of these data. They have no routine distribution mechanism although data can be provided on request.

France has an archive on PC. There is quality control carried out, and the archive includes data from other countries. There is a facility for data distribution on request.

There are known to be some national processing facilities, but the information about these is lacking.

What needs to be done

a. The project needs a manual that describes the quality control processing carried out on the data.
b. The project needs to decide on how data quality will be recorded. Will it use the IGOSS/GTSPP scheme or some other?
c. The project needs to decide what information will be needed to monitor data collection, receipt, and transfer. GTSPP experience suggests that all of these processes be time stamped.
d. The project needs to decide if changes in data are permitted, under what circumstances, and if original values will be retained. GTSPP permits changes at the discretion of participants, but original values are stored in a history structure. GTSPP records both changes to values and changes in flags. It records reasons for these changes and who made the change and when.
e. The project needs to decide what, if any, place there is for real-time data. GTSPP experience is that even after many years, the real-time data still form a significant fraction (of order 20%) of the archive. We do not know if this is true for TSG data, but it is prudent to assume that the same will be true.

f. Assuming real-time data are to be archived, a scheme for duplicate resolution needs to be designed. There is no reason to believe that this will be any easier than GTSPP's experience. The project should consider the adoption of a unique data tag as a way that the identification of duplicates can be eased. Alternatively, duplications between real-time and delayed mode could be done on a "cruise" basis rather than station by station. That is, instead of comparing each station's information, as delayed mode cruises come to the archive, the ship identification, approximate area and approximate time be used. When a match is found, all real-time data from the cruise is noted as being a duplicate.

g. MEDS currently acquires data from a single GTS stream. GTSPP experience shows that by acquiring data from other streams, a further 5% can be added to archives. Given delays between receipt of real-time and delayed mode data, the project needs to recommend whether or not multiple data streams should be pursued for real-time inputs.

h. The project should consider a way to tag data as being the best copy available. GTSPP does this through the history structure, but this may be too clumsy. It may be that a time stamp of last changes and a unique data tag needs to be implemented. If a unique tag is adopted, all participants must preserve the tag unchanged.

i. The project needs to clearly state the processing roles and responsibilities of participants.

j. The project must establish what monitoring reports are needed. Each proposed report must have a specific purpose. The following suggestions are made.

- Track timeliness of receipt of real-time data. This information will be used to identify where gains might be made in increasing speed of access to data. They will periodically use this information to try to encourage data providers for faster provision of data.
- Monitor the quality of TRACKOB reports and provide a monthly reporting to operators.
- Monitor the completeness of data derived from different sources on the GTS and will provide regular reports to its partners and other interested groups to try to improve identified deficiencies.
- Monitor the quality of data processed and will contact data providers of problems found.

6. Archives

This component deals with the nature of the archives, security of holdings, data provision and historical data acquisition.

What is done now

ICES appears to have a substantial amount of surface salinity data in its "underway" data files going back to the 1900s. This covers the area of the North Atlantic. There are maps of the stations on line and in viewing them, you can infer from the ship tracks that more data exist in areas at least in the South Atlantic as well. Presently it is not known what is the number of surface salinity measurements nor details of their accuracy.

There is a surface salinity database operated in Brest. This operation is not well funded and it is not known what is the content of the archives.

MEDS operates a global archive for TRACKOB data starting in 1989. They carry out very simple data quality checks only.

It is believed that the most sophisticated management of TSG data presently occurs at Noumea. They equip their ships with TSGs, collect the data, calibrate the instruments and apply quality control to ensure the data are good. Their archive contains mostly data from their own ships, but they do have some data from other countries.

NOAA places some TSG data on an ftp server in Miami.

There is thought to be a substantial volume of TSG data residing in oceanographic institutes and perhaps other organizations around the world. The quantity and quality of these holdings are not known.

Surface salinity data are collected by a variety of instruments and techniques. Although TSGs can acquire a considerable volume of data, there are existing programmes that also have surface salinity measurements. The UOT/GTSPP/SOOP programme deals in temperature and salinity profiles in an end-to-end system. Some surface drifters and moored platforms report salinity both at the surface and in profiles. These data come to the RNODC for drifting buoys and also make their way into the GTSPP archives. Profiling floats are expected to make substantial contributions to salinity profile measurements. Tide gauges make surface salinity measurements as well. The Surface Salinity Project will need to consider all of these sources, look at how the data are managed in comparison to requirements for data and product delivery and decide what sources will be archived within the Project and what sources will be used in product generation.

What needs to be done

a. The project needs to assess the state of existing archives, their contents, data quality, accuracy, and metadata to determine if the contents meet the OOPC constraints. It needs to know which are the centres and institutions that have expertise and a desire to contribute. Possible contacts should be all of the known contributors of TRACKOB data (France, Germany, Japan, U.S.A.). Other contacts are countries/agencies with interest, or making data collections though not reported in real-time (Australia, Canada, ICES, WDCs, SEAS). Representation should be made to IODE and JCOMM for member support and contribution to the project.

b. The project needs to organize the global archiving efforts into a more coherent system. There is more than one model that can be applied and these should be explored. At the same time, the Project should assess the need for a data rescue project to try to recover data. Perhaps liaisons can be made with the GODAR Project.

c. The project needs to decide on how data quality will be indicated in the archives. It could adopt the model of GTSPP with a list of tests applied, and failed but with each observed value having only the indication of the quality. Alternatively, it could take the model of the World Ocean Database where each observation has flags saying whether it passed or failed each test. There may be other models as well that should be considered.

d. The project needs to decide if other observations besides those returned from the TSG will be archived. This includes surface salinity data collected at lower accuracies than the OOPC criteria. Also, it is common for meteorological variables to also be collected, and it is also desirable to have all of these associated. This project could choose to try to keep all of these variables in the same archive (when they are submitted with the TSG data) or to adopt an exchange format that permits easy integration with meteorological data held in other archives. Such a format may be BUFR, or netCDF. If this project chooses not to include other measurements, what will be done if/when such associated data accompany the TSG observations?

e. The project will need a strategy for handling metadata not initially anticipated. This means the archive and exchange format must be easily expandable. GTSPP experience suggests a format that uses names in tables to indicate the content of a file. An alternate strategy is to use a completely self-describing format such as BUFR. The new BUFR Master Table 10 would have most or all of the entries needed for this.

f. The project needs to make liaisons with agencies that have a serious interest in the surface salinity data. It is known that space agencies in Europe and the U.S.A. are interested in building space-based salinity sensors. These groups need to be contacted to make the project known and to request support and guidance on data and information to be collected in the archives (perhaps format issues as well).

g. The project needs representation from science centres that actively use the data. These centres may be a formal part of the data review, value adding process, or regular users who provide feedback to the project on an on-going basis.

h. The project needs to decide the nature of the archive. The GTSPP model is a continuously managed archive where data enter as soon as possible. At any time, a user may request data and the highest quality, highest resolution will be provided. In practice this means data provided are at various levels of quality and resolution. An alternate mode would have the archive keep the data at different levels of quality and resolution separately and permit users to choose for themselves what they wish.

7. Products

This component encompasses both the provision of data to users and scientific analysis products. It also addresses project documentation, such as project plans and brochures, to increase the visibility of the project.

What is done now

MEDS places on its web site monthly maps of where TRACKOB data received were collected. ICES has maps of underway data collected since the early 1900s. There may also be other products at other sites.

What needs to be done

a. The project needs to decide how products will be made available. Candidate techniques are through web sites, CDs, or other distribution media. It is suggested that 2-3 years after the formal start of the project, a CD or other "hard copy" of the project be issued. This will assist in project visibility and contributor recognition.

b. The project needs to consider what is the data distribution format to users.

c. The project needs to define the suite of data flow products to help contributors and users know what has been received. It is suggested that the following be adopted.

• Maps of the location and times of real-time and delayed mode data in the archive. The maps will be produced by the appropriate archive centres and on a schedule that make sense given the data flow.

• On-line catalogue capabilities. This may take the form of files prepared from the archives or on-line search facilities. It is recommended that this product be available on the Internet. This product is designed to permit a user to ascertain if data of interest exist in the archives.

d. The project needs to define the suite of scientific products and how these will be assembled. Where surface salinity data are not archived by the Project, appropriate links to the archive centres will need to be made so that data and derived products that include all surface salinity data can be generated. It is suggested that the following products be adopted.

• Scientific analyses such as surface salinity fields on suitable time and space scales. It is expected that the scientific partners on the project will want to carry this out. Proper reference to the project should accompany any such products.

• Any computed fields shown in graphical products may also be suitable for provision as digital files.

e. The project needs to define the suite of "advertising" products and documentation needed. It is suggested that the following be adopted.

• A formal project plan be written. This will be used at the start of the project to help attract partners and to better define the scope of the project. The project plan should define the roles and responsibilities of partners.

• A brochure or similar one page document be designed. The object is to have a visually pleasing, eye-catching device to advertise the project and encourage participation. This should be designed as early in the program as possible and receive wide distribution.

• Meeting reports and documentation should be maintained in electronic form. Participants to meetings will contribute all formal documentation in electronic form, and a partner in the project must be designated to carry out the editorial and library functions for these.

• Any manuals produced by the project, such as detailed descriptions of processing, should be maintained in electronic form. Distribution can be by web site or through international publishing such as done by IOC.

Annex A

A SEAS Station Tag

New SEAS software is being designed that will have a major change in how data are collected by this program. Presently, the full resolution profile is collected on a shipboard PC. The data are stored on disk, and software constructs a BATHY or TESAC report that is sent ashore for distribution on the GTS. Under the new plan, the full resolution profile will be sent ashore and then the building of the appropriate GTS report will take place. The report will be sent out onto the GTS, but the high-resolution data will go directly to the archives. This opens up an opportunity to better manage duplications in data arriving at data centres.

Multiple copies of a single profile are known to occur in large archives. It is easy to understand how this can happen, even with the efforts that are taken to identify duplications. Using the example of the SEAS program, data first reach an archive as a GTS report. Then, after the ship returns to a port, the high-resolution data are retrieved from the computer on the ship and sent to the archives. In the mean time, the report will have circulated on the GTS, may have been picked up by a number of centres and quality control applied. Sometimes this quality control will have made modifications to the report including position, time or observed values.

The high resolution, delayed mode data will also pass through quality control. It is sometimes noted that there were clock problems on the computer recording the data so a correction to observed time may be made. Position information may also have been in error and so these may be corrected.

Over time, both the GTS report and high resolution profiles will be exchanged with other data centres, and they, too, may well put the profiles through quality control procedures. Then the data may return to the original centre. In all of these steps, changes may be made to the profile which makes the identification of duplicates difficult to do.

The Upper Ocean Network Review recognized that such problems exist and suggested attaching a unique tag to every station collected. As the station is handled, no matter what else is done, the tag is carried and never changed. Under this scheme the identification of duplicates is simply a matter of looking at the tags on stations. In order for this to help identify the GTS report and the high resolution profile from which it came, the tag must be attached to all forms of the profile, both high and low resolution.

The proposal of the SEAS program to send the high resolution data ashore opens an opportunity for tagging the GTS report as well so that when these reports come to the archive where the high resolution profile resides, the match can be made to the original high resolution profile. There is also the added bonus that if a high-resolution profile turns out to be bad, but a "cleaned" version was sent to the GTS, the report based on the erroneous data can be identified and removed from the archives.

If such a tag is built, it should be attached as soon as possible after the data have been collected. The Argo program is considering attaching such tags to its data and has suggested the tag be created by the concatenation of a 2 character country code, a 2 character PI code, a 4 character float ID and a 4 character profile number.

The tag need not be meaningful in the way that has been proposed by Argo. Its primary requirement is that it be unique. For SOOP, it is suggested that the tag be generated as a concatenation of the Lloyds number of the hull, a 4 character year, and a counter of the number of minutes since the start of the year (excluding the Lloyds number this is a 10 digit number). All of this is completely programmable once the Lloyds number is set. It also has the attraction that all historical data can be retroactively tagged (if not by using a Lloyds number, then by some other "cruise number" identifier according to commonly available cruise identification schemes). An alternative to the Lloyds number could be the ship call sign.

This tagging could be applied by the new SEAS software as follows. When a profile is collected, the unique tag is generated and linked to the data file. When the data are sent ashore, a GTS report is created and sent to the GTS. At the same time, the full resolution file is sent to NODC and MEDS. MEDS uses the file to identify the incoming report from the GTS to attach the tag and store this in the GTSPP archives with the reports. These data then go back to NODC, in the file that contains other BATHYs and TESACs. NODC uses the tag to compare to the high-resolution file to immediately flag the BATHYs and TESACs as duplicates. MEDS uses those profiles that lie in its area of interest to complement its archives directly. It is important that all data centres receiving these data preserve the tag unchanged.

It can be argued that since the high-resolution data appear at the GTSPP CMD simultaneous to the GTS report, there is no need to attach the tag to the BATHY or TESAC by MEDS. However, MEDS will be handling the real-time data stream and these data are distributed to users directly. If users build archives as well, it is important that the tags are attached. We wish to be sure that these reports at MEDS (and whoever receives data from MEDS) are unequivocally linked to their original high-resolution data. The only way to do this, and to properly monitor the exchange of data on the GTS is for MEDS to attach the tag to the real-time messages.