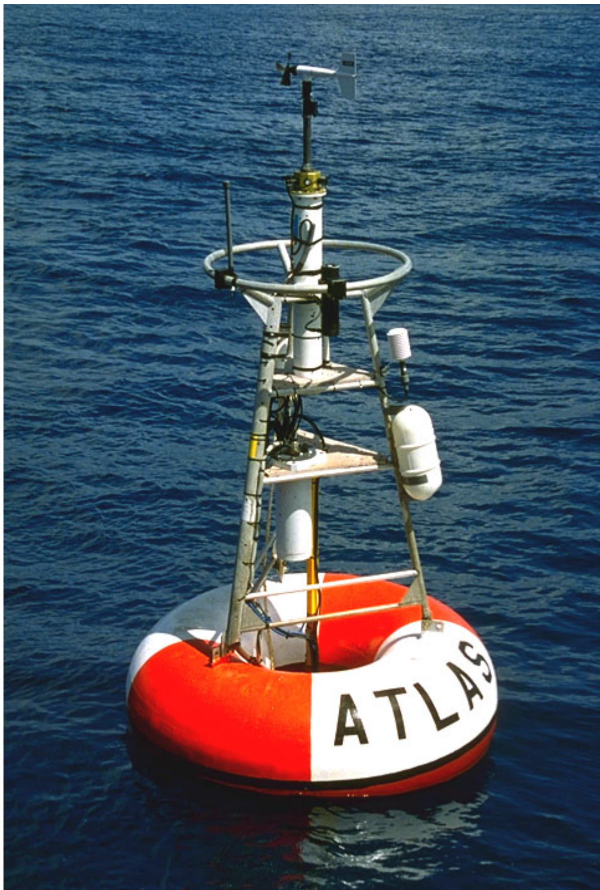
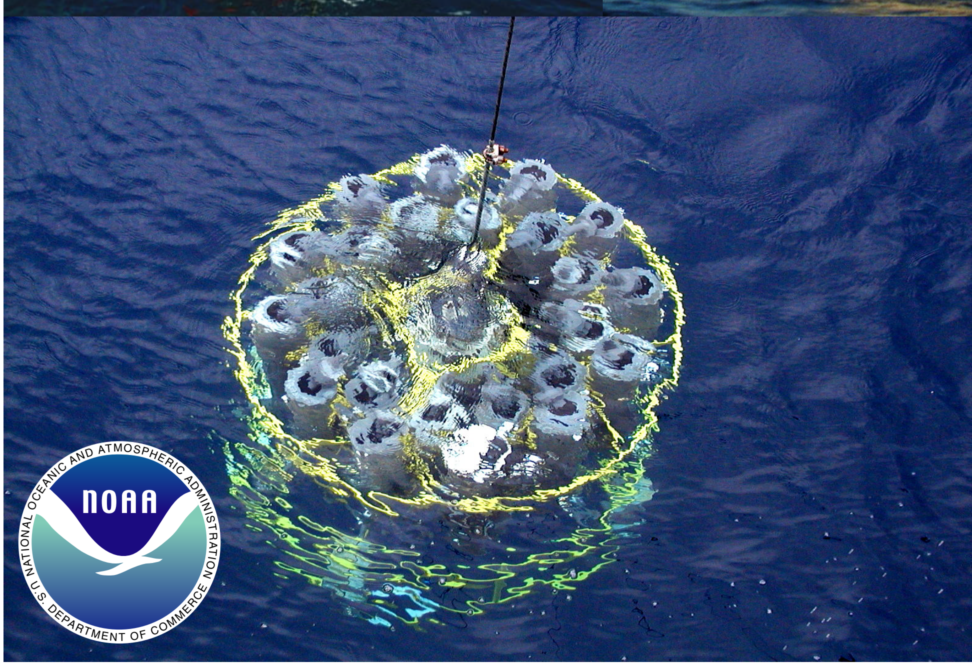
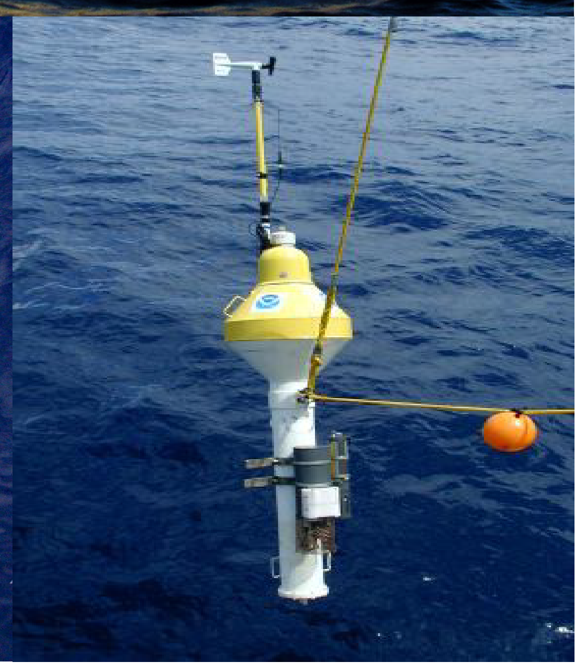
A Plan for Data Management of In Situ Large-Scale Oceanic Carbon Observations

Report of the Data Management Workshop
Pacific Marine Environmental Laboratory
Seattle, Washington
25-26 October 2001

Edited by Richard A. Feely
and Christopher L. Sabine







NOAA OAR Special Report

A Plan for Data Management of In Situ Large-Scale Oceanic Carbon Observations

Richard A. Feely[1] and Christopher L. Sabine[2] (editors)

Contributing Authors:
Thomas Boden, Robin Brown, John Bullister, Francisco Chavez, Margarita Conkright, Andrew Dickson, Lisa Dilling, Scott Doney, Marjorie Friedrichs, David Glover, Nicolas Gruber, Steve Hankin, Dennis Hansell, Matthew Harrison, Greg Johnson, Robert Key, Joanie Kleypas, Alex Kozyr, Charles McClain, Frank Millero, Calvin Mordy, Raymond Najjar, Tsung-Hung Peng, Detlef Stammer, James Swift, Robbie Toggweiler, and Rik Wanninkhof

[1]Pacific Marine Environmental Laboratory
7600 Sand Point Way NE
Seattle, WA 98115

[2]Joint Institute for the Study of the Atmosphere and Ocean (JISAO)
University of Washington
Box 351640
Seattle, WA 98195

February 2002

Contribution 2445 from NOAA/Pacific Marine Environmental Laboratory

NOTICE

Contribution No. 2445 from NOAA/Pacific Marine Environmental Laboratory

For sale by the National Technical Information Service, 5285 Port Royal Road
Springfield, VA 22161

Contents

A Plan for Data Management of In Situ Large-Scale Oceanic Carbon Observations

Richard A. Feely[1] and Christopher L. Sabine[2] (editors)

1. Executive Summary

The primary goal of the Data Management Workshop was to formulate a plan for data policy, institutional arrangements, and mechanisms for dealing with new and historical oceanic CO_2 data sets. The overall objective is to provide the scientific community with easy access to high-quality historical and near real-time CO_2 and related data sets from the large-scale ocean carbon observing system proposed in the LSCOP report. The workshop consisted of 31 scientists, data managers, and program administrators from federal agencies, universities, and private research institutions. Two major data types were considered in the workshop: (1) discrete data from the large-scale Repeat Hydrography Program and time series stations; and (2) high-frequency data streams from shipboard underway measurements and autonomous measurements from fixed moorings and drifting buoys. The workshop participants advocated the formation of a CO_2 Science Team involving CO_2 scientists and data managers to collaborate in the development of standardized procedures for data reporting and quality assessment. They also emphasized the need for rapid data distribution to the scientific community via the Internet and CD-ROMs. The data access system must include a "Live Access Server" for on-line browsing that will include graphics tools to provide plots and visualizations of the data.

The following outline lists the data management requirements for producing QCed versions of these data, and for assuring their easy access to the scientific community. Rough estimates of the associated costs are included.

1. Discrete Data (including measurements made on discrete samples at time-series stations)
    (a) Processing of shipboard data
        i. CO_2 Science team will coordinate CO_2 measurement program, establish measurement protocols and data formats, and run intercomparison exercises as necessary to ensure highest-quality measurements (\$30–40K/yr).
        ii. One data management person on each survey cruise will be responsible for merging and initial QC of all shipboard data.
    (b) Production of locally QCed data

[1]NOAA/Pacific Marine Environmental Laboratory, 7600 Sand Point Way NE, Seattle, WA 98115

[2]Joint Institute for the Study of the Atmosphere and Ocean (JISAO), University of Washington, Box 351640, Seattle, WA 98195

  i. PI will be responsible for post-cruise processing of shipboard data.
  ii. CO_2 Science team will assist PIs by developing techniques and tools for post processing and will oversee and coordinate public access to locally QCed data ($25K/yr).

(c) Production of regionally QCed data

  i. The National Data Center will be primarily responsible for assessing quality of new data sets with respect to other high-quality data sets in that region. They will also serve research quality data to the public ($50K/yr).
  ii. CO_2 Science team will assist the National Data Center in quality assessment by providing scientific oversight.
  iii. The World Oceanographic Data Center will archive all data products and assist in data quality assessment ($50K/yr).
  iv. A "Live Access Server" will be developed to distribute the data via the Internet ($50K/yr).

2. Semi-autonomous and autonomous underway and buoy data

(a) Processing of data

  i. CO_2 Science team will coordinate CO_2 measurement program, establish measurement protocols and data formats and run intercomparison exercises as necessary to ensure high-quality measurements ($30-40K/yr).
  ii. CO_2 Science team will assist PIs by developing techniques and tools for post-measurement data-quality assessment, but PI will be responsible for implementing these procedures as part of their measurement responsibilities.
  iii. Data will be transferred to National Data Center within 1 month of reaching the PI.

(b) Distribution of data and production of integrated datasets

  i. Data Management Group will be responsible for checking that the data received from the PI are correctly formatted.
  ii. The National Data Center will make data available publicly as soon as practical (typically within 1 week of receipt) ($50K/yr).
  iii. CO_2 Science team will assist Data Management Group by providing scientific oversight and indicating how best to integrate such datasets.
  iv. The World Oceanographic Data Center will ultimately archive and distribute all national and international data products ($50K/yr).
  v. A "Live Access Server" will be developed to distribute the data via the Internet ($50K/yr).

2. Introduction

The U.S. Carbon Cycle Science Plan (Sarmiento and Wofsy, *A U.S. Carbon Cycle Science Plan.* USGCRP, Washington, D.C., 69 pp., 1999), developed under the auspices of the U.S. Global Change Research Program (USGCRP), represents the beginning of intensive planning for the next decade of global carbon cycle research in the United States. This document provides an evaluation of the then-current state of knowledge of carbon processes in the atmosphere and oceans including its anthropogenic aspects, and suggests a course of coordinated federal action for advancing carbon cycle science.

The Plan emphasizes the need for coordinated and complementary programs of basic and applied research from the U.S. Federal agencies with interests and responsibilities in global carbon cycle science. As a direct outgrowth of this original Plan, the U.S. Carbon Cycle Science Program (CCSP) was established to foster community interest and planning in the area of carbon cycle research. Under the auspices of the CCSP, a recent NOAA report by Bender *et al.* (2001)[3] provides a U.S. implementation plan for large-scale observations of CO_2 and ancillary properties in the atmosphere and oceans in direct support of upcoming research programs, such as the Climate Variability Program (CLIVAR) and the CCSP.

In support of this effort the NOAA Office of Global Programs commissioned a workshop at the Pacific Marine Environmental Laboratory of NOAA on the requirements for data management, synthesis, and assimilation of the potentially very large numbers of oceanic CO_2 and related observational data that will become a part of the ocean component of the CCSP. The primary goal of the Data Management Workshop was to formulate a plan for data policy, institutional arrangements, and mechanisms for dealing with new and historical oceanic CO_2 data sets. The ultimate objective is to provide the oceanographic community with easy access, via the internet and CD-ROMs, to high-quality near real-time CO_2 and related physical, chemical, and biological data sets from the large-scale ocean carbon observing system proposed in the Large Scale CO_2 Observing Plan (LSCOP) report. It will also be desirable to provide equivalent access to high-quality historical (pre-2001) data that can be used for comparison purposes.

The workshop consisted of 31 scientists, data managers, and program administrators from federal agencies, universities, and private research institutions. The workshop was divided into nine plenary talks and nine discussion sessions, each led by a discussion leader and rapporteur (see attached agenda and participant list). Each discussion leader was responsible for a written report from which this summary was assembled. Two major data types were considered in the workshop: (1) discrete data such as will come from the large-scale CO_2/CLIVAR Repeat Hydrography Program (see `http://www.aoml.noaa.gov/ocd/repeathydro/`); and (2) data streams from semi-autonomous shipboard measurements and from auton-

[3] Bender, M., S. Doney, R. Feely, I. Fung, N. Gruber, D.E. Harrison, R. Keeling, J.R. Moore, J. Sarmiento, E. Sarachik, K.B. Stephens, T. Takahashi, P. Tans, and R. Wanninkhof, A large-scale CO_2 observing plan: Oceans and Atmosphere (LSCOP), NOAA Special Report, in press.

omous measurement systems on fixed moorings and drifting buoys. The following summarizes the workshop recommendations for these activities.

3. The CO_2 Science Team

The idea for a CO_2 Science Team comes from the highly successful CO_2 science team (supported by DOE and NOAA) that participated in the WOCE Hydrographic Program global survey of CO_2. This science team will be responsible for standardizing the techniques used for both the discrete and autonomous measurement programs. The workshop participants felt that it was important for the science team to be formed as soon as possible and to meet regularly throughout the program. One of the first tasks of this science team would be to generate a document that clearly outlines

- the standard set of analytical methods that will be used,
- standardized data formats,
- minimum metadata reporting requirements,
- procedures for quality control and assessment,
- steps for post processing,
- standardized guidelines for flagging data.

This document will provide the guidelines for U.S. participants in the program. This document should be published and distributed widely to the international community. The hope is that this document will encourage international partners that are also making carbon measurements on repeat sections, Volunteer Observing Ships (VOS), moorings, and drifting buoys to adopt similar protocols. If they do, it will make future merging and quality assessment of such non-U.S. data sets much easier. International partners that are interested in working with the U.S. should also be invited to participate in and contribute to the CO_2 Science Team. Evaluation of developing technologies and outreach programs, such as methods comparison studies, should also be a part of the CO_2 Science Team's purview. These activities will help ensure that the program stays state-of-the-art and will encourage international participation and collaboration.

The workshop participants noted that the science team approach could work well for other elements of the repeat section and autonomous measurements programs and that the CO_2 Science Team should make a point to include discussion of the hydrographic, nutrient, and tracer components in their meetings. The group generally did not feel that one large multi-component meeting would be the most efficient, but noted the need for clear coordination and communication between the science teams responsible for the various components.

These meetings should be held at regular intervals to encourage coordination between the CO_2 community and other U.S. and international communities. Such periodic meetings would also allow the CO_2 scientists to adapt the program strategies and approaches over time to take advantage of developing technology, or new information gained from the ongoing program. There should be enough flexibility built into the program to allow for decisions as to which section lines will be occupied by which group, or to adapt to constantly changing VOS ship schedules and unforeseen opportunities for participation in non-U.S. cruises.

The CO_2 Science Team would also facilitate the processing of shipboard data to the locally QCed level. It was pointed out that at-sea data processing is the most effective but can be expensive. At a minimum, there should be a data management person on board—for the entire shipboard analysis program, not just CO_2. The CO_2 Science Team will assist the PI by providing data and metadata requirements, standards, etc., and by helping the PI to oversee the process. Thus, the PIs doing individual cruises must include the cost of such data processing in their proposals to conduct the at-sea work.

It was suggested that some level of quality assessment could be performed most efficiently at a central facility, perhaps in collaboration with the data management group(s). The workshop participants also agreed that the data should be released to the public as quickly as possible after a cruise. One problem is that the carbon data cannot be fully processed until the hydrographic and nutrient data are available. It was noted that the routine hydrographic and nutrient data should be available in near final form within 5 weeks of cruise completion. Given that time frame, the group agreed that the discrete carbon data could be made available within 6 months of cruise completion and autonomous data could be made available within about 1 month of data collection. The data should be served on the World Wide Web with the necessary caveats about the data quality at that point. PIs will be encouraged to submit data reports that can be cited by the data users.

4. Discrete Water Column Measurements

The goal of the water column data management discussion was to develop a structure that will facilitate the timely processing and public release of high-quality data associated with the repeat survey program outlined in the LSCOP report. The proposed occupation of 15 lines over a 10-year repeating cycle will be a lower level of effort than the 1990's global CO_2 survey, but the long time frame and ambitious objectives of the program make the need for a good data management plan important from the outset. The plan must be able to maintain the continuity of the program while at the same time keeping the flexibility to incorporate improvements in data collection and analysis techniques and be able to respond to surprise findings. The workshop participants emphasized the importance of keeping the PIs involved in the processing of the data even after the cruise, working in concert with the data management group. They thus proposed a two-stage data management structure. This structure is designed to accommodate three basic data types:

- Shipboard Data—Complete data set resulting from a single cruise.

These data would include the CO_2 measurements and other appropriate ancillary information available at the completion of the cruise.

- Locally QCed Data—Complete data set that has undergone post cruise processing and has been examined for consistency with other parameters collected on that cruise. The PI will have the primary responsibility for this work in collaboration with the CO_2 Science Team.
- Regionally QCed Data—Complete data set, for which the "locally QCed" data has been further examined for consistency with other data from that region and is pronounced suitable for melding with other "regionally QCed" data sets (perhaps after some identified adjustment has been applied).

The workshop participants advocated that the CO_2 Science Team would facilitate the production of high-quality shipboard and locally QCed data as a first stage of the proposed data-processing scheme. The regional level of data assessment will then be handled primarily through the data management center(s) in collaboration with the CO_2 Science Team.

By dividing the data processing into two distinct processes to be handled by two communicating groups, we can standardize the approaches and reduce duplicated efforts. The group felt that higher level data products, such as the production of gridded carbon fields from the regionally QCed data, were still research issues at this point and should not yet be put into an "operational" structure.

5. Autonomous and Semi-Autonomous Measurements

The LSCOP report envisions a variety of autonomous and semi-autonomous measurement systems, particularly for the determination of the $p(CO_2)$ of the surface ocean and of a variety of ancillary properties. These devices range from shipboard systems (e.g., $p(CO_2)$ systems on VOS) that measure surface ocean properties while a ship is underway, moorings and drifting buoys along with their associated measurement systems. The proposed increase in the number of such systems over the next decade will greatly improve our surface seawater $p(CO_2)$ database.

The focus of discussions at the Workshop was thus on how to implement a data system that would provide rapid availability of calibrated data of a known quality, even as the number of measurement systems supplying such data increases with time. The workshop participants felt that, in a well-functioning system, it should be possible to have the results made available to other scientists within 1 month of the date when the data "reaches dry land." For autonomous shipboard systems, this would be the date when the full data set is recovered from a shipboard system, e.g., at the end of a round trip; for a buoy-mounted system it would be the date when the data were telemetered to the operator from the instrument.

Achievement of this rapid turnaround will require investments both in the data acquisition systems and in the data management areas. Investments in instrument design and day-to-day operating procedures should be aimed at minimizing the need for further "data processing"; i.e., ensuring that the resulting measurements of $p(CO_2)$ and of any ancillary parameters are well calibrated and of known quality. Initial investments in data management should focus on developing increased automation of the data-quality assessment procedures used to evaluate such measurements. As the program continues, it is envisaged that the "unit cost" of these measurements will decrease with time, as a consequence of the increased reliability of the instrumentation and the automation of the data handling.

To make this practical, data must be stored in a standardized format that makes all the measurements (including those of ancillary parameters), together with their associated metadata, accessible electronically. Once the data and metadata are in this standard form, they can be passed as a package to a Data Management Group who will have the responsibility for verifying the data format and for making the data available to the community as quickly as is practical.

Achieving this utopian vision thus requires clear documents describing

1. the details of the proposed standardized data format, specifying what constitutes essential ancillary information (e.g., time, place, S, T, ...),
2. the details of an electronically searchable metadata structure describing such data,
3. the data-quality assessment procedures that are used to select (or flag) potentially problematic results,
4. the data submission schedule, and the role of the Data Management Group in handling and distributing these data, and in preparing any regional data products.

The workshop participants recommended that the CO_2 Science Team should take the responsibility for preparing these documents in concert with a designated Data Management Group. This initial set of documents would then be updated as needed, and would be made openly available to potential participants, both from the U.S. and abroad.

The necessary focus on measurement quality will also require documentation of the various measurement techniques being used, including details of their calibration and quality control procedures. Wherever practical an attempt should be made to standardize the measurement procedures being used by different groups, and to work to improve their efficiency.

Furthermore, it will be essential to implement a structured system of inter-laboratory method comparisons to ensure that the various measurement systems being employed in the field are capable of yielding comparable results. The intent would be to have the Data Management structure available to those contributors (both U.S. and international) who are prepared to work to conform to the proposed reporting formats, and who can adequately document their procedures and data quality.

Ultimately, as sufficient data are gathered to make basin-scale maps of time-varying properties such as surface $p(CO_2)$, it will be necessary to advise the Data Management Group how best to integrate the various datasets to achieve such a goal. At this time, this task is still a research project and needs significant scientific input from individuals with regional knowledge. However, it seems likely that, as the years go by, it may become practical for such datasets to be prepared by the designated Data Management Group in collaboration with the CO_2 Science Team.

6. The Data Management Groups

A national data center, such as the Carbon Dioxide Information and Analysis Center (CDIAC), and a world oceanographic data center (e.g., National Ocean Data Center (NODC)) should be involved in the data management process at all levels. They will participate in the CO_2 Science Team meetings described above and will provide assistance in generating the locally QCed data wherever possible. Regional QC of the data will be primarily performed by the data centers. As soon as the locally QCed data are released to the public, they will also enter into this second level of QC processing. At this point, they will also work closely with the CLIVAR hydrographic data center to ensure that the carbon data are properly integrated with the other hydrographic and tracer measurements.

There was consensus on the need for two complimentary organizations/ facilities for the processing and archival of carbon data and related parameters. Specifically, the workshop participants recommend the following responsibilities for the existing organizations:

6.1 The National Data Center

A National Data Center should provide a short- to medium-term repository for carbon data. Their duties and responsibilities would include acquiring carbon data from investigators, acquiring/confirming associated metadata, format conversions, quality control/quality assurance with feedback from investigators, version control for submitted datasets providing data access "tools" as appropriate, preparation and distribution of datasets and data products to meet requirements by the carbon data community.

These data will be evaluated with respect to both U.S. and international data, and both modern and historical data. Although the National Data Center will have primary responsibility for processing and maintaining the regionally QCed research-quality data sets, NODC will be directly involved at all levels of processing, lending their unique skills to this work. Once all parties are satisfied with the quality of these data they will be posted to the National Data Center WWW site as a research-quality data set, available along with other U.S. and international research-quality data.

At this point, the preliminary locally QCed data will be removed from the National Data Center web site to minimize confusion over different data versions. The National Data Center will also produce their printed Numeric Data Packages (NDPs) describing the regionally QCed data sets, to ensure availability of data even to those without web access. Over the long term both NODC and the National Data Center will maintain the research-quality data sets. The role outlined above also takes advantage of strong links with international investigators and will help with the integration of atmospheric, terrestrial, and ocean carbon data at the research level.

6.2 The World Oceanographic Data Center

The U.S. World Data Center A for Oceanography (NODC/WDC A) should provide the permanent archival for ocean carbon and hydrographic data. The duties and responsibilities would include acquiring data from data centers, acquiring/confirming associated metadata, providing quality assessment with feedback to the National Data Center, integration into large oceanographic databases, allowing easy retrieval of general oceanographic data and carbon data, as required, data exchange with other international data centers, and preparation of large datasets, and their distribution to the user community via the Internet and CD ROMs.

This split in activities and responsibilities makes the best use of each organization's strengths, exploiting strong working relationships at the investigator level and NODC's expertise with very large databases and international data exchange. Investigators will not need to interact with both agencies. The two organizations will need to coordinate version control and data exchange procedures to make this process as efficient as possible.

There is also a need to make historical carbon data available to researchers. Many of the important data holdings remain in the hands of a few investigators and are at risk of loss to the community when these investigators retire. The integration of historical data with modern data will need a cautious approach. There are serious known problems with some historical data, due primarily to limitations in methodology, a lack of appropriate standards, and insufficient metadata. There is substantial research required to evaluate the accuracy and precision of some historical datasets and to assess their applicability to questions of changes in the ocean carbon system over time. In the short to medium term, it would be useful to develop an inventory of historical datasets and assign a priority for their recovery and review.

7. Data Accessibility for Carbon Cycle Ocean Measurements

The ocean carbon database must be maintained in a manner that makes it accessible to a range of scientists, both inside and outside of the immediate ocean carbon data measurement community. The data should be publicly accessible through the Internet, as well as made available at suitable intervals on media that can be distributed to those that lack Internet access. The web-based user interface for the data should be uniform across data versions, measurement types, and to the greatest degree possible, should include historical data and international datasets.

Two types of data subsetting and downloading are required: (1) "large-scale" subsetting in which the entire holdings for a particular group of parameters can be queried, based upon constraints such as latitude, longitude, and depth range, date range, and/or seasonality; and (2) "per cruise" subsetting, in which users recover the measurements from a particular cruise or mooring. The downloaded subsets should be made available in a variety of user-specifiable formats. The list of formats to be supported should address the range of applications commonly used by interested communities and may need to be adapted with time as applications change.

The data access system must include on-line browsing and graphics tools that allow users to quickly determine when and where measurements are available and provide plots and visualizations to give the user a "quick look" at the data. Such tools should enable a user to make quick comparisons of repeat sections, underway tracks, and cross-over points. They should enable a user to visualize changes to the data between update versions.

The Live Access Server (LAS) is a proven system that addresses the community's needs with respect to ease of use, uniformity of interface, data download and visualization capabilities, and comparison/fusion capabilities. Both the discrete and underway/mooring data management discussion groups recommended use of LAS as a web interface. Prototype systems are already in operation at PMEL (e.g., `http://ferret.wrc.noaa.gov/underway/main.html`). Enhancements to LAS are suggested to allow a comparison between measured and gridded data types, integrated presentation of data with metadata, and on-the-fly merging of data from different sources.

Audit tracking of changes to the database is desirable as a tool to locate the source of changes in results that are generated over time. Given the need for continual updates to the databases it is recognized that thorough audit tracking may be a costly demand to place upon the data management system. To keep costs within acceptable bounds it was agreed that check points ("time stamping") should occur at intervals—either fixed intervals of time or intervals measured in changes to the database. Determination of the intervals to be used will occur only after the initial database management strategy is determined.

Metadata management, which includes free text content such as shipboard logs, should be fully integrated into the data access system. It should be possible to locate specific items of metadata based upon parameter type and cruise (or location/date) without the need for a person to read through large amounts of text.

In addition on-line "fusion" with historical data and databases maintained by international partners should be possible. The long-term responsibility for integration of U.S. ocean carbon measurements with historical and international data sets is the responsibility of the NODC; however, there is a need for quick-look "fusion" of data sources as a normal part of ongoing research activities. The process of fusion through a Web interface should be designed to ensure that the user is aware that data of varying quality are being merged. The resulting output products (files and graphics) should clearly document the multiple sources of data that were fused.

Two classes of gridded data were identified. (1) Gridded reference fields—

created through research-driven, proposal funded efforts—should be made available through the data access system in such a way that the observations and gridded products can be compared. (2) For modelers it may be desirable to generate gridded fields on the fly from observations, given suitable user inputs of grid resolution and gridding parameters. The data management system should provide a stable programmers' interface, so it is accessible by machine (other programs) as well as by users. This provides an essential level of adaptability when considering alternative accessibility techniques over the lifetime of the program.

8. Numerical Modeling and Data Assimilation

Given the large inherent space and time variability of the inorganic carbon system in the ocean, the ability to directly observe and document such variations will remain limited by necessity. As a consequence, reaching the goal of characterizing the regional to global distribution and seasonal to decadal scale variability of carbon sinks and sources in the ocean from observations requires the use of interpolation, extrapolation, and other diagnostic techniques. Diagnostic modeling (inverse models up to full data assimilation systems) provide an ideal means to address this issue by generating complete, dynamically consistent ocean carbon fields that incorporate data when and where they are available. Diagnostic modeling therefore provides a natural framework for integrating the limited in situ observation with the other elements of ocean carbon cycle research such as satellite remote sensing, process studies, and prognostic modeling. Diagnostic models and their products have a long history in the atmospheric sciences in connection with weather forecasts and reanalysis of climate variability and are increasingly used in physical oceanography to estimate the state of the ocean circulation. One such project is the ECCO Project (Estimating the Circulation and Climate of the Ocean), which is now routinely making their reanalysis products available to the community (see `http://www.ecco.ucsd.edu`). However, diagnostic mathematical methods have only very recently begun to be used in ocean carbon-cycle research and the experience is therefore limited. Nevertheless, it is expected that inverse modeling approaches will develop rapidly over the next years and become standard tools for the analysis of the state of the ocean carbon cycle and many other applications (see JGOFS Report from IGBP/EU Meeting in Paris). This section outlines the data policy and data management requirements from the perspective of diagnostic modeling.

Figure 1 shows the main components of the envisioned diagnostic modeling framework and how they interact with the other elements of an integrated ocean carbon cycle program. The two central components are formed by the ocean carbon data centers and the ocean carbon assimilation systems. Together they provide the structure by which the data streams originating from the various observation platforms are collected, assimilated, and turned into higher-level data products.

The data management centers play a crucial role in this diagnostic modeling framework, since they act as the collection points for the various types

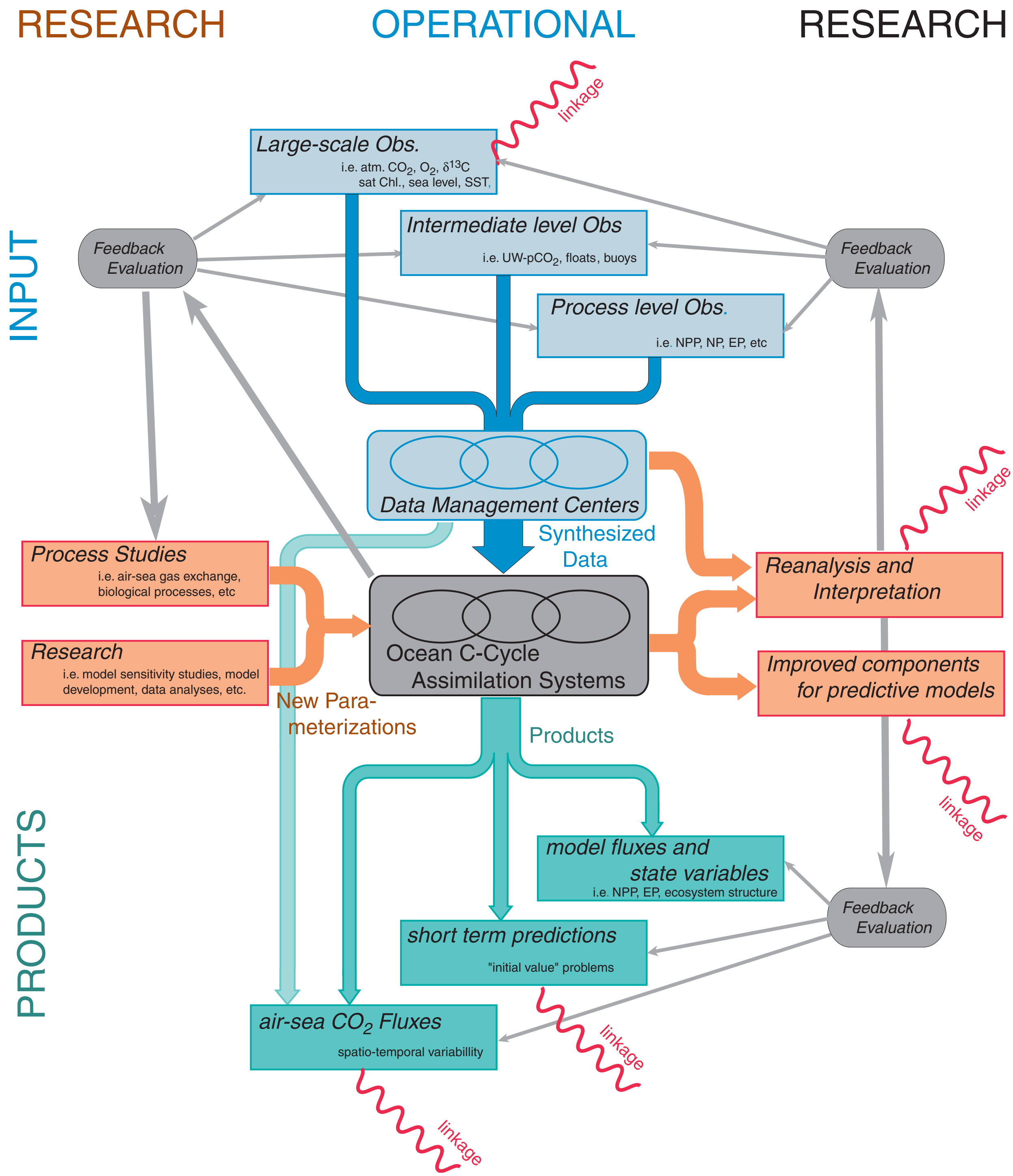


Figure 1: Conceptual diagram of the envisioned diagnostic modeling framework. At the center is an array of diagnostic models that integrate the different types and spatio-temporal resolution of the data into their model and produce "best" estimates of the present and past state of the CO_2 fluxes across the air-sea interface and of the state of the carbon system in the ocean in general. The preparation of the data that stream into these diagnostics need the establishment and support of dedicated data management centers that collect the data, perform additional quality control, and provide long-term availability of these data, etc.

and levels of data streams. For many types of data, particularly for those collected on space-borne platforms, such centers are already in existence. However, for many other data streams, for example, those associated with the rapidly increasing number of underway $p(CO_2)$ data, such data centers need to be established and supported. The data synthesis efforts at these data centers should also include quality control procedures that extend beyond the initial quality control done at the level of the individual observations. This includes, for example, investigation of the internal consistency of the data as well as testing for long-term precision and accuracy of the data. High priority should also be given to fully documenting the various data products and streams (metadata). This is particularly important, since re-analysis projects in the future with substantially more sophisticated models than those existing today require that high-quality and well-documented data are taken now. The multi-tier data management approach proposed in this document appears to address the needs from the perspective of diagnostic modeling.

The requirements of diagnostic modeling studies with regard to data policy are less stringent and depend on the particular aim of the project. Forecast studies will require very rapid dissemination of the data, whereas re-analysis projects don't have such needs. For most diagnostic modeling studies, the timeline of data dissemination proposed in this document seems adequate. Once diagnostic models have advanced to the point where forecast studies could become important, faster availability of some of the data streams might be helpful. Nevertheless, one needs to emphasize that fast dissemination should not occur at the expense of data quality. Ensuring highest quality already during the collection of the data and thereby eliminating the need for extensive post-calibration will be a very effective manner to obtain fast and reliable data for inverse modeling studies.

The scope of diagnostic modeling extends beyond the optimal estimation of the state of the ocean carbon cycle. Such studies can also help to address gaps in the knowledge and pinpoint areas in space and time that need additional observations (through network optimization studies) as well as indicate the need for new types of observations. Close interaction between observational and the modeling communities are therefore of great benefit for both communities and should be supported.

9. Coordination with International Programs

There was extensive discussion of the relationship between U.S. and non-U.S. programs. There was agreement that there was value to international cooperation, although there are some issues such as timeliness of data availability that must be resolved. It was agreed that interagency and national data sets should be made available with data from other countries. Attempts to merge U.S. and non-U.S. data into a centralized database should weigh several factors, the most important being the availability of information necessary to assess the quality of the data. The availability and submission of accompanying metadata is essential to the successful integration of any outside data set. Adherence to data reporting, metadata reporting, and quality control requirements as set forth in this workshop report will help facilitate the successful integration of data from other countries. Just as the workshop participants agreed it was important to evaluate the quality and consistency of U.S. carbon measurements, the participants emphasized the importance of attempting to evaluate the quality and consistency of international measurements.

Discussion and recommendations could be grouped into two categories, those aimed at improving international measurements before data submission to a central database and those aimed at evaluation after submission. Recommendations aimed at improving international measurements directly included making planned "procedures manuals" available to the international community, promoting the availability and use of Certified Reference Materials (CRMs), and establishing formal intercomparison exercises. To aid evaluation of international measurements after submission, it was recommended that international measurements be checked and flagged, using the same strategy adopted for assessing U.S. measurements, and that these assessment records too should become part of the centralized database. Inevitably researchers worldwide will compare U.S. measurements to non-U.S. measurements and these comparisons, where possible, should be reflected in the documentation for the central database.

Appendix I

Acronyms

CDIAC	Carbon Dioxide Information Analysis Center (DOE)
CCSP	U.S. Carbon Cycle Science Plan
CLIVAR	Climate Variability and Predictability (WCRP program)
CO_2	Carbon dioxide
CTD	Conductivity/Temperature/Depth profiler
DIC	Dissolved Inorganic Carbon
DOE	Department of Energy
GOOS	Global Ocean Observing System (WMO)
NOAA	National Oceanic and Atmospheric Administration
NODC	National Oceanographic Data Center
NSF	National Science Foundation
OAR	Oceanic and Atmospheric Research (NOAA)
OGP	Office of Global Programs (NOAA)
$p(CO_2)$	partial pressure of carbon dioxide
PMEL	Pacific Marine Environmental Laboratory (NOAA)
TA	Total Alkalinity
TCO_2	Total dissolved inorganic carbon
VOS	Volunteer Observing Ships
WOCE	World Ocean Circulation Experiment

Appendix II

Ocean Carbon Cycle Data Management Workshop Agenda

October 25–26, 2001
Pacific Marine Environmental Laboratory
Seattle, Washington

25 October

0800–0830	Coffee	
0830–0840	Welcome	Eddie Bernard
0840–0850	Introduction and Logistics	Dick Feely, Marilyn Roberts
0850–0920	Agency Perspectives: NOAA, NSF, NASA Future program directions	Lisa Dilling, Mike Johnson, Don Rice, Chuck McClain
0920–0945	Overview of Ocean Carbon Programs	Dick Feely
0945–1000	Break	
1000–1030	Water Column Data and QA/QC Protocols	Plenary Talk: Chris Sabine
1030–1100	Underway Data and QA/QC Protocols	Plenary Talk: Rik Wanninkhof
1100–1130	Mooring data and QA/QC Protocols	Plenary Talk: Francisco Chavez
1130–1200	Data Assimilation into Models	Plenary Talk: Niki Gruber
1200–1300	Lunch	
1300–1330	Live Access Servers and Data Management	Plenary Talk: Steve Hankin
1330–1400	Metadata Requirements	Plenary Talk: Margarita Conkright
1400–1430	OA/QC Procedures at CDIAC	Plenary Talk: Alex Kozyr
1430–1445	Break	
1445–1600	Breakout Group I: Underway Data Requirements	Discussion Leader: Rik Wanninkhof; Rapporteur: Andrew Dickson
1445–1600	Breakout Group II: Water Column Data Requirements	Discussion Leader: Chris Sabine; Rapporteur: Jim Swift
1445–1600	Breakout Group III: Mooring Data Requirements	Discussion Leader: Nick Bates; Rapporteur: Francisco Chavez
1600–1615	Break	
1615–1700	Breakout Group Reports	Breakout Group Leaders
1700–1730	General Discussion	Dick Feely
1730	Adjourn	
1900–2100	Dinner at Anthony's Home Port	

26 October

0800–0830	Coffee	
0830–0930	Should there be one or more central facilities responsible for archiving all carbon and related parameter data? If so, how should the data center(s) be structured to meet the needs of the community?	Discussion Leader: Robin Brown; Rapporteur: T.-H. Peng
0930–1000	How should the data management center(s) interact with the PIs to assure that high quality CO_2 data are being archived? How should the scientists be acknowledged for including their data in the databases?	Discussion Leader: Andrew Dickson; Rapporteur: Chuck McClain
1000–1030	Break	
1030–1130	Will there be public access to new data sets that still need QA/QC work before inclusion into long-term databases? Should we develop improved WWW-based tools for data display/distribution? What information should be provided by the server?	Discussion Leader: Steve Hankin; Rapporteur: Dave Glover
1130–1300	Lunch	
1300–1400	Should the data management center(s) be responsible for providing gridded data sets for modeling purposes? Will the gridded data sets be developed in a uniform manner?	Discussion Leader: Joanie Kleypas; Rapporteur: Bob Key
1400–1430	Should the historical data sets be integrated with the new Carbon Cycle Science Program database? If so, how?	Discussion leader: Robin Brown; Rapporteur: Margarita Conkright
1430–1500	Break	
1500–1530	How should the interagency and national data sets be integrated with data from other countries?	Discussion Leader: Tom Boden; Rapporteur: Margarita Conkright
1530–1630	General Discussion and Summary	Dick Feely
1630	Adjourn	

Notes: Each discussion leader is required to provide a written report of their individual session. Dick Feely and Chris Sabine will summarize the reports into a final workshop document.

Appendix III

List of Participants

Nicholas R. Bates
Bermuda Biological Station for Res., Inc.
17 Biological Station Lane
Ferry Reach, GE 01
BERMUDA
Tel: 441-297-1880
Fax: 441-297-8143
E-mail: nick@sargasso.bbsr.edu

Thomas Boden
Oak Ridge National Laboratory
P.O. Box 2008, MS 6335
Oak Ridge, TN 37831-6335
USA
Tel: 865-241-4842
Fax: 865-574-2232
E-mail: bodenta@ornl.gov

Robin Brown
Ocean Science and Productivity Division
Institute of Ocean Sciences
P.O. Box 6000
Sidney, BC V8L 4B2
CANADA
Tel: 250-363-6378
Fax: 250-363-6310
E-mail: BrownRo@pac.dfo-mpo.gc.ca

John Bullister
NOAA/PMEL/OCRD
7600 Sand Point Way NE
Seattle, WA 98115-6349
USA
Tel: 206-526-6741
Fax: 206-526-6744
E-mail: bullister@pmel.noaa.gov

Alexander S. Bychkov
North Pacific Marine Science Organization
c/o Institute of Ocean Sciences
P.O. Box 6000
Sidney, BC V8L 4B2
CANADA
Tel: 250-363-6364
Fax: 250-363-6827
E-mail: bychkov@pices.int

Kenneth G. Caldeira
Lawrence Livermore National Laboratory
Mail Stop L-103
7000 East Avenue
Livermore, CA 94550
USA
Tel: 925-423-4191
Fax: 925-422-6388
E-mail: kenc@llnl.gov

Francisco Chavez
Monterey Bay Aquarium Research Institute
P.O. Box 628
Moss Landing, CA 95039-0628
USA
Tel: 831-775-1709
Fax: 831-775-1645
E-mail: chfr@mbari.org

Margarita Conkright
NESDIS HQTR - E/OC5
Bldg: SSMC3 Rm 4350
1315 East-West Hwy
Silver Spring, MD 20910-3282
USA
Tel: 301-713-3290 x193
Fax: 301-713-3303
E-mail: Margarita.Gregg@noaa.gov

Catherine E. Cosca
NOAA/PMEL/OCRD (JISAO)
7600 Sand Point Way NE
Seattle, WA 98115-6349
USA
Tel: 206-526-6183
Fax: 206-526-6744
E-mail: cosca@pmel.noaa.gov

Michael DeGrandpre
Department of Chemistry
The University of Montana
Missoula, MT 59812
USA
Tel: 406-243-4118
Fax: 406-243-4227
E-mail: mdegrand@selway.umt.edu

List of Participants (cont.)

Andrew Dickson
Marine Physical Laboratory
Scripps Institution of Oceanography
University of California, San Diego
Mail Code 0902
9500 Gilman Drive
La Jolla, CA 92093-0902
USA
Tel: 858-534-2582
Fax: 858-456-9079
E-mail: adickson@ucsd.edu

Lisa Dilling
NOAA/OAR - R/OGP
1100 Wayne Ave., Suite 1210
Silver Spring, MD 20910-5603
USA
Tel: 301-427-2089 x106
Fax: 301-427-2073
E-mail: dilling@ogp.noaa.gov

Scott C. Doney
Climate and Global Dynamics
National Center for Atmospheric Research
P.O. Box 3000
Boulder, CO 80307-3000
USA
Tel: 303-497-1639
Fax: 303-497-1700
E-mail: doney@ncar.ucar.edu

Rana A. Fine
RSMAS/MAS
University of Miami
4600 Rickenbacker Causeway
Miami, FL 33149-1098
USA
Tel: 305-361-4722
Fax: 305-361-4917
E-mail: rfine@rsmas.miami.edu

Marjorie A.M. Friedrichs
Center for Coastal Physical Oceanography
Old Dominion University
Crittenton Hall
Norfolk, VA 23529
USA
Tel: 757-683-5560
Fax: 757-683-5550
E-mail: marjy@ccpo.odu.edu

Dave Glover
Woods Hole Oceanographic Institution
Dept. of Marine Chemistry and Geochemistry
MS #25
Woods Hole, MA 02543
USA
Tel: 508-289-2656
Fax: 508-457-2193
E-mail: david@whoi.edu

Dana Greeley
NOAA/PMEL/OCRD
7600 Sand Point Way NE
Seattle, WA 98115-6349
USA
Tel: 206-526-6693
Fax: 206-526-6744
E-mail: greeley@pmel.noaa.gov

Nicolas Gruber
Center for Earth Systems Research
University of California, Los Angeles
405 Hilgard Avenue
Los Angeles, CA 90095
USA
Tel: 310-825-4772
Fax: 310-206-3051
E-mail: ngruber@igpp.ucla.edu

Steve Hankin
NOAA/PMEL/OCRD
7600 Sand Point Way NE
Seattle, WA 98115-6349
USA
Tel: 206-526-6080
Fax: 206-526-6744
E-mail: hankin@pmel.noaa.gov

Dennis Hansell
RSMAS/University of Miami
Div. of Marine and Atmospheric Chemistry
4600 Rickenbacker Causeway
Miami, FL 33149
USA
Tel: 305-361-4078
Fax: 305-361-4689
E-mail: dhansell@rsmas.miami.edu

List of Participants (cont.)

Matthew Harrison
NOAA/GFDL
PO Box 308
Princeton, NJ 08542-0308
USA
Tel: 609-452-6579
Fax: 609-987-5063
E-mail: Matthew.Harrison@noaa.gov

Gregory Johnson
NOAA/PMEL/OCRD
7600 Sand Point Way NE
Seattle, WA 98115-6349
USA
Tel: 206-526-6806
Fax: 206-526-6744
E-mail: gjohnson@pmel.noaa.gov

Michael Johnson
NOAA/OAR - R/OGP
1100 Wayne Ave., Suite 1210
Silver Spring, MD 20910-5603
USA
Tel: 301-427-2089 x169
Fax: 301-427-2073
E-mail: johnson@ogp.noaa.gov

Robert Key
Geosciences Department
Princeton University
B80 Guyot Hall
Princeton, NJ 08544
USA
Tel: 609-258-3595
Fax: 609-258-1274
E-mail: key@wiggler.Princeton.edu

Joanie A. Kleypas
UCAR
P.O. Box 3000
Boulder, CO 80307-3000
USA
Tel: 303-497-1316
Fax:
E-mail: kleypas@ucar.edu

Alexander Kozyr
CDIAC/Environmental Sciences Division
Oak Ridge National Laboratory
U.S. Department of Energy
Building 1509, Mail Stop 6335
Oak Ridge, TN 37831-6335
USA
Tel: 865-576-8449
Fax: 865-574-2232
E-mail: ako@cdiac.esd.ornl.gov

Charles R. McClain
NASA/Goddard Space Flight Center
Mailstop 970.2, Building 28, Room W107
Greenbelt, MD 20771-0001
USA
Tel: 301-286-5377
Fax: 301-286-1755
E-mail: mcclain@calval.gsfc.nasa.gov

Frank Millero
RSMAS/University of Miami
Division of Marine and Atmospheric Chemistry
4600 Rickenbacker Causeway
Miami, FL 33149-1098
USA
Tel: 305-361-4731
Fax: 305-361-4689
E-mail: fmillero@rsmas.miami.edu

Calvin Mordy
NOAA/PMEL/OCRD
7600 Sand Point Way NE
Seattle, WA 98115-6349
USA
Tel: 206-526-6870
Fax: 206-526-6744
E-mail: mordy@pmel.noaa.gov

Raymond G. Najjar
Department of Meteorology
503 Walker Building
The Pennsylvania State University
University Park, PA 16802-5013
USA
Tel: 814-863-1586
Fax: 814-865-3663
E-mail: najjar@essc.psu.edu

List of Participants (cont.)

Tsung-Hung Peng
NOAA/AOML - R/AOML/OCD
4301 Rickenbacker Causeway
Miami, FL 33149-1026
USA
Tel: 305-361-4399
Fax: 305-361-4392
E-mail: peng@aoml.noaa.gov

Marilyn Roberts
NOAA/PMEL/OCRD
7600 Sand Point Way NE
Seattle, WA 98115-6349
USA
Tel: 206-526-6252
Fax: 206-526-6744
E-mail: roberts@pmel.noaa.gov

Detlef B. Stammer
Physical Oceanography Research Division
Scripps Institution of Oceanography
9500 Gilman Drive
La Jolla, CA 92093-0230
USA
Tel: 858-822-3376
Fax: 858-534-9820
E-mail: dstammer@ucsd.edu

James Swift
WOCE Hydrographic Program Office
Physical Oceanography Research Division
SIO/University of California, San Diego
Mail Code 0214
9500 Gilman Drive
La Jolla, CA 92093-0214
USA
Tel: 858-534-3387
Fax:
E-mail: jswift@ucsd.edu

Robbie Toggweiler
NOAA/OAR - R/GFDL
P.O. Box 308
Princeton, NJ 08542-0308
USA
Tel: 609-452-6659
Fax: 609-987-5063
E-mail: jrt@gfdl.gov

Rik Wanninkhof
NOAA/AOML/OCD
4301 Rickenbacker Causeway
Miami, FL 33149-1026
USA
Tel: 305-361-4379
Fax: 305-361-4392
E-mail: wanninkhof@aoml.noaa.gov