US EPA ARCHIVE DOCUMENT

United States Environmental Protection Agency
Office of Research and Development
Washington, DC 20460
August 1990

1990 Demonstration Project

Data Management System for the Near Coastal Demonstration Project - Volume I

Environmental Monitoring and Assessment Program

Data Management System
for
Near Coastal Demonstration Project

The Environmental Monitoring Assessment Program
Near Coastal Task Group

By

Jeffrey S. Rosen[1]
Jay Beaulieu[1]
Melissa Hughes[1]
Henry Buffum[1]
Jane Copeland[1]
Raymond Valente[2]
John Paul[3]
Fred Holland[4]
Steven Schimmel[3]
Charles Strobel[2]
Kevin Summers[3]
K. John Scott[2]
Jeffrey Parker[2]

1 Computer Sciences Corporation (CSC)
2 Science Applications International Corporation (SAIC)
3 U.S. Environmental Protection Agency (US-EPA)
4 Versar

Acknowledgements

This document is the product of a large national team effort. The authors wish to thank many individuals for extensive inputs. This work is made possible through the efforts of the EMAP Steering committee. The overall design and approach have been developed with extensive input from the EMAP Information management committee chaired by Eugene Meier. 3 members of the IMC have made significant inputs to the approach discussed in this document they are: Rodney Slagle of Lockheed, Edward Barrows of NSI, and Charles Liff of Utah State University.

Extensive input to this approach came from three organizations within the National Oceanographic and Atmospheric Agency (NOAA). Specifically, Daniel Basta, director of the Office of Strategic Assesment, James Price, Data Base Administrator of the National Status and Trends program, and Brian O'Gorman from the National Marine Fisheries Service Laboratory in Woods Hole Massachusetts. J.A. Ranasinghe and Jeffrey Frithsen of Versar, Inc. supplied data base structures, variable names and species codes used as starting points for the Benthic Community approach outlined in this document.

This document was reviewed over the course of 6 months. Ms. Katherine Peres of Lockheed, Las Vegas, reviewed and helped in the editing of this docuemnt. Mr. James Pollard of Lockheed Las Vegas helped write the Quality Assurance section of this document. Ms. Patricia Gant of the US-EPA, Environmental Research Laboratory, Narragansett reviewed the document.

Support in the production of this document came from Ms. Collette Brown, Ms. Patricia Bussiere, Ms. Agnes O'Rielly, and Ms. Dinalyn Spears.

Table of Contents

APPENDICES

PART I - OVERVIEW

1.0 Introduction

This document presents the rationale, approach, objectives, and plan for establishing an information management system to meet the needs of the Near Coastal task group of the Environmental Monitoring and Assessment Program (EMAP-NC). The Near Coastal task group is one of six elements of EMAP, a nationwide program being conducted by the U.S. Environmental Protection Agency's (EPA) Office of Research and Development (ORD). A conceptual overview of EMAP has been written by the Office of Modelling, Monitoring Systems, and Quality Assurance (EMAP 1990). The Goal of the EMAP-NC task group is to periodically assess the status and trends in ecological condition of the Nation's near coastal ecosystems. The program is specifically designed to assess changes in ecological condition over broad biogeographical regions (e.g., the Virginian Province, the Mid-Atlantic region, the Gulf of Mexico) and over long time periods (e.g., years to decades).

The goal of all the EMAP task groups (i.e. elements) is to document the condition of the Nation's ecosystems in an integrated manner, on a continuing basis. The ecosystems include forests, wetlands, coastal waters and the Great Lakes, inland surface waters, agricultural lands, and arid lands. Although EMAP is designed and funded by ORD, other offices and regions within EPA (e.g., Office of Marine and Estuarine Protection, Region III), and other federal agencies (e.g., Office of Oceanography and Marine Assessments of the National Oceanic and Atmospheric Administration (NOAA); the U.S. Forest Service, and the U.S. Fish and Wildlife Service (FWS)) have contributed to its development and will participate in the collection and use of EMAP data. When fully implemented, EMAP will form a complex national ecological monitoring network with each ecosystem task group sampling approximately 800 sites across the United States. Because a large proportion of the actual

field sampling during the implementation phase will be conducted by other federal, state, and local agencies, EMAP must develop interagency agreements with these other agencies which include the transfer and sharing of data.

EMAP is not a compliance or enforcement monitoring program; nor is it a diagnostic monitoring and assessment program. Site-specific field studies, bioassays, and fate and effects modeling programs are generally used to support compliance determinations, enforcement actions, and diagnostic programs. EMAP is designed to provide the information required to formulate environmental protection policies of the 1990's and beyond by providing answers to the following questions:

- o What is the status, extent, and geographical distribution of the Nation's important ecological resources?
- o What proportion of these resources are declining or improving; where, and at what rate?
- o What are the likely factors that are contributing to declining condition?
- o Are control and mitigation programs achieving overall improvement in ecological conditions?
- o Which resources are at greatest risk to pollution impacts?

The information generated by EMAP will be used by the following:

- o Decision makers at all levels of government who set environmental policy;
- o Resource managers and regulators who require an objective basis for allocation of resources and prioritization of actions, especially those directed toward protecting and enhancing ecological resources;

- o Researchers working on the development and understanding of ecological indicators and processes; and
- o Those interested in evaluating the effectiveness of the Nation's environmental policies for protecting and enhancing ecological resources (e.g., the EPA Administrator and Senior Management staff, Congress, and the public).

2.0 Functional Statements

2.1 Functional Statement for the Near Coastal Task Group

In many coastal regions, water and sediment quality and the abundance of living resources are perceived to have declined over the past 10 to 15 years, despite the implementation of more strict control programs. Increasingly, reports appear in the popular press (e.g., Morganthau 1988; Toufexis 1988; Smart et al. 1987) and scientific journals describing declining near coastal environmental quality, as exemplified by the following:

- o Increases in the frequency, duration, and size of water masses that do not contain sufficient oxygen to sustain living resources (e.g., USEPA 1984; Officer et al. 1984; Parker et al. 1986; Rabalais et al. 1985; Whitledge 1985);

- o Accumulation of contaminants in sediment and in the tissues of fish and shellfish to levels which threaten humans and the vitality of fish and shellfish populations (OTA 1987; NRC 1989);

- o Declines in the amount and quality of ecologically important habitats (e.g., wetlands and SAV) that are associated with high population levels of waterfowl, shorebirds, fish, and shellfish (Frayer et al. 1983; The Conservation Foundation 1988; Orth and Moore 1983);

- o Increased evidence that many restoration and mitigation efforts

have not replaced losses of critical habitats (Sanders 1989; The Conservation Foundation 1988;

- o Increased incidence of pathological problems in fish and shellfish in systems that have high levels of chemical contamination (Sinderman 1979; O'Connor et al. 1987; Buhler and Williams 1988; Capuzzo et al. 1988);

- o More favorable conditions for and increased frequency and persistence of algal blooms and associated decreases in water clarity (USEPA 1984; Pearl 1988);

- o Increased incidence of closures of beaches, shell fishing grounds, and fisheries because of pathogenic and chemical contamination (Smart et al. 1987; Food and Drug Administration 1971, 1985; Hargis and Haven 1988; Broutman and Leonard 1988; Leonard et al. 1989); and

- o Increased incidence of human health problems from consumption of contaminated fish and shellfish or swimming in contaminated waters (Fein et al. 1984; Jacobson et al. 1983, 1984; Malins 1989).

The above symptoms of the declining environmental quality of near coastal systems are specific to particular areas; nonetheless, they are characteristic of the kinds of problems facing coastal areas from New England to Alaska.

The specific objectives of EMAP-NC are as follows:

- o Provide a quantitative assessment of the regional extent of coastal environmental problems by measuring pollution exposure and ecological condition;

- o Measure changes in the regional extent of environmental problems

for the Nation's near coastal ecosystems;

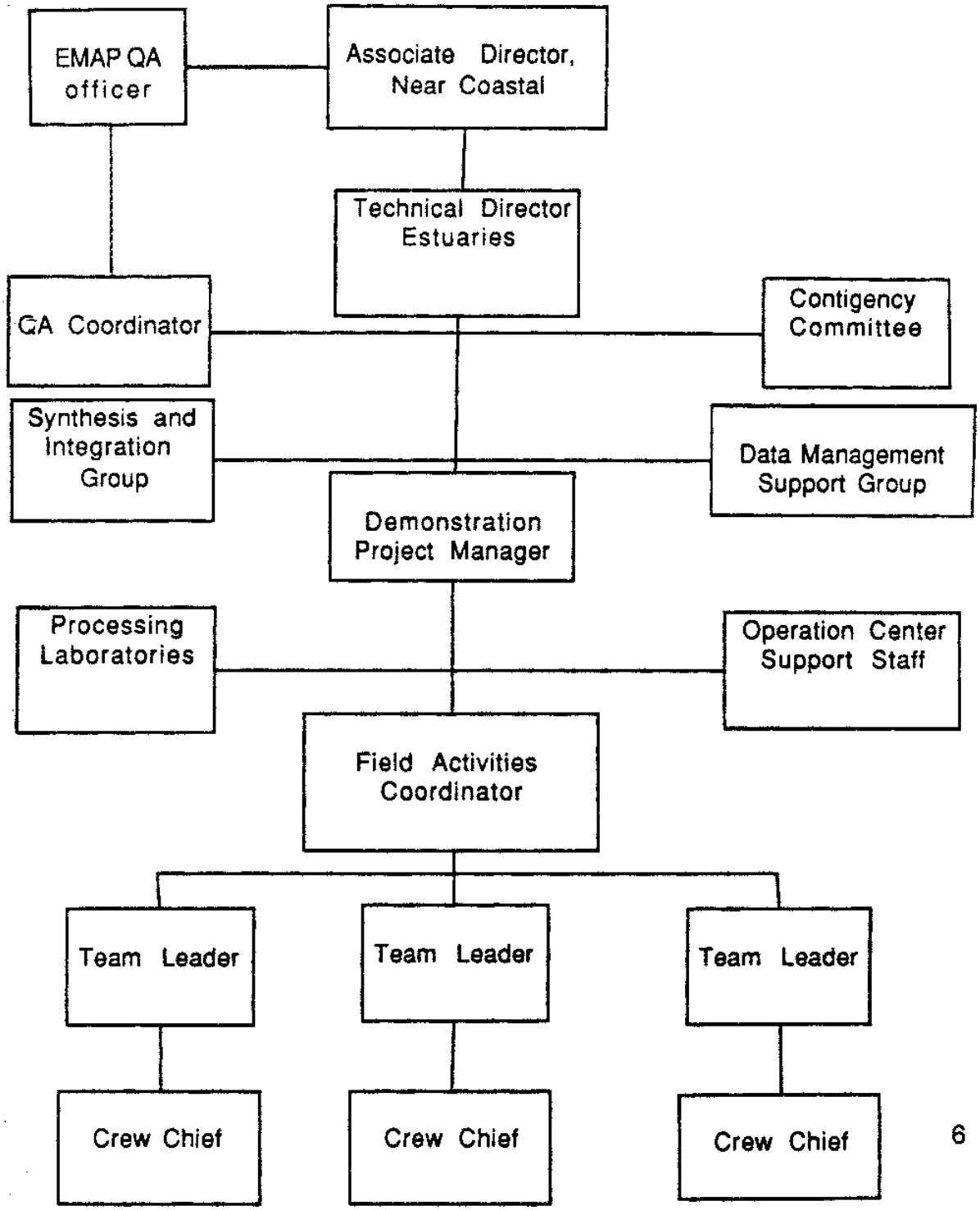
- o Identify and evaluate associations between the ecological condition of the Nation's near coastal ecosystems and pollutant exposure, as well as other factors known to affect ecological condition (e.g., climatic conditions, land use patterns); and

- o Assess the effectiveness of pollution control actions and environmental policies on a regional scale (e.g. large estuaries like Chesapeake Bay, major coastal regions like the mid-Atlantic and Gulf Coasts) and nationally.

Detailed discussion of these objectives can be found in the Draft EMAP - Near Coastal Program Plan (reference).

The Administrative structure of the Near Coastal Task group is presented in Figure 1.

Figure One - Administrative structure of EMAP Near Coastal Demonstration Project.



2.2 Functional Statement for the Near Coastal Information Management Team

The Near Coastal Task Group will be collecting a large volume of data. For example, during the EMAP-NC demonstration project (FY 1990 - Virginian Province), more than 5,000 samples will be collected at over 200 sample sites. In addition, many of the parameters (i.e., indicators) that will be measured (e.g., contaminants in fish flesh) involve the collection of many samples and the generation of large amounts of data for each sampling site. As the sampling program is expanded to include other provinces, the quantity of data that will be collected will increase exponentially. The ability of the EMAP-NC program to manage and disseminate the large amounts of information that will be collected will have a major influence on the success of the program. Development of an adequate information management system is, therefore, as important to the success of EMAP-NC as is collection of the data (NRC 1990a).

The large volume of data will be used for a wide variety of analysis and interpretations. Analyses will range from tabular summaries and statistical comparisons to evaluations of spatial distributions using Geographical Information Systems (GIS). An established goal of EMAP-NC is the publication of statistical summaries within nine months after the completion of data collection. A computerized data management system is, therefore, required to ensure that EMAP-NC data are made available for the analyses in a timely manner.

EMAP-NC will be conducting a range of simultaneous activities (e.g., sample collection, laboratory processing, statistical analyses) over broad geographical areas. In order to identify problems, develop alternative plans, control costs, and modify schedules, managers within EMAP will require frequent (daily and weekly) reports on the status of each program activity. Therefore, a project management information system is needed to organize and track EMAP-NC project management data.

3.0 Near Coastal Task Group Information Processing

Data management within EMAP will occur at three levels of organization. Each level will function independently but their activities will be coordinated to form an integrated data management system that covers all ecosystems and related EMAP activities. The three levels of organization are as follows:

- Regional - each region (e.g., Virginian Province) within a task group (e.g., near coastal);
- National - each task group (i.e., ecosystem type), with data aggregated over all regions; and
- Program-wide - entire EMAP program, integrated over all task groups (i.e., national evaluations across multiple ecosystem types).

3.1 Relationship to the EMAP Information Management Committee

EMAP information management activities are coordinated by the EMAP Information Management Committee (IMC). The EMAP-NC Task Group Information Manager (TGIM) is a member of the IMC and is responsible for ensuring that information management activities within EMAP-NC are consistent with EMAP objectives and with activities occurring in other ecosystem task groups. The EMAP-NC TGIM will use the IMC as an advisory group in the development, establishment, and maintenance of the Near Coastal Information Management System (NCIMS). The EMAP-NC team will participate in the development of standards for EMAP data processing through representation on the IMC. The EMAP-NC data management team will adhere to all standards developed by the IMC.

3.2 The Near Coastal Information Center

The NCIMS will be a distributed data base that consists of a central site and multiple remote nodes (see Figure 2). The core of the distributed EMAP-NC data management system is the Near Coastal Information Center (NCIC) located at the Environmental Research Laboratory-Narragansett (ERL-Narragansett). Personnel at this facility are responsible for maintaining a comprehensive data inventory, data set index, code libraries, and data dictionary for EMAP-NC. They will also maintain and disseminate EMAP-NC data and ensure that appropriate data are incorporated into the NCIMS. The Near Coastal Information Processing Center will also support the data processing requirements of the remote nodes and the exchange of data with other agencies and organizations.

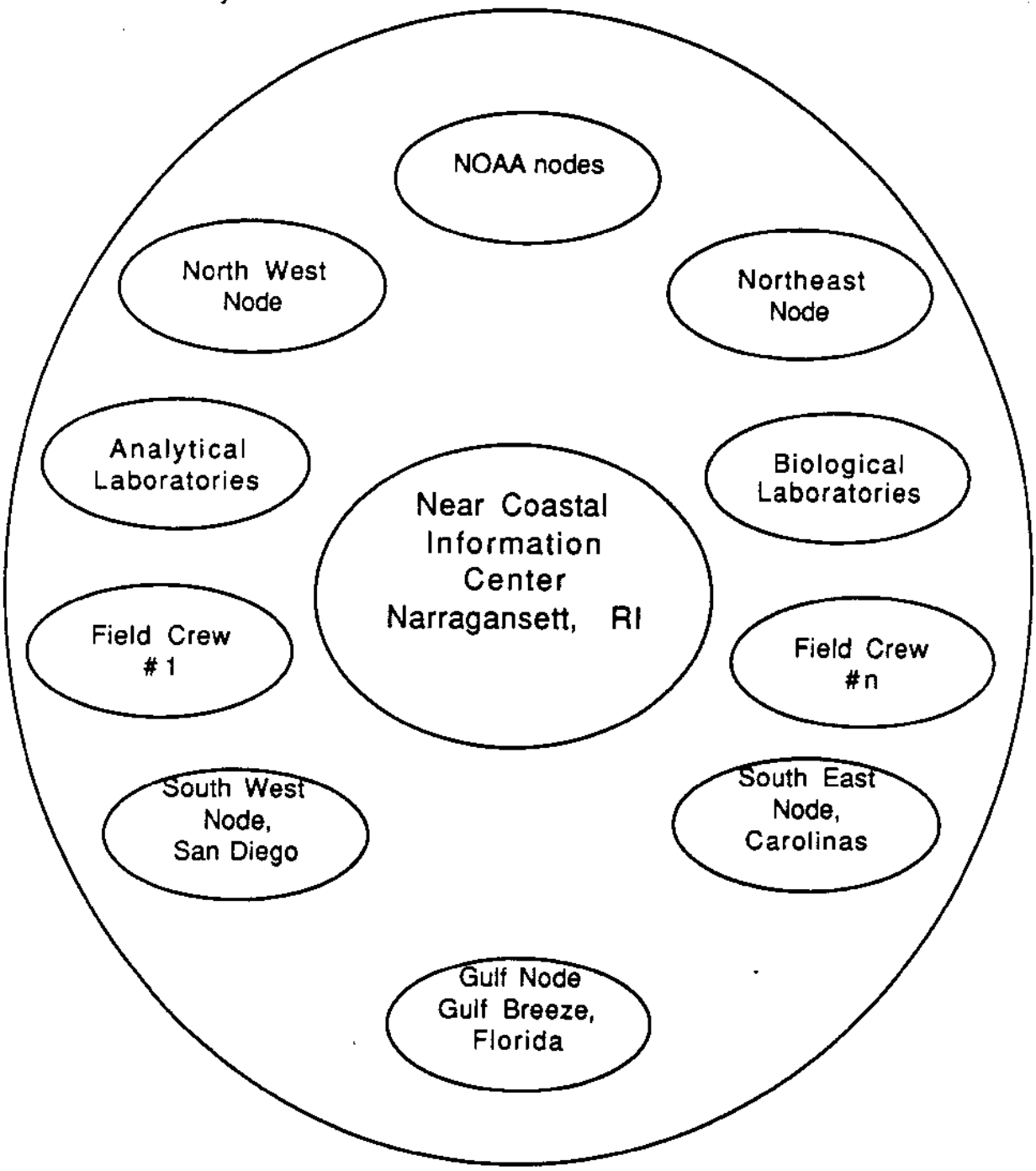
The NCIMS must have the flexibility to handle the array of data types resulting from sample collection and processing. It must also support a variety of analysis, presentation, and reporting activities. For the 1990 Demonstration Project, the Statistical Analysis System (SAS) will be used as the data management system. SAS will also be used for most statistical analyses. SAS has been selected as the data management system because no relational data base system is available to EPA through current contacting mechanisms. When a relational data management system is available to EPA through the Office of Information Resource Management (OIRM), the EMAP-NCIMS will be converted to the selected relational data base system. This will allow users more flexible and efficient access to EMAP-NC data. After the conversion to a relational data base system, SAS will continue to be used as a principal data analysis tool.

3.2.1 Remote Nodes

The two major remote nodes are (1) field teams, responsible for collecting samples and making primary measurements, and (2) laboratories, responsible for processing samples.

The conceptual structure in figure two assumes adequate funding, for hardware, software and personnel. Additionally, this design assumes that the proper resources can be procured by agency mechanisms.

Figure two - Conceptual diagram of the Near Coastal distributed information system.



Specific data management activities that will occur at the remote nodes are:

1 - Field Teams

- o Data collection;
- o Initial calculation of parameter values;
- o Initial entry of data into electronic format;

2 - Remote Laboratories

- o Preliminary data analysis and summarization;
- o Quality assurance for sample tracking, sample preparation, and analytical techniques; and
- o Transfer of appropriate data, in specified electronic formats, to the NCIC, indicating progress on analyses, summarizing results, and identifying potential problems.

4.0 - User Requirements

The users of EMAP NC data can be put into six general categories.

1 - Near Coastal Core Research Group (NC-Core) - Includes those individuals and groups charged with designing, implementing and interpreting the data from the field sampling programs. This group is limited to the individuals at the implementation sites, the field crew members, the technical director (or his designees), and the implementation contingency committee.

Requirements - This group needs to have access to a comprehensive data set including:

- o Project Management Information
- o Sample Tracking
- o Shipment Tracking
- o Raw data files
- o QA/QC reports
- o Field Logs
- o Logistics
- o Summary reports
- o Maps
- o Verified and validated data sets

This group requires access to the data on as close to a real time basis as possible. The data need not be quality assured prior to access by the members of this group. All raw data used by this group must be used with the understanding that the data have not been verified or validated. This group needs access to all data described in the other categories.

2 - Near Coastal EMAP Team - This group includes individuals and groups who are involved in the Near Coastal effort but are not involved in the day to day operations of the field programs. These participants include: NOAA EMAP personnel, logistical support personnel, GIS support personnel, EMAP QA/QC personnel, program reviewers, and headquarters personnel.

Requirements - This group requires access to summary information regarding logistics and project management. They require access to some raw data files but only those which have been validated and verified. They do not require real time access nor do they need to have access to a comprehensive data set.

3 - EMAP Research Group - Includes all researchers directly involved in the design, implementation and analysis of the national EMAP program. These individuals include members of other task groups,

members of the synthesis and integration team, and members of other agencies directly involved in EMAP (e.g. The National Forest Service, U.S. Geological Survey (USGS), Fish and Wildlife, etc.).

Requirements - This group requires final summaries regarding logistics and project management. They require access to some raw data files but only those which have been validated and verified. They do not require real time access nor do they need to have access to a comprehensive data set. They need data in a context which can be integrated with data from other disciplines. Document summaries with interpretation and graphic outputs will be most useful.

4 - Other Federal and State agencies involved in environmental monitoring. This group includes the EPA regions, other EPA offices, state agencies, universities and other research organizations.

Requirements - This group requires access only to validated and verified data integrated to the station level. They will not require any data regarding the logistics and project management. They will need summarized characterization data for each station sampled. They will also need access to an index of available data both EMAP generated and Historical data. They will require access to some raw data files but only those which have been validated and verified. They need data in a context which can be integrated with data from other disciplines. Document summaries with interpretation and graphic outputs will be most useful.

5 - Legislatures and Environmental Managers.

Requirements - This group will need summarized and interpreted data. They will not require any data regarding the logistics or project management. This group of users will be best served by published reports, maps, and an on-line summary system.

6 - The general public

Requirements - This group will need summarized and interpreted data. They will not require any data regarding the logistics or project management. This group of users will be best served by published reports, maps, and an on-line summary system. When there is a requirement for detailed data special reports/presentations will be prepared by the NCIMT.

4.1 NCIC Products and Services

To meet these requirements the NCIMT will develop a distributed data base system. The system will contain the following major components.

- o A project management system - tracks samples, shipments, equipment, personnel, results and expectations.
- o A field computer system - facilitate the calibration of field equipment, navigation, sample collection, logging of information, data entry, and sample delivery to analytical laboratories.
- o A communications system - allows file transfer and program upgrades to occur during the sampling program. The communications systems will facilitate the transfer of data from the field computers to the NCIC VAX computer. Similarly the communications system will allow the field coordination team to transfer instructions, changes, and queries to the field crews.
- o A data processing system (Parser)- processes incoming data from the field crews. It will ensure that the expected files are received and that the formats of those files are correct. The Parser will also facilitate the input of data from data sheets and the verification that the data in electronic form agree with the results entered on the field data sheets. The Parser will prepare the field data which have been received electronically or entered manually, for addition into the data management system. The

parser will also generate a daily report indicating what work was performed the day before, which field teams communicated with the Field Operation Center (FOC) and what sample shipments were sent and received.

- A data management system - The data management system will facilitate the entry, validation, access, dissemination, aggregation and maintenance of the NC data. The data management system will be a tiered system. The first tier will contain data which have not been aggregated and have only been partially validated. These data will be accessible only to group one described above. All other users will be prevented from accessing these data. The second tier of the data base will contain the fully validated data base. It will also contain the first level of data aggregation and verification. This level will be available to groups two, three and four described above. The third tier will contain the fully validated data aggregated to the level of all results for a given station on a given date (an *EVENT*). These data will be available to the all groups except groups five and six. The fourth tier will contain summary data aggregated at the station level. These data will be completely validated and verified. These data will be available to all groups except group 6.

Reports and summary data will be made available in hard copy, in magnetic media and via Compact Laser Disks (CD ROM). The summary results will also be transferred to a system developed by NOAA called COMPAS. The COMPAS system runs in Hypercard on an Apple MacIntosh. These data summaries will be available to the general public.

4.2 Contents of the Near Coastal Information Management System

At a minimum, the NCIMS will contain the following information:

- A record of each sampling event;
- CTD profiles for salinity, temperature, water depth, dissolved oxygen concentration, pH, transmissometry (an estimate of water clarity), fluorometry (an estimate of algal abundance), and photosynthetically active radiation (PAR);
- Concentrations of contaminants, organic content, physical sediment characteristics, and apparent redox potential discontinuity (RPD) depth of sediments for each sampling site;
- Grain size and organic content for each grab sample processed for benthic community parameters;
- Benthic counts and biomass by sampling event and species;
- Counts and sizes for bivalve species collected by the bivalve dredging program;
- Counts and size measurements for fish species, concentrations of contaminants in fish flesh for targeted fish species, gross pathological disorders for targeted fish species at the base sampling sites and for a subset of all species at the indicator testing and evaluation sites, detailed histopathology information for fish that were found to have gross pathological disorders, and detailed histopathology for a subset of all species at indicator testing and evaluation sites;
- Raw and summarized data of dissolved oxygen concentration, salinity, temperature, pH, and tidal stage (as indicated by change

in water depth) for the continuous dissolved oxygen monitoring sites;

- Summary data collected with the sediment interface camera, including the depth of the RPD, benthic community index, surface sediment characteristics, and general description for the sedimentary environment for selected indicator testing and evaluation sites; and

- Summary data resulting from standard toxicity tests of: (1) water samples collected at the indicator testing and evaluation sampling sites, and (2) sediment samples collected at all stations.

Data will be stored in SAS data libraries by indicator and topical area (e.g., benthic species composition and biomass, contaminant concentration in fish flesh, ancillary physical/chemical data collected for each station sampling event).

4.3 The Near Coastal Data Set Index

A directory of data sets and libraries (Data Set Index) that are available at the NCIC will be developed. This index will provide users with important information about the contents of each data set. It will also describe how to access a particular data set. Information in the Data Set Index will include the following:

- A description of the data set and its purpose;

- Spatial and temporal sampling information about the collection site (e.g., length of record, geographical location);

- A list of the variables measured (e.g., salinity, sediment characteristics, numbers of benthic species, abundance of each species, biomass by major taxa, etc.);

- Name, address, and telephone number of the scientist working on EMAP-NC that is most informed about the data set;
- A description of the storage format of the data;
- An indication of whether the data is a subset/superset of other data sets (i.e., Does it belong to a particular data library?);
- The location of the data (i.e., where it physically resides);
- An assessment of the quality of the data including results of quality assurance evaluations conducted on it;
- Identification of and directions for access to other data sets that contain similar or related information; and
- Information on how to access the data set including names of contacts, approximate costs, and length of time required to access the data set.

The Data Set Index will be updated weekly. Potential users will have access to the most current version.

Historical data sets will be evaluated to determine whether they contain data of value to EMAP-NC. Those that contain useful information and are available will be incorporated into the NCIMS as data sets or data libraries. Historical data that are used frequently will be converted into SAS data sets. Information for historical data sets available from the NCIMS will be included in the Data Set Index.

4.4 Geographic Information System (GIS)

A major requirement of the Near Coastal Information Processing Center capabilities will be to create maps and perform geographically

based analyses. Therefore, the data generated for EMAP-NC will be referenced to a spatial entity, such as a latitude and longitude. Spatial analyses will be accomplished using Arc/Info, a GIS that is used throughout EPA. Arc/Info is also used by most of the other federal and state agencies participating in the EMAP program. Arc/Info is a powerful tool which includes extensive analytical capabilities and interfaces with a number of other major software products, including SAS and ERDAS (a common software tool for processing data collected by satellites). Arc/Info is not user friendly. Therefore, user friendly interfaces for routine data analysis and display will be developed by the Near Coastal Information Processing Center (assuming adequate resources and staff are available). In the absence of a user friendly interface, the requests for maps will be fulfilled by the Near Coastal Information Management Team (NCIMT). EMAP-NC data analysts will work with other data management groups within EMAP (e.g., the Las Vegas GIS group) and other agencies (e.g., NOAA) to develop standards and coverages for GIS applications. Standards will be developed for data accuracy, naming conventions, and documenting and archiving completed maps.

The initial base map for the Virginian Province will be at a scale of 1:100,000. Overlays for this base map delineating Virginian Province sampling locations, sampling plan (i.e., anticipated sampling dates, sampling crews, sample types), major road networks, and locations of facilities to which the field crew may require access will be a part of the NCIMS.

4.5 Project Management

EMAP-NC program managers need frequent and accurate status reports about field collection and laboratory processing activities. The Project Management component of the Information System is used for this purpose. The Project Management Information System has two major elements: (1) a communications system for rapidly transferring information between field crews, processing laboratories, and the NCIMS, and (2) a sample tracking system for monitoring the status of sampling

events, shipments of samples and status of analyses, on a real time basis. These two elements of the Project Management Information System are discussed below.

The computer programs that are used to communicate with field crews and sample processing laboratories have been developed by the NCIMT. This system electronically tracks the transfer of samples from the field teams to the processing laboratories. The field data systems include communications, sample tracking, and data entry. Functionally this system contains (1) navigational assistance, (2) a system for recording all events and observations made by field crews and transferring these data to the NCIMS, (3) bar code readers for rapidly and effectively entering sample identification information, (4) communications capabilities for data retrieval from a broad range of electronic data logging/recording devices, and (5) access to a data base of logistical information (e.g., boat repair facilities). In addition, by the summer of 1991, this system will conduct routine quality assurance checks (e.g., validation of station identification information) as well as provide documentation for sampling events (e.g., latitude and longitude of sampling site).

4.5.1 Communications

Field crews and processing laboratories submit data to the EMAP-NC information management center in established time frames using standard formats. The communications software described above facilitates this information exchange. For example, software in the NCIMS automatically logs remote computers into the central processing center, then performs file transfers into predetermined directories. Initial processing of the data will be initiated automatically. When processing is complete, the NCIC will be notified and requested to acknowledge that it is aware the data are ready for additional processing.

Data in the NCIMS that will be available to the field crews via the communications link includes the following:

- Logistical Information -- locations of boat ramps, overnight delivery offices, dry ice suppliers, airports, bus stations, hospitals, police stations, marinas with boat repair facilities, Coast Guard stations, motels, restaurants, gas stations, automotive repair centers, etc. This information will be available in text format and as maps processed by the GIS system.

- Sampling Locations -- including latitude and longitude, LORAN coordinates, sample identification numbers, expected sediment and water quality characteristics, estuary class, and station classification (e.g., base sampling site, indicator testing and evaluation sampling site, etc.) for each sampling location.

Some of these data are available in the prototype field systems being used in the 1990 Near Coastal Demonstration Project. All these capabilities will be available to the Near Coastal Task Group for the 1991 field sampling program (assuming adequate and timely funding).

4.5.2 Sample Tracking Information

The sample tracking system will track samples from their initial collection through completion of all analyses and/or processing. To accomplish this, each sampling event and sample type will be assigned a unique identification number. These numbers will be entered into the NCIMS prior to collection of data. Sample numbers will be bar coded to facilitate data entry by the field crews.

Information entered for each sample in the sample tracking system that are available for retrieval and review include:

- Sampling site name (cross referenced to a Station Data Base);

- The time the sample was collected including date, hour, and duration of sampling effort;

- Type of sample (e.g., grab samples to be processed for benthic species composition and biomass, fish tissue sample to be processed for contaminant concentrations);

- Identification of the individual/team that collected the sample;

- A list of the analyses and processing activities which are planned for that sample and the status of those analyses and activities (e.g., collection completed, analyses completed);

- Directions to files containing "raw" data generated for each sample (e.g., CTD profiles); and

- Directions to text files containing descriptive information about the sampling event (e.g., field team comments).

When samples are transferred from field crews to analytical laboratories, a record of the exchange is entered into the sample tracking system, both upon release and upon receipt of the materials. The identity and disposition of any sample can therefore be established by checking the sample status in the NCIMS. The status of all analyses and results will also be available through the sample tracking system. When all processing for a sample is complete, a "flag" will be set. The Demonstration Project Manager will be automatically notified via electronic mail that the sample is complete. The GIS system will be linked to the sample tracking system to display the status of sampling activities.

4.6 - Processing and storage of Indicator Data

All data received by the NCIC will be quality assured, using

procedures described in Chapter 8.0 of the NC Research Plan, and converted into SAS data sets. The data sets will be stored in data libraries by indicator type. Following initial data processing, the EMAP-NC Synthesis and Integration Team will perform the required data analyses and produce summary data bases. Examples of the types of information stored in summary data bases are dissolved oxygen summary data (e.g. percent of values below 2 mg/l for continuous dissolved oxygen monitoring stations), cumulative distribution functions for each indicator by estuarine class and for sub-populations of interest (e.g., salinity zones), and means and standard deviations for all indicators by estuarine class and sub-populations of interest. The NCIMS will maintain data and relevant analysis results in both raw and summarized form. This will eliminate costly redundant analyses.

4.7 Data Access and Transfer

The EPA VAX network will be the main means of access to data in the NCIMS. All data base design work and documentation, including the code libraries, the data dictionary, standard operating procedures for data handling, and Geographical Information System (GIS) standards and base coverages will be available over this network. Selected users who do not have access to the VAX network require appropriate authorization to access the EPA VAX network. This authorization can be granted by ADP Coordinators at each site. Once this access authorization is approved, the users will be provided direct dial access to the NCIMS. Access authorization to individual EMAP data sets for these users will be established under the direction of the Technical Director of EMAP-NC.

Data will be made available to different types of users, on different time scales, based on the quality assurance level of the data. Detailed definitions of the groups and their access rights and priorities are specified in section 6.3 of this document. Any departure from the rules specified in section 6.3 must be approved by the Technical Director and the QAO.

Ultimately the data, reports, and findings of the EMAP-NC will be important to many other groups within EMAP, the scientific community, and the general public. Data that are available to the entire EMAP community will be transferred to and maintained on the EPA National Computer Center VAX Cluster. Public access to EMAP-NC data will be through the National Computer Center.

Historical data sets or data collected by other organizations that may be important, though not likely to be used regularly, will be documented, processed, and quality assured; they will, however, not be incorporated into the data sets that are available on the VAX cluster. Data sets which are not likely to be used or that contain data that cannot be quality assured will be maintained on tape. These data will be documented but will not be made available through the NCIMS. The amount of confidential data or data for which the quality is suspect or cannot be determined that is available through the NCIMS will be limited.

All data in the NCIMS made available for general use will be in read only format, allowing users to access the data without compromising the integrity of the data base. Requests to obtain copies of or access to data in the NCIMS will be submitted to the EMAP-NC Senior Data Analyst. The EMAP-NC Senior Data Analyst in conjunction with the EMAP-NC Technical Director will develop a schedule for providing access to these data. The release schedule will depend on the availability of personnel to process the data and the urgency of the request.

PART II SYSTEM DETAILS

5.0 Introduction

This document describes a theoretical but detailed design for the EMAP Near Coastal Information System. The design is intended to be comprehensive for the needs of the 1990 Demonstration Project being conducted in the Virginian Province. Because of the dramatic time constraints available to implement this system, the design is intended to have sufficient complexity to function properly and be manageable, but at the same time be simple enough to be implemented in 10 months by a staff of five computer specialists. Additional resources are available to implement the field computer system. Specifically, the Boat System including the navigation module will be a joint development between the Near Coastal Information Management Team and Science Applications International Corporation (SAIC).

5.1 - Personnel

The following positions were filled between August 1989 and June 1990. A chart of the NCIMC staff is presented in figure 3.

Task Group Information Manager - responsible for: establishing functional requirements; system design; liaison with other ecosystems and agencies; member of IMC; development of data management standards for EMAP-NC; and ensuring that EMAP-NC adheres to standards developed by IMC.

Senior Data Librarian - Responsible for the tracking and organization of all field data; maintain reports on the status of data, and results. Predeployment - catalog all stations, events and sample numbers.

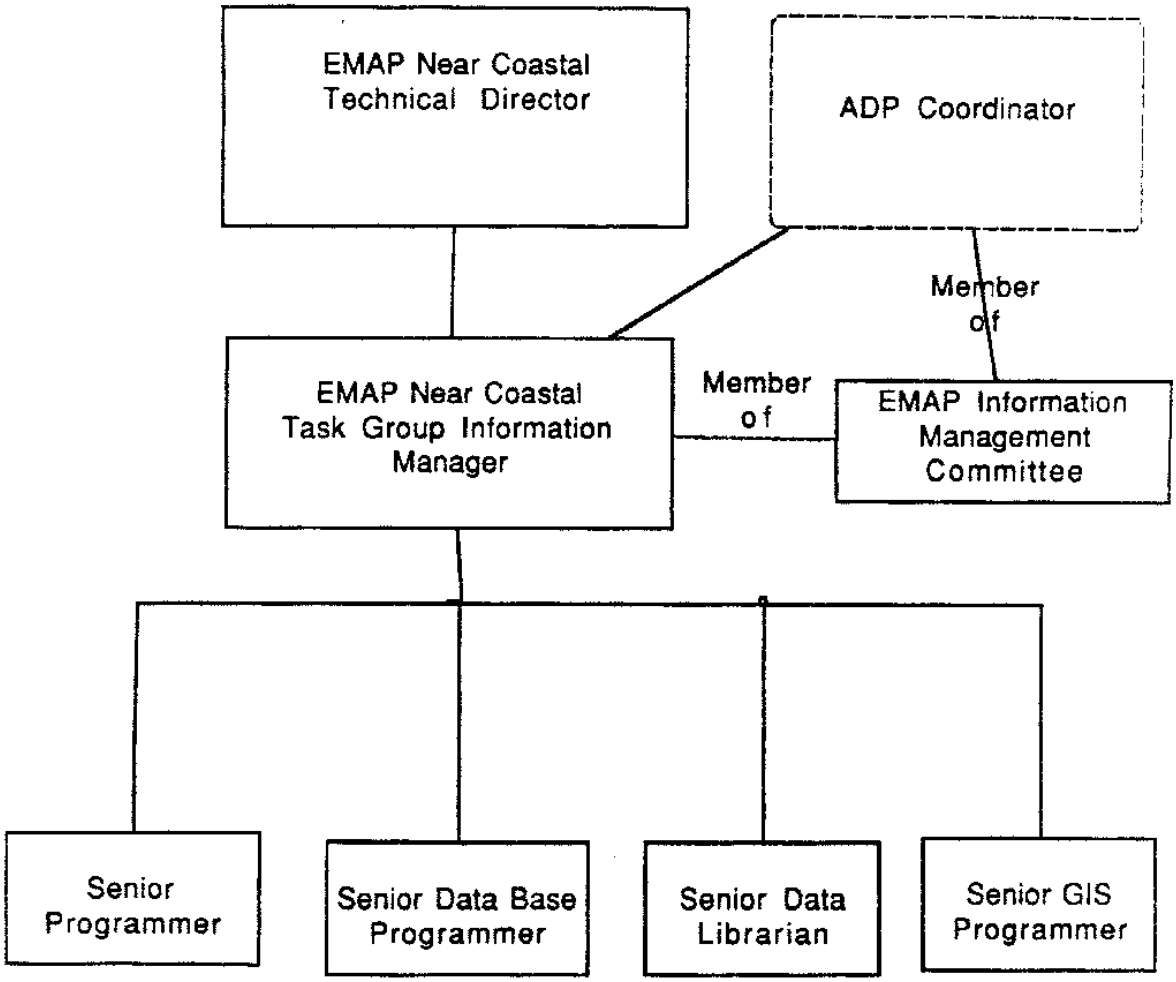
Senior Data Base Programmer - Responsible for the detailed data base design; implementation and maintenance of the data bases,

tables and interfaces necessary for the NCIS.

Senior Programmer - Responsible for the development of the field computer system; communications; the processing of all incoming data; development of data entry programs, security on the VAX information system.

Senior Geographic Information System Specialist - Responsible for design and implementation of the Near Coastal GIS. Member of the EMAP GIS team. Interact with the other members of the team to facilitate the easy conversion of logistics and results data into the GIS system.

Figure three - The Administrative structure of the Near Coastal Information Management Team for the Demonstration Project.



5.2 Major Components of the System

The table below shows the major components of the system which are addressed in this document. Each component is addressed in a separate section as noted below:

Component	Section
Quality Assurance and Control	6
Sample tracking and identification	7
Field Systems	8
Data Processing	9
Data Bases	10
Directories and file structures	11
Geographic Information Systems	12
Security	13

6.0 Quality Assurance and Quality Control Guidelines

6.1 Field Navigation and Data Logging System

Portable microcomputers modified to withstand the rigors of use on small boats represent an important component of the data management system for the Near Coastal program. The software on these machines will provide navigation and real time positioning of the boat, and control all sampling activities, sample logging, and data storage through an interactive menu. The software to be used is a modification of the Integrated Navigation and Survey System (INSS) developed by Science Applications International Corporation. The INSS is a simple, automated, menu-driven software package with complete logging facility; it has been used successfully on numerous environmental field programs during the past decade.

All data will be entered into the computer system but will also be entered on paper data forms. For the 1990 Demonstration Project, the paper forms will be considered the primary data record. The computer system will be considered a test system and a backup to the paper forms. (see appendix A.)

6.2 QA/QC for Management of Information and Data

Two general types of problems will be resolved in developing QA/QC protocols for information and data management. First, correction or removal of erroneous individual values. Second, inconsistencies that damage the integrity of the data base. The following features of the NCIS will provide a foundation for the management and quality assurance of all data collected and reported during the life of the project.

6.2.1 Standardization

A systematic numbering system will be developed for unique identification of individual samples, sampling events, stations, shipments, equipment and diskettes. The sample numbering system will contain codes which will allow the computer system to distinguish among several different sample types (e.g., actual samples, quality control samples, sample replicates, etc.). This system will be flexible enough to allow changes during the demonstration project, while maintaining a structure which allows easy comprehension of the sample type.

Clearly stated standard operating procedures will be given to the field crews with respect to the use of the field computer systems and the entry of data in the field. Contingency plans will also be stated explicitly in the event that the field systems fail.

6.2.2 Preparation of equipment and sample containers

All field equipment is prelabelled with serial numbers and equipment tracking numbers. Envelopes are prepared for each class of sampling event. Each envelope contains all the required barcode labels, data sheets and diskettes for the appropriate class of station.

6.2.3 Data entry, transcription, and transfer

To minimize the errors associated with entry and transcription of data from one medium to another, data will be captured electronically. When manual entry is required, the data will be entered twice by different data entry operators and then checked for non-matches to identify and correct errors. In many instances, the use of barcode labels will eliminate the need for manual entry of routine information.

Each group transmitting data to the NCIC will be given a separate account on the Near Coastal VAX 3300. Standard formats for data transfer will be established by NCIMT. A specific format will be developed for each file type within each discipline. If data are sent to the NCIC in formats other than those specified, the files will be deleted and the sending laboratory or agency asked to resubmit the data in the established format.

The communications protocols used to transfer data electronically will have mechanisms by which the completeness and accuracy of the transfer can be checked. In addition, the group sending the information will specify the number of bytes and file names of the transferred files. These data characteristics will be verified upon receipt of the data. If the file cannot be verified, a new file transfer will be requested. Whenever feasible, a hard copy of all data will be provided with transfer files.

The data files transmitted from the field will be fixed format text files. These files will be "parsed" by the system. The parsing process involves transferring records of similar type into files containing only those type of records. For example, observations on fish species and size will be copied from the original log file transmitted from the field to a "fish" data file. After the records have been parsed from the field log files, the individual data files will be checked automatically for erroneous values, as described in the following section. Records in the field log file which are not entered into the data base (e.g., comments in text form) will be archived for documentation or future extraction.

6.2.4 Automated data verification

Erroneous numeric data will be identified using automatic range checks and filtering algorithms. When data fall outside of an acceptable range, they will be flagged in a report for the quality assurance officer (QAO) or his designee. This type of report will be generated daily and will detail the files processed and the status of the QA checks. The report will be generated both on disk and in hardcopy for permanent filing. The QAO will review the report and release data which have passed the QA check for addition to the data base.

All identified errors must be corrected before flagged files can be added to a data base. If the QAO finds that the data check ranges are not reasonable, the values can be changed by written request. The written request will include a justification for changing the established ranges. After such changes are made, the files may be passed through the QA procedure again. In the event that the QA check identifies incorrect data, the QAO will archive the erroneous file and request that the originator correct the error and retransmit the data.

Data base entries which are in the form of codes will be compared to lists of valid values (e.g., look up tables) established by experts for specific data types. These lists of valid codes will be stored in a central

data base for easy access by data base users. When a code cannot be verified in the appropriate look up table, the observation will be flagged in the QAO report for appropriate corrective action (e.g., update of the look up table or removal of the erroneous code). If the QAO finds the need for additional codes, they can be entered by the senior data librarian.

6.2.5 Sample tracking

Samples collected in the field will be shipped to analytical laboratories. All shipping information required to adequately track the samples (sample numbers, number of containers, shipment numbers, dates, etc.) will be transmitted by phone to the NCIC at the end of each sample day, using modems built into the portable field computers. Once the field crew have transmitted the data, it will be the responsibility of the NCIMT to confirm that the samples arrived at their destination. To facilitate this, the receiving laboratories will be required, upon receipt of the samples, to record and similarly transmit all tracking information (e.g., sample identification numbers, shipment numbers and the status of the samples) to the information center, using either microcomputers or the VAX. The NCIMT will generate special programs to create fixed format records containing this information.

6.2.6 Reporting

Following analysis of the samples, summary data sets will be transmitted from the laboratories. These will include sample tracking information, results, quality assurance and quality control information, and relevant text. If the laboratory has assigned internal identification numbers to the samples, the results will include the original sample number and the internal number used by the laboratory. The analytical laboratories will be responsible for permanent archiving of all raw data used in generating the results.

6.2.7 Redundancy (Backups)

All files in the NCIS will be backed up regularly. At least one copy of the entire system will be maintained off site, to enable the information management team to reconstruct the data base in the event that one system is destroyed or incapacitated. The information on the hard drive will also be copied to diskettes at the end of each day of sampling. At the NCIC, incremental backups to removable disk will be performed on all files which have changed on a daily basis. In addition, backups of all EMAP directories and intermediate files will be performed on a weekly basis.

All original data files will be saved on-line for at least two years. After this the files will be permanently archived on floppy diskette. All original files, especially those containing the raw field data, will be protected so that they can only be read (i.e., write and delete privileges will be removed from these files).

6.2.8 Human Review

All discrepancies which are identified by the computer will be documented in hard copy. These discrepancy logs will be saved as part of the EMAP archive. All identified discrepancies will be brought to the attention of the QAO, or his designee, who will determine the appropriate corrective action to be taken. Data will not be transferred to the database until all discrepancies have been resolved by the QAO. Once data have been entered into the data base, changes will not be made without the written consent of the QAO, who will be responsible for justifying and documenting the change. A record of all changes will be entered into a data set index and kept in hard copy.

6.3 Documentation and Release of Data

Comprehensive documentation of information relevant to users of the NCIMS will be maintained and updated as necessary. Most of this

documentation will be accessible on-line in data bases which describe and interact with the system. The documentation will include a data base dictionary, access control and data base directories (including directory structures), code tables, and continuously-updated information on field sampling events, sample tracking, and data availability.

A limited number of personnel will be authorized to make changes to the Near Coastal data base. All changes will be carefully documented and controlled by the senior data librarian. Data bases which are accessible to outside authorized users will be available in "read only" form. Access to data by unauthorized users will be limited through the use of standard DEC VAX security procedures. Information on access rights to all EMAP-NC directories, files and data bases will be provided to all potential users.

The release of data from the NCIS will occur on a graduated schedule. Different classes of users will be given access to the data only after it reaches a specified level of quality assurance. Each group will use the data on a restricted basis, under explicit agreements with the Near Coastal Task group.
The following four groups are defined for access to data:

I. The Near Coastal central group including the information management team, the field coordinator, the logistics coordinator, the demonstration project coordinator, the QA officer and the field crew chiefs.

II. Near Coastal primary users - ERLN, prime contractors (e.g. VERSAR and SAIC), Gulf Breeze Personnel, NOAA Near Coastal EMAP personnel, EMAP quality assurance personnel.

III. EMAP data users - All other task groups within EPA, NOAA, and other federal agencies.

IV. General Public - University personnel, Other EPA offices (includes

regional offices), other federal, state and local governments.

The following table summarizes the policy of the Near Coastal task group with respect to the distribution of data. The Roman numerals in the table refer to the groups listed above.

Synthesis level		No QA/QC 1	Machine QA/QC 2	Human QA/QC 3	Technical data analysis 4
RAW	A	I *	I,II*	I,II,III*	I-IV
FIRST SUMMARY	B	I *	I,II*	I,II,III*	I-IV
FINAL SUMMARY	C	I *	I,II,III*	I,II,III*	I-IV

* = Explicit restrictions on the uses and dissemination of the data must be made and agreed to by all participants in these groups.

Requests for premature release of data will be submitted to the information management team. The Task Group Information Manager and the QAO will determine if the data can be released. The final authority on the release of all data, however, is the technical director of EMAP Near Coastal.

The long term goal for the Near Coastal information management team will be to develop a user interface through which all data will be accessed. This will improve control of security and monitoring of access to the data. The user interface will also help ensure that the proper data files are being accessed.

7.0 Sample tracking and identification

A key function which must be met by the EMAP NCIMS is the tracking of samples, shipments and results. Samples and results will originate in the field. The results will be transferred to the Near Coastal Field Operations Center (NC-FOC) along with the sample identification information for the physical samples collected. Samples will be shipped to appropriate laboratories according to procedures outlined in the EMAP Near Coastal Research plan. To ensure unbiased processing of samples, only the SAMPLEID will be sent with each sample. The identification and the status of the samples, and the contents of each shipment, will be tracked in the data base system. The laboratories doing the analyses will return their results to the NCIC with the appropriate sample number. Results will be matched with the appropriate sample identification information by cross referencing the SAMPLEID numbers of the samples with those found in the results data files.

7.1 Primary identification of samples

The following variables will be used to identify sampling events, samples and shipments:

STATION - The station ID is the prime variable which will be used to relationally couple all appropriate output data.

DATE - The date, as YYMMDD, that the sample was collected.

EVENT - Each visit to a station on a given date will be defined as an event.

SAMPLE IDENTIFICATION (SAMPLEID) - Each sample generated and each measurement made during an event will be assigned a specific SAMPLEID. A portion of the SAMPLEID will contain the event number.

DISKETTE (DISKID) - Each diskette used in the field for data collection

will be given a unique sample number.

SHIPMENT - Since shipments will include samples from a number of different events, each shipment will be assigned a unique number.

FILENAME - The results of some field activities will be data files. Each of these files will be assigned a file name and will be associated with a particular date or a particular event.

7.2 Sample and shipment numbering system

7.2.1 Station Identification

All stations which are being sampled in the Virginian Province have been assigned station names. The name is constructed by arbitrarily numbering the stations from 1 to 259. Each station number is then prefixed with "VA90-". Therefore, the Station Identifications go from VA90-001 --> VA90-259. Each station identification is associated with a latitude, a longitude, and a series of codes identifying the station classification. The station classification codes and the definition of each code can be found in appendix I.

7.2.2 Event Numbers

Unique Event Numbers have been assigned to each Station Class during each Interval, to Quality Assurance/Quality Control Samples and Extra Events (Table 1). The Station Class Event Numbers range from 1 - 249 and 320 - 694. The Event Numbers for the Quality Assurance/Quality Control Sediment Samples range from 250 - 319.

7.2.3 SAMPLEID Numbering

EMAP Near Coastal Samples are numbered with SAMPLEID's which consist of four to six digits. Each SAMPLEID is composed of an event number and a sample number. An event is defined as a visit to a station on

a specific date. Each visit to a station must have a unique number associated with it because stations are visited more than once during the three Intervals. The event numbers range from one (1) to three (3) digits and the sample numbers are consistently three (3) digits. Event numbers always precede the sample number when forming a SAMPLEID. Event numbers range from 1 - 799, while the sample numbers range from 000 to 549. For example if we were sampling EVENT number 131 all samples from this event would fall in the range 131000 - 131549, where 131 is the event number, 000 - 549 are the sample numbers.

Each sample which is going to be collected has been assigned a unique sample number (Table 2). This sample number will always be associated with the same sample during each event. Sample numbers for activities on the boat range from 000 to 099. Sample Numbers 100 - 149 have been assigned to fish chemistry samples for fish trawl 1. Sample numbers 150 - 299 have been assigned to fish pathology samples for fish trawl 1. Sample numbers 300 - 349 have been assigned to fish chemistry samples for fish trawl 2. Sample numbers 350 - 499 have been assigned to fish pathology samples for fish trawl 2. Sample numbers 500 - 549 have been assigned to bivalve chemistry samples.

Sample numbers are assigned primarily to samples which must be tracked physically. These types of samples were assigned sample numbers first (000 - 026) because they require barcode labels associated with them. The barcode labels associated with an event had to be sequential for printing purposes.

Some Sample Numbers, however, are associated with activities which have other data related to them. These include sample numbers associated with CTD and Hydrolab data retrieval, fish trawls and bivalve tows. These activities did not require a sample number for each piece of data associated with it. The CTD and Hydrolab data consist of DO, salinity, temperature and pH measurements, among others. Individual length measurements are associated with both the fish trawl and bivalve tow sample numbers.

For each event there is a unique range of SAMPLEID's composed of an event number and a sample number (i.e., 185000 - 185013). These SAMPLEID's identify the samples and data associated with the event (Appendix A - Table 3.). The length of the sequence of SAMPLEID's associated with an event depends on number of activities performed during the visit to the station. Because the activities performed at a Station vary from Interval to Interval, some events will have more SAMPLEID's associated with them than others. Also, the SAMPLEID's associated with an event will not necessarily be in sequence as the SAMPLEID's assigned to an event are dependent on the activities performed at the station and if the samples must be tracked physically.

7.2.4 Blind Sample Numbering

According to the Quality Assurance Quality Control Plan established for the 1990 Demonstration Project, at 10% of the stations sampled, quality control samples will be taken. These samples must be unidentifiable to the analytical laboratories which are analyzing them. The laboratory must not be aware of the station from which they were sampled and they must not know that these are quality control samples. To ensure this, a separate numbering scheme has been developed for these samples. In the interest of maintaining the blind nature of these samples the scheme will not be presented here.

7.2.5 Shipment and Equipment Barcode Numbers

Barcode numbers have also been assigned for equipment and shipment of samples. The equipment barcode numbers range from 810000 to 810199. Equipment barcodes were attached to CTDs and Hydrolabs. The barcode labels for the shipment of samples will be used to track packages of multiple samples shipped from the field to the laboratory performing the analyses. These numbers range from 820000 to 820999.

7.2.6 Forms

Forms for all sampling activities have been created for the 1990 Demonstration Project. Each form corresponds to a specific activity and falls into one of three general categories: Reconnaissance, Samples, or Shipment. Each form contains data specific to that activity which will eventually be entered into appropriate SAS data sets. All sample identification and tracking are recorded on data sheets. This information is also entered into the field computer system where appropriate. Examples of the forms and the documentation associated with each particular form are located in Appendix A.

7.2.7 Distribution

Barcodes and forms for each interval are distributed to field crews packaged in envelopes marked with the station class, event number and interval. The appropriate number of envelopes for each interval were made up according to the Station Classes visited during an Interval. Each station class involves a subset of all the activities which can occur at a particular station. The contents of each envelope reflects the activities which should be accomplished for a particular station class during each interval. The contents include the required labels, data forms and diskettes based on the station classification.

The Crews are sent out with the proper number of envelopes for each station class in each interval. Each crew is also given additional envelopes to cover any changes in plans. At each sampling event in the field, the crew randomly selects an envelope for that particular Station Class and thereby, assigns an Event number. The Event number will be related to a Station on a particular date in the EMAP Near Coastal Data Base.

7.2.8 Diskettes

One disk will be used per day to enter data into the field computers. This disk will contain data from one to several events. The same disk will be used the next day in the mobile laboratory to record data from the events sampled on the boat the previous day. Therefore, all the data from an event will be on the same disk. Each diskette is given a unique six digit number which is used to track the diskette and the events on the diskette.

7.2.9 Filenames

Specific file names will be assigned to files on both the field computers and on the NCIC VAX. The location and name of each file will determine what event and parameters it is associated with. Details of these structures and naming conventions can be found in Appendix B.

7.3 Data Form Tracking

After a crew finishes a five day period in the field, data sheets are returned to the EMAP NCIC. A Tracking number is assigned each set of data sheets completed for an event. The number is written in the upper right hand corner of each data sheet.

The data sheets are assembled in chronological order and assigned a three part tracking number. The first segment of the number is a sequential number, beginning with one (1) and prefaced by an F (Form). The first segment of the tracking number is assigned arbitrarily, but generally indicates the chronological order in which the Events were conducted. The second segment is the crew number (1, 2 or 3) and the third segment is the Interval number (1,2 or 3). Segments are separated by dashes. As more data sheets arrive from one crew, the tracking number starts with "n + 1", "n" being the last number assigned to a set which came in previously. Thus, the code F3-1-1 indicates the third set of data sheets (first segment) received from Crew 1 (second segment) in Interval 1 (third segment); F12-3-2 indicates the twelfth set of data sheets

received from Crew 3 in Interval 2. The first segment also generally indicates that the Data Sheets are from the third or twelfth Event conducted by the crew.

Information from each set of forms relating to an event is recorded on two data sheets. A data sheet tracking form was created to record the form number, the date the forms were received, the event number, the event date, the station sampled, the station class and the crew number. A decision is made on whether or not all pieces of information expected from an Event were collected and comments are entered. There is also space on the form for cross-referencing the information on the form to a specific disk. A data sheet information sheet was also created to record the specific types of information collected during the event. On this sheet, the form number, the event date and the date the forms were received is repeated for cross reference.

This system not only logs in the data sheets, but facilitates finding data sheets from a specific event, station or completed on a specific date. Once data sheets are assigned a tracking number, they are divided by Crew and placed into folders. If information is sought for a specific station, the station number can be found on the data sheet tracking form and related to the form number written on the data sheets. The forms can then be easily located by finding the forms in one of the Crew folders.

7.4 Diskette Tracking

A similar system has been developed for tracking disks containing information from events. As diskettes arrive from the field, the disks are assigned a three segment tracking number. The first segment is a sequential number, starting with 1 and prefaced by a D (Disk). These numbers are arbitrarily assigned to a disk and are not related directly to the event date as is generally the case with the forms. The second and third segments are the same as that of the forms, crew number and Interval, respectively. When disks arrive, the directory is scanned to

obtain a list of stations whose information is on the disk. The tracking number, disk number, stations and events on the disk are recorded on a label attached to the disk.

A disk tracking form is also created to record pertinent information contained on the disk. This information is the same as that recorded for data sheets, except that a disk number and a disk received date is recorded first and eventually a form number is recorded which corresponds to the event and station. Also, most disks have multiple entries, indicating that information from more than one event is on the disk. The disks will be scanned for content and the types of files present recorded on a disk information data sheet. The disks are also separated by crew and field in diskette boxes in order of the disk tracking number.

8.0 The Field Computer System

The field computer system is designed to facilitate the collection and organization of the tracking information, results gathered in the field, observations, and communications.

8.1 Hardware

Each of four 24-foot boats will be equipped with a 386 laptop computer which is designed to withstand the rigors of field work. The purpose of this computer will be to interface with navigational instruments (Loran C and GPS) and data collection instruments (CTD and Hydrolab Oxygen meters), and for logging and interfacing data and station position information. Additionally, each field team will have a backup computer based in its mobile laboratory. The Mobile lab computer will be used for data entry, calibration, down loading of field data and communications with the NCIC.

The computer chosen for this application is the GRiDCase model 1530. This model has a 386 processor. It will be equipped with 2 Mb RAM,

a 2400 baud modem, a 1.4 Mb floppy drive, a 40 Mb hard drive, DOS 3.3 (both in ROM and on the hard drive), an 80387 math co-processor and an expansion cartridge providing an additional two serial ports (for a total of three serial ports). A portable printer will also be on-board (Diconis).

8.2 Functional Components

The requirements discussed below will be built as application programs on the field computers. The people in the field will not be computer experts and their access to the software needs to be as simple as possible. The systems which will be developed include:

- o Boat system
- o Shore system
- o Shipment system
- o Communications System.

8.3 User Interface

The entire field computer system will be run from a common interface. The interface will be menu driven and will contain menu choices which will facilitate all of the computer work which must be done by the field crews. The interface was developed using the C programming language and it runs under the DOS operating system. Documentation of the interface and all field programs can be found in Appendices C, D, and E. The detailed components of the field system are described below.

8.4 Boat System

The boat system consists of all the programs which are run on the EMAP-NC research vessels. Many of these programs can also be run from the Shore system to allow the crews to perform data entry, down loading of electronic data and sample tracking on shore when conditions are not optimal for computer work on-board.

8.4.1 Calibration and daily setup

In most cases data collected aboard the research vessels will be relative to a sampling event. Each day, however, a number of activities occur which are pertinent to all the activities performed that day. This includes calibration of the Loran unit and the CTD unit. These data are transferred to directories which associate the calibration information with a particular instrument and date (see appendix B for details of the directory structures).

8.4.2 Navigation system

The navigation system helps the crew locate the station which will be sampled. Once at the station the system stores the coordinates of the station and the route taken to get to the station (as Way points). The latitude and longitude of each station is stored in a data base which can be recalled at a later date. When a crew must return to a station which it has already sampled the system will direct the pilot to the station.

The navigation system contains modules for calibrating all navigational equipment, storing date, time and locational fixes, uploading way points to the navigational instruments, navigating to station, and calculation of error on the calculated latitude and longitude coordinates. All outputs from the navigation system are sent to either the stations data base (stored in the boat directory) or to the field LOG file which is described in greater detail below.

8.4.3 Logging System

All the software modules which are part of the boat system write records to a LOG file. The structure of the LOG file is given in appendix E. The LOG file is stored in a subdirectory which is created at the time of sampling. The subdirectory is named according to the station and the date, eg. "C:\VA90-168\900724\" is the subdirectory for the samples collected on July 24, 1990 at station VA90-168. The log file is named

with a time stamp indicating when the file was opened. The form of the file is MMDDHHMM.LOG, e.g. the file C:\VA90-168\900724\07241312.LOG, is the log file for station VA90-168, sampled on July 24, 1990, opened at 1:12 p.m.

The entries in the LOG file indicate what activities were done at each station on each date. Each line in the LOG file indicates at what location and time it was written by the software. The formats for the records depend on the activity to which they pertain. For each activity the output record is in a predetermined, fixed format. The user activates an editor which allows them to enter a comment record into the LOG file at any point during the sampling process. These entries, accompanied by a location and time fix, are stored directly in the log. These observations can include weather, sea conditions, or anything the crew member wishes to include in the station information file. The user is able to flag any record in the log file if the entry is in error, questionable or if it fails a QA/QC check. When the data are reviewed at the NCIC, a flag is raised and reference made to the log entries containing crew flagged information.

A separate "maintenance log" is maintained in the subdirectory C:\EQUIP. This contains information about general boat and vehicle maintenance as well as equipment maintenance. This log is maintained on the hard drive and on the floppy drive for redundancy.

8.4.4 CTD system

The water column is profiled using a SeaBird model 25 Sealogger CTD. This instrument measures salinity, temperature, depth, DO, pH, transmissometry, fluorometry and PAR (photosynthetically active radiation). Data are recorded as voltage and then converted, using SeaBird's software, to appropriate units. Transmissometry and fluorometry values remain as voltage and are post-calibrated following determination of suspended solids and chlorophyll 'a' in filtered water samples.

From the main menu the user selects "CTD". The computer asks for calibration/verification information. The CTD will be deployed off the side of the boat, just below the surface and data will be recorded. At the same time, from the same depth, water samples are collected. Temperature (thermometer), salinity (refractometer) and DO (micro Winkler) are determined. The observed quality control data are entered into the computer. Data from the CTD are uploaded into the computer and compared to the quality control values. Water samples are taken for suspended solids and chlorophyll 'a'. Samples are filtered on-board. The computer requests a sample number for these samples. All QA/QC data and sample tracking information entered into the computer are written to fixed format records in the LOG file.

Following verification/calibration, the CTD unit is dropped through the water column while recording data. The unit is then connected to the computer's serial port and the data file is uploaded using the SeaBird's software. The data file is then reviewed by the crew to ensure that the data look reasonable and that the Oxygen does not go below 0.5 mg/liter. Boat personnel will need this for a decision making process described below. Following viewing, confirmation that the cast is being accepted will be entered and recorded in the LOG file. The data files are stored on both the hard drive and the floppy drive under the file specification: \STATION\YYMMDD\MMDDHHMM.CTD. This file specification is entered into the LOG file and given a tracking SAMPLEID. The file specification will direct the NCIMT to download this file and associate it with the proper event number.

8.4.5 Dissolved Oxygen System

During the 1990 Demonstration Project, dissolved oxygen data will be collected using Hydrolab DataSonde III units. These units will collect data for a ten day period. Each ten day period will involve deploying, retrieving, maintaining, and downloading data from the oxygen meters. Each meter will have to pass a quality control check at the beginning of

the deployment and at the end of the deployment. Each task involves different processes and therefore, the recording and storing of different types of data.

One Hydrolab will be deployed at each of 30 Long Term Dissolved Oxygen (LTDO) stations. These stations will be revisited every ten days. At each visit the meter which was deployed at the station will be retrieved. Its data will be downloaded, saved and analyzed. Depending on the general characteristics of the data, another unit will be deployed or the LTDO station will no longer have a Hydrolab deployed.

8.4.5.1 Hydrolab Calibration

These units will be calibrated the day before deployment. The crews confirm that the calibrations are done and stable.

8.4.5.2 Hydrolab deployment

At each station one Hydrolab is deployed at one meter above the bottom. In the event that the CTD cast shows that the bottom water is 0.0 mg/l DO, the unit is not deployed. This is noted in the DO subsystem and output as a record in the LOG file. The serial number and the EPA property tag number of the meter being retrieved is entered into the computer system. These numbers are output into the LOG. A series of quality control checks are performed for dissolved oxygen, temperature, and salinity. The results of these checks are entered into the computer system and output to the LOG file for the event.

Communications with the hydrolab unit are done using PROCOMM Plus communications software. The hydrolab meter is programmed to collect data with a fixed starting time and sampling interval. The meter is then lowered over the side. The user can enter comments, observations, etc. using the comment log described above.

8.4.5.3 Hydrolab Retrieval and Maintenance

The LTDO Stations are revisited every ten days. The data are downloaded and the unit is cleaned and recalibrated after each visit.

For a maintenance visit, the crew registers on-site and records comments on the condition of the mooring. The Hydrolab is then recovered and a general comment on the condition of the instrument is made. The computer prompts for:

1 - extent of fouling,
2 - Condition of the membrane
3 - Condition of the connection to the mooring.

The computer prompts for an instrument Serial number and an EPA property tag number. These can either be entered manually or read in using the barcode reader. These numbers are output to the LOG file, accompanied by a FIX. The actual data file which is transferred from the Hydrolab is transferred to the appropriate station and date subdirectory where it is given a file name composed of a time stamp and an extension of either DOM (Dissolved Oxygen Monitoring) or DOA (Dissolved Oxygen ASCII). For example C:\VA90-168\900724\07241420.DOM, is the DOM file for the hydrolab unit retrieved at station VA90-168, on July 24, 1990, file opened at 2:20 p.m.

8.4.5.4 Data evaluation

At each maintenance visit, the data will be downloaded from the instrument which has been previously deployed, and evaluated. These data will have to be plotted as O2 concentrations versus time. Any questionable observations are entered by the field crew into the LOG file.

8.4.6 Fish Trawl System

The crew enters a starting point and an ending point. The navigation system takes a time and a location FIX every 30 seconds during the trawl. At the end of each trawl the computer calculates the distance traveled and the length of the trawl. These data are output to the LOG file. The computer also asks the user for the average boat speed as recorded by the boat's Knot meter. This is also be sent to the LOG file. The crew can hit a function key which cancels a trawl and resets the system. When this key is activated the data from the original aborted trawl is maintained but a separate code is affixed to the records denoting that it is an aborted trawl. Following the recovery of the net the crew begins identifying, counting and sizing organisms. All data recording is done on log sheets. The data management system is being modeled after the NOAA National Marine Fisheries Service data system used for the Virginian Province. The crew records the species found, the number per species, the lengths of a subsample of fish, and any gross pathology observed. A hand written sample number is included in the sample bag with for fish with gross pathology. This sample number is also recorded with the species and size of the fish on the data sheet. All Data are entered into the log file by the crew member in the field laboratory the next day.
All fish data are written to and reside in the appropriate LOG file.

8.4.7 Bivalve Dredging

A bivalve dredge will be towed retrieving bivalves greater than 2 centimeters. The bivalve data will be recorded on paper data forms. The data entry into the log will occur in the mobile lab on the second day using the shore system. The data entered will include the EVENT NUMBER and a species code. The computer will then prompt the user for the measurements made on each bivalve. This process will be repeated for each species caught in the dredge. Each individual will be output as a single record in the appropriate LOG.

8.4.8 Benthic biota

Three grab samples are taken for species enumeration and biomass determination. Each of these samples has a 1" core removed for grain size analyses prior to the removal of the sample from the grab.

To record the samples taken the crew enters a return to indicate the location and time when the core is taken. For all cores which are accepted, the computer prompts the user for water depth, Redox potential depth, and a comment. A record is written to the LOG file for all cores taken regardless of whether they are accepted or not. Only the cores which are accepted have complete data records written to the written the LOG file. The record includes the sample number, a FIX, the water depth and comment.

8.4.9 Sediment chemistry and toxicity

Additional grabs are taken for sediment chemistry and toxicity testing. Multiple grabs will be needed to include sufficient sediment for one sample.

Similar to the system for the Benthic Biota samples described above, when a sample is taken, the carriage return is pressed and a record of the coring attempt is written to the log file. Each grab is recorded separately. Once all samples which are to be collected for a single composite have been homogenized, the crew enters sample numbers for chemistry and for toxicity. The system allows comments to be entered with each SAMPLEID. The SAMPLEID and the comment are written to the LOG file.

8.4.10 Additional modules

The interface is written to accommodate further development including the addition and deletion of indicator suites. Additional modules

exist for additional sample identification information, e.g to perform REMOTS camera profiling. As indicator suites change the system will be modified to meet the needs of EMAP-NC. Similarly, as the Near Coastal task group moves to other provinces the system will be modified to meet the specific needs of each province while maintaining a central look, feel and approach.

8.5 The Shore system

On the day following a field sampling event, the crew in the mobile lab performs fish dissection, data entry, shipment of samples, calibration communications and data backups. Each task which is performed in the mobile laboratory is addressed in a module of the field computer system. The shore system creates a log file for each day. This Lab log file is stored in a subdirectory C:\LAB\yymmdd\mmddhhmm.LAB, where yymmdd = the date the work was performed in the mobile lab, mmddhhmm = the month, day, hour, minute that the file was created. Each record in the LAB LOG file contains the event number from which the sample was taken.

8.5.1 Instrument calibration

The shore system contains modules to calibrate the CTD units and the Hydrolabs. These systems establish connections with the instruments and then transfer real time data to the computer's screen. The only information written to the LAB LOG file is an indication of whether the instrument was successfully calibrated along with its Serial number an EPA tracking number.

8.5.2 Boat system capabilities

All programs which are accessible in the boat portion of the system can also be accessed in the mobile laboratory. This enables the crew to catch up on activities which could not be accomplished on the previous day

such as data entry, file transfers, etc.

8.5.3 Backups and Software updates

The Interface contains an entire section for the backup of data located on the hard disk. All data collected in the field and generated in the mobile laboratory are written to both the hard drive and to a floppy diskette. The floppy diskettes are the principle electronic record for each event. The copies on the hard drive are backups to the floppy records. The backup capabilities accessible within the interface facilitate:

- o additional backups of the data
- o transfer of data between the boat computer and the mobile lab computer
- o loading of software modifications sent to the field crews from the NCIC

8.6 Shipment Systems

Samples are shipped from the field teams to laboratories for archival, processing, and analysis. A record of the shipment and the receipt of samples is maintained in the NCIMS. The system includes the date and time the sample was sent and its condition on receipt. This allows the field coordinator to determine if the samples arrived in an acceptable condition for proper analysis. Samples which are lost or destroyed are noted in the tracking data bases to ensure that results are not expected for these samples. Samples which arrive in a condition which is questionable are flagged so that their results can be qualified.

8.6.1 From the field laboratories to the remote laboratories

When processing of all samples from the previous day is complete, the crew prepares shipment cases for those sample types with a sufficient number of samples to warrant a shipment. The shipment module prompts for a shipment number and a carrier tracking number. Each SAMPLEID is then entered (using the bar code reader). A fixed format

shipment file is created detailing when the sample was sent, who packed it, the carrier, all tracking information, and a list of the sample numbers included in the sample. This information is output to the LAB LOG.

8.6.2 From the remote laboratories to the NCIC

A program has been written and is being distributed which will allow the laboratories receiving samples to enter the same information entered by the field crew. In addition, the program will request the input of a condition for each sample. The program outputs a file in a fixed format similar to the file sent up by the field crew. This file can then be sent via electronic mail to the NCIC.

8.7 Communications Systems

The communication system is designed to facilitate two way communications between the field crew and the NC-FOC personnel. The communications system allows the crew members to set the computer up in a motel room and put it into a program mode. The program mode waits until a preestablished time and then automatically dials the NCIC, connects and initiates the transfer of the LOG's and messages from the Field computer to the NCIC VAX3300. When this transmission is complete the VAX initiates the transfer of messages and software updates to the field computer.

8.8 Directory and file structures for the field systems

Detailed explanations of the directories and the files in the directories can be found in Appendix B.

9.0 Processing of Incoming Data

Data will be arriving at the NCIC in a number of different forms, via a number of different avenues. The data forms include paper data forms, electronic files, magnetic media and voice communications. The different avenues include the telephone lines, electronic mail, US mail, private carriers, and hand carried by EMAP-NC personnel. All the data must be logged, tracked and eventually entered into the NCIMS. Details of the tracking of data sheets and diskettes have been discussed in sections 7.3 and 7.4.

9.1 Data Entry

All data which is received on paper forms which has not been entered into electronic files will be manually entered using programs which are similar to the programs being used in the field. These programs will create files similar to those created in the field. All manually entered data will be compared to the original data sheets. The QA Officer will personally check 10% of the data entered by each data entry person to ensure proper data entry. All entered data will be processed along with the data transmitted from the field.

9.2 Uploading Electronic meter data

All CTD data files and Hydrolab data files will be delivered to the NCIC on diskettes. These files will be uploaded to the VAX for initial QA/QC validation and will be archived in fixed directories (see section 11 - directories and file structures).

9.3 Processing and Quality Assurance of Field Data

All data files collected in the field or entered manually from the field data forms will pass through a series of programs which will check the completeness and the validity of the data. These processing programs will produce:

- Reports indicating what has been accomplished and what communications have occurred.

- Reports indicating potential problems in the data which will require human intervention.

- Structured data files which will be used as input to the main NC Data Bases.

These programs will also transfer data from processing directories to archive directories where the data files will be kept and maintained on the VAX in the NCIMS. The flows of data are summarized pictorially in Appendix F.

All data files will be partitioned into fixed format data files in which similar data is coalesced and prepared (parsed) for entry into the NCIMS. For example, all the quality control water temperature will be copied from the field LOG to a file containing only the quality control water temperatures. The output file containing these data will then be evaluated to ensure that the maximum and minimum values for water temperature are reasonable. If they are reasonable then the file is loaded into the appropriate data base. If problems arise in the Quality assurance data check then the file is stored in a temporary directory where it waits for human intervention to determine the next step.

10.0 Data Bases

10.1 Introduction

The EMAP NC data base system will be constructed using the Statistical Analysis System (SAS). SAS has adequate capabilities to simulate a relational data base system with sufficient flexibility to deal with the 1990 NC Demonstration Project. A schematic of the overall data base design is presented in Figure 4. It illustrates the general approach which is being taken. At the heart of the system are the tracking data bases and the identification data bases. The identification data bases define the sampling units, the stations, and the individual sampling events. The tracking data bases track individual samples, measurements or data files from conception in an experimental design, to final analysis, reporting and data archiving. These data bases are explained in great detail below. The conceptual structure also indicates the exchange of data with other agencies and into other information dispersal systems, e.g. ODES, COMPAS and the NASA Data Directory.

10.2 - General Principles

This design attempts to adhere to standards established by the Narragansett Bay Project Data Management Team. Two major components of the system are modeled after major NOAA programs with extensive data bases. For the results of analytical chemistry, the structures used by the NOAA Status and Trends Program is adhered to as closely as possible. The fish population data and the pathology data have been modelled after the Northeast National Marine Fisheries service data base. There will be instances in which the EMAP NC system requirements are incompatible or the established standards are not applicable. In these cases the designs have been modified to meet the requirements of EMAP-NC. Where established standards are not followed the intention is to begin resolving the inconsistencies during FY91.

The data base structure is based on a sample tracking system which allows cross referencing of all samples with the stations and the dates

that they were taken. The numbering systems are outlined in section 7.2. By cross referencing the different data bases using the key variables SAMPLEID, CHEMID, EVENT, and SHIPMENT, all results can be logically linked. These linkages can be used to create a logical data base which contains all the information pertinent to a given sampling event. Figure 5 presents a gross diagram of the structure and relationships of the different data bases.

Figure 4 -

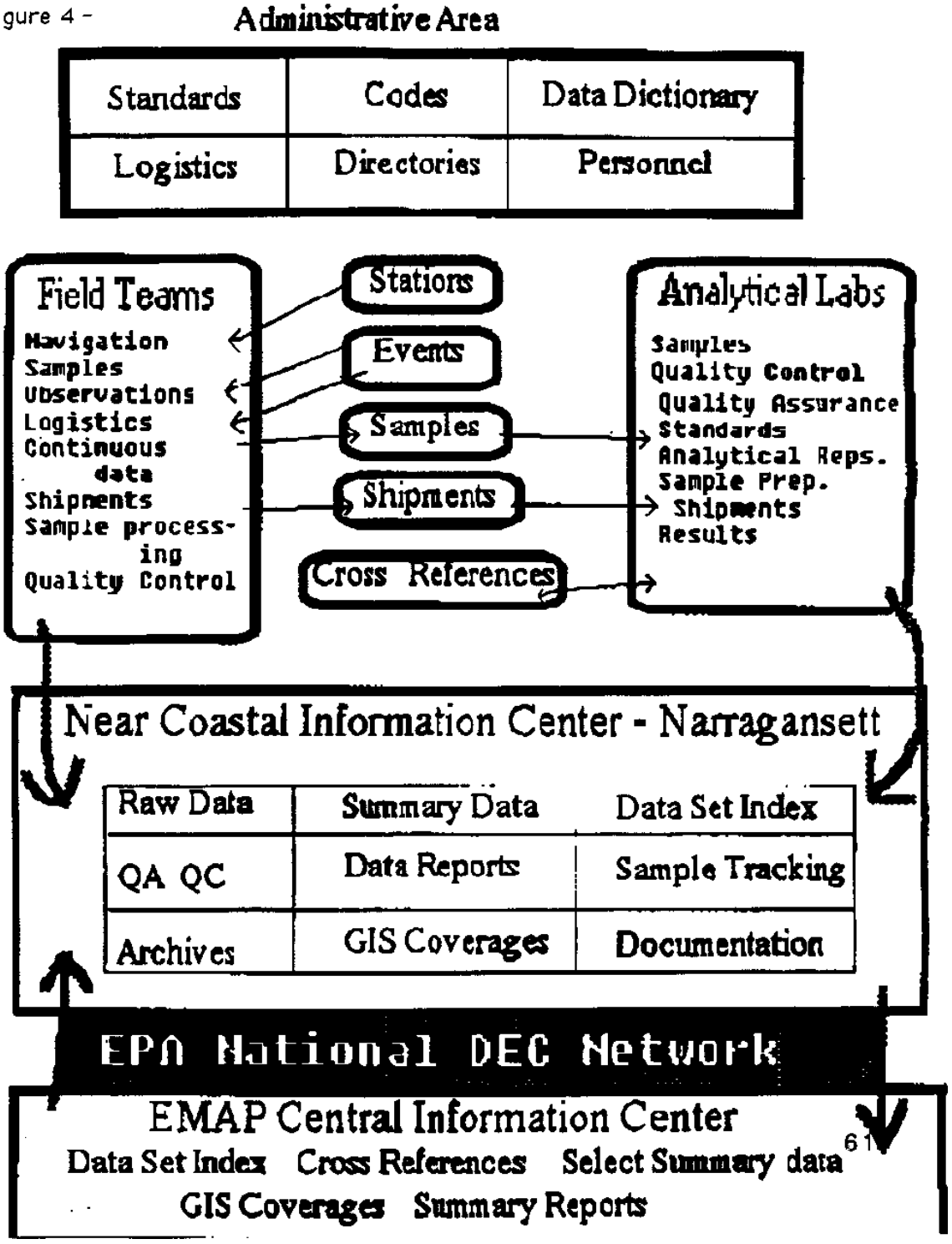
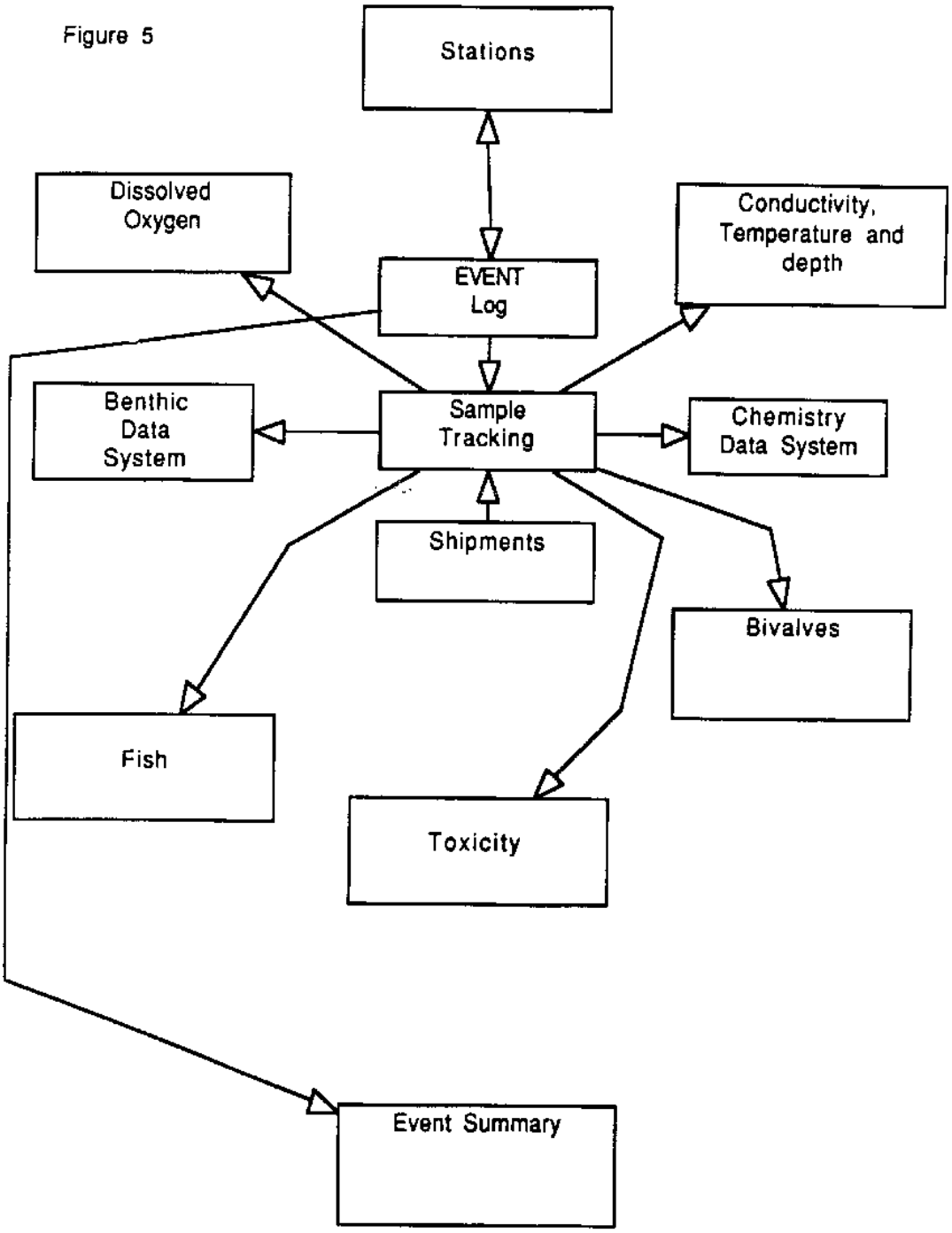

Administrative Area
Standards
Codes
Data Dictionary
Logistics
Directories
Personnel
Field Teams
Navigation
Samples
Observations
Logistics
Continuous data
Shipments
Sample processing
Quality Control
Stations
Events
Samples
Shipments
Cross References
Analytical Labs
Samples
Quality Control
Quality Assurance
Standards
Analytical Reps.
Sample Prep.
Shipments
Results
Near Coastal Information Center - Narragansett
Raw Data
Summary Data
Data Set Index
QA QC
Data Reports
Sample Tracking
Archives
GIS Coverages
Documentation
EPA National DEC Network
EMAP Central Information Center
Data Set Index
Cross References
Select Summary data
GIS Coverages
Summary Reports

Figure 5


Stations
Dissolved Oxygen
Conductivity, Temperature and depth
EVENT Log
Benthic Data System
Sample Tracking
Chemistry Data System
Shipments
Bivalves
Fish
Toxicity
Event Summary

This design uses codes as infrequently as possible. Wherever reasonable, full descriptive words are used in place of codes. Codes are used for quality assurance purposes and ease of data handling (e.g. in the field) in certain instances. These include very long names (e.g., taxonomy, chemistry), and existence of universally accepted codes (e.g., State names).

The final SAS data sets for the EMAP-NC data base management system will be stored in four SAS libraries, each of which resides in a separate subdirectory. The four libraries are **MAS**: the Master Library, **ONL**: the Online Library, **RES**: the Results Library, and **SYN**: the Synthesized Data Library. The Master Library MAS holds Station data and Station and Event Classification data. These data will remain relatively constant throughout the data collection phase and will only be updated occasionally. These data sets are central to the database design, and will be merged with other data sets very frequently. For these reasons they reside in the Master Library.

The Online Library "ONL" is used for tracking and dictionary information. These datasets must be kept as current as possible, and will be updated frequently. Data in this library will not be released to the public; they are for in-house tracking only. The Results Library holds data elements that are collected on field data sheets and from analytical labs. These data sets will be updated as new data is collected. Data sets in this library will be updated after QA/QC checks have been run on new data. The Synthesized Data Library "SYN" will hole summary data sets derived from the data stored in the Results Library. These datasets will be designed for specific analytical purposes.

10.2.1 User Interface to the NC data bases

An additional directory called **PRODUCTION** will contain programs which are the interface to the SAS data management system. These programs will allow standard queries of the data base and will also process standard requests for data. All data bases will be accessible to

experienced SAS users directly throgh SAS.

10.3 - Standard Codes

All established codes will be stored and retrieved from a directory of CODES. These codes will reside in

EMAP$:[SASDATA.MASTER.LOOKUP] = CODES$

These data bases of codes will be used to quality control incoming data and to resolve codes for reporting purposes. The structures of these data bases and the way in which they will be related to the data bases which use the codes will be discussed below. Standard codes for Fish and invertebrates can be found in appendices G and H respectively. Standard codes for identification of stations can be found in appendix I.

10.4 - The Design Tables

The tables below describe the detailed structures of, and relationships among data bases. The tables list the:

- **FIELD NAME** - The name of the field. This will not be changed without discussion with the senior analyst.
- **VARIABLE TYPE** - A = Alpha Numeric, I = Integer, R = Real
- **VARIABLE FORMAT** - The format for the values in the variable
- **INDEX** - An indication if the field is a INDEX or KEY variable. If the variable is noted as a YES1 (Y1) it means that it is a variable which will be cross referenced with other data bases and that it will be required as a unique identifier within this data base. If it is noted as a YES2 (Y2), it will be extensively cross referenced but it is not required as a unique identifier in this particular data base.
- **DEFINITION** - A brief definition of the purpose and contents of the variable.

10.5 MASTER Data Sets

The following datasets resided in a library named MAS. They are master data sets, containing data that will not change over time, except in rare cases. These data will be available to all users for reading.

10.5.1 STATIONS data set

This data base describe the stations as they are set up conceptually. It will contain data on location, classification of the station, and general characteristics expected at the stations. This data base is cross referenced with many other data bases including:

- o EVENTS
- o SAMPLOG

These cross references are done via the field STATION.

Variable Name	Type	Format	Index	Definition
STATION	A	A8	Y1	EMAP Station Name
STA_LATD	I	I2	N	Degrees of Latitude
STA_LATM	I	I2	N	Minutes of Latitude
STA_LATS	R	F6.3	N	Seconds and Decimal Seconds of Latitude
STA_LNGD	I	I2	N	Degrees of Longitude
STA_LNGM	I	I2	N	Minutes of Longitude
STA_LNGS	R	F6.3	N	Seconds and Decimal Seconds of Latitude
STA_SYS	A	A6	Y2	System in which Station is Located
STASTATE	A	A2	Y2	State Code
STA_SAL	A	A10	N	Expected Salinity Classification
STA_DEP	R	F6.2	N	Expected Depth at Station (meters)
STA_BOT	A	A10	N	Expected Bottom Type
STA_HEX	R	F12.3	Y2	EMAP Hexagon Number
STA_INPB	R	F5.2	N	Inclusion Probability

Inherent in the design of the sampling program is the fact that a

particular station can have a number of different classifications. In order to deal with this two data bases will have to contain information on the the type of station. The stations data base will have an additional look up table for STA_CLAS. Its structure is defined below. The event data base LOCEVENT, which will contain one record for each sampling event will also have a lookup table for the EVENT CLASS (EVNTCLAS) which will indicate the reason each station was visited for that event.

10.5.2 Station Classification Data Base

This lookup table contains all the station classifications assigned to a station. A relational join to the STATIONS data base will result in all the station classifications being assigned to each station.

Variable Name	Type	Format	Index	Definition
STATION	A	A8	Y1	EMAP Station Name
STA_CLAS	A	A4	Y1	Station Classification Code

10.5.3 EVNTCLAS Data Set

This is also a look up table which resolves the sampling event classification for each event at each station.

Variable Name	Type	Format	Index	Definition
EVNTNUM	I	I6	Y1	Event Number (Randomly Assigned)
EVNTCLAS	A	A8	Y2	Purpose of Visit

10.5.4 Data Dictionary Data Sets

The Data Dictionary is a SAS dataset that will contain one record for every unique variable in the EMAP Near Coastal Database. These records will contain information about the variables as they are defined by SAS: VARNAME is the variable name, VARLABEL is the SAS label for the

variable, LIBNAME is the SAS Library name in which the variable is stored, MEMNAME is the dataset name, VAR_FMT is the SAS format assigned to the variable, and VAR_TYPE and VAR_LEN are the defined data type and length of the variable. Two additional variables VAR_MIN and VAR_MAX will be entered for numeric variables if they have minimum and maximum possible values. A copy of this data set will be kept in the Results Directory (RES) for access by all interested personnel.

The Data Dictionary may be updated during the data collection phase as analytic labs return results files with new variables. As new datasets are defined and added to the database, new entries will be added to the dictionary in two separate files. The DATASETS file contains 1 record for each SAS dataset in the database system. The DATADICT file contains one record for each variable in each SAS dataset. The two files can be linked on the variable DATASET.

DATA DICTIONARY data set

Variable Name	Type	Format	Index	Definition
VARNAME	A	A8	Y1	Variable Name
VARLABL	A	A40	N	Variable Label/Description
LIBRARY	A	A3	N	SAS Library Location
DATASET	A	A8	Y2	SAS File Name
VARSHARE	A	A1	N	Variable Exists on Multiple Files?
VARTYPE	A	A1	N	Variable Data Type
VARFMT	A	A10	N	Variable Format
VAR_LKUP	A	A8	N	Lookup File (SAS format)
VAR_MIN	N	R10.3	N	Minimum Acceptable Value
VAR_MAX	N	R10.3	N	Maximum Acceptable Value

10.5.5 DATASETS - Data Set Index - Listing of available data sets

This is an index of the available data sets in the NC EMAP system. It contains one record for each SAS dataset in the Data Base. The structure of this data base will be expanded when historical data sets are included in the INDEX.

Variable Name	Type	Format	Index	Definition
LIBRARY	A	A3	Y2	SAS Library Name
DATASET	A	A8	Y1	SAS Dataset Name
FILEDESC	A	A25	N	Description of Dataset

10.6 Tracking Data Sets

The following datasets reside on the ONL library and are used for tracking data and describing data elements. These data sets are not available to the general user community until after all analyses on a station are complete. At that point, the data will be moved from these "LOCKED" data sets to identically structured data bases in the RES directory.

10.6.1 LOCEVENT database details

One record is stored in this dataset for each sampling event. It can be linked to EVNTCLAS records on the variable EVNTNUM. The Event record will be updated as new data for the event is collected and tracked. When all data are complete for an event, the data in this data set will be transferred to a data base called EVENT in the RES directory.

Variable Name	Type	Format	Index	Definition
EVNTNUM	I	I6	Y1	Event Number (Randomly Assigned)
STATION	A	A8	Y2	Station Identifier

INTERVAL	I	I4	Y2	Sampling Interval of event
EVNTDATE	I	YYMMDD	Y2	Date of Event
TIMEON	I	HHmm	N	Time Sampling Event Began
TIMEOFF	I	HHmm	N	Time Sampling Event was Completed
EVNTDISK	I	I7	N	Diskette Number containing raw data
EVNTCREW	I	I2	Y2	Crew Number that performed Event
EVNTLOG	A	A40	N	File Spec. of Event Log file
EVNTCOMP	A	A15	N	Was Event Successfully Completed?
LOCERROR	A	A20	N	Estimated Error in Locational System
T_WTRTOX	A	A10	N	Status of H2O Toxicity Data
T_CTDRET	A	A10	N	Status of CTD Retreival Data
T_HYDDEP	A	A10	N	Status of HydroLab Deployment
T_HYDRET	A	A10	N	Status of HydroLab Retreival Data
T_BENINF	A	A10	N	Status of Benthos Samples
T_SEDGRN	A	A10	N	Status of Sediment Chemistry Samples
T_SEDQA	A	A10	N	Status of Sediment QA Samples
T_FTRAWL	A	A10	N	Status of Fish Trawl Data
T_BIVTOW	A	A10	N	Status of Bivalve Samples
T_COMP	A	A10	N	Flag to indicate if all Analyses Complete

10.6.2 LOGLOCK Data Set details

This dataset will contain one record for every SAMPLEID recorded by the data collection crews. The record will define the sample type, and will contain tracking information for the sample as it is collected. Records for samples that are shipped to analytic labratories will be linked to the SHIPMENT dataset on the variable SHIPNUM, providing a trackiing record of the sample. Detailed information on the sample, as well as the results of labratory analysis, will be saved in datasets in the Results Library. The LOGLOCK record can be linked to those datasets on the variable SAMPLEID, and can be linked to the Event and Station level datasets (LOCEVENT, STATION) on the variables EVNTNUM or STATION. This data set serves the purpose of linking results data for a sample to the Station it was collected at, as well as saving all tracking information for the sample.

When all the analyses expected for a station are complete these data will be transferred to a data set called SAMPLOG in the RES directory.

Variable Name	Type	Format	Index	Definition

SAMPLEID	I	I9	Y1	Sample Id.
EVNTNUM	I	I6	Y2	Event Number
STATION	A	A9	Y2	Station Identifier
SAMPTYPE	A	A10	Y2	Sample Type
SHIPNUM	I	I9	Y2	Shipment Number
REC_COND	A	A10	N	Condition of Sample when Received
SAMPSTAT	A	A10	N	Current Satus of Sample (Collected, Shipped, Received, etc.)
STATDATE	I	I6	Y2	Date of Last Change in Status

10.6.3 SHIPMENT data set

This dataset, stored on the ONL library, will contain one record for each shipment of samples sent from the sampling crews to the analytic labratories. The record will contain shipment identification information and the dates the shipment was sent and was received. There will only be one record per shipment although many samples will be shipped at once. The Shipment record can be linked to the dataset LOGLOCK on SHIPNUM to list all samples included in a shipment. This data will be transferred to an identical data set on RES once all analyses for the shipment are complete.

Variable Name	Type	Format	Index	Definition
SHIPNUM	I	I6	Y1	Shipment Box Number
COURIER	A	A25	N	Shipment Courier Service
COUR_ID	A	A25	N	Courier Id. Number
SHIPDATE	I	YYMMDD	Y2	Date Shipped
SHIPTIME	I	HHMM	Y2	Hour Shipped
SHIP_BY	A	A25	N	Person who packed Shipment
SHIPDEST	A	A25	N	Destination of Shipment
REC_DATE	I	YYMMDD	N	Date Shipment Received
REC_TIME	I	HHMM	N	Hour Shipment Received
REC_BY	A	A25	N	Person who unpacked Shipment
REC_COMP	A	A1	N	Shipment Complete?

10.7 The Results Library

Data collected on the field data sheets will be stored in the results library in separate datasets. Each dataset will be indexed on SAMPLEID. Every record in each of the results datasets will have a matching record in the Sample Tracking dataset LOGLOCK, which will also be indexed on SAMPLEID. The sample tracking dataset will serve as a means of linking samples to events and stations. Results data from Analytic labs will also be stored in separate data sets on the results library, and will be indexed on SAMPLEID.

10.7.1 BOATLOC - Boat Location Data

This dataset contains additional information on entire sampling events (or "visits") that is not stored in the event tracking file LOCEVENT. There is a one-to-one relationship between records in this file and in LOCEVENT. The records can be joined on EVNTNUM.

Variable	Type	Format	Index	Description
EVNTNUM	I	I6	Y1	Event Number
BOATNAME	A	A20	Y2	Boat Name
CRWCHIEF	A	A20	Y2	Crew Chief
CRWMEMB1	A	A20	N	Crew Member #1
CRWMEMB2	A	A20	N	Crew Member #2
LANDCRW	A	A20	N	Land Crew Member
VISITOR1	A	A20	N	Visitor #1
VISITOR2	A	A20	N	Visitor #2
EVNTPURP	A	A20	N	Purpose of Visit
EVNTLATD	I	I3	N	Event Latitude: Degrees
EVNTLATM	R	R5.2	N	Event Latitude: Minutes
EVNTLATS	R	R5.2	N	Event Latitude: Seconds
EVNTLNGD	I	I3	N	Event Longitude: Degrees
EVNTLNGM	R	R5.2	N	Event Longitude: Minutes
EVNTLNGS	R	R5.2	N	Event Longitude: Seconds
TRASH	I	I1	N	Trash Present at Station
WOOD	I	I1	N	Trash at Station: Wood
PLASTIC	I	I1	N	Trash at Station: Plastic
OTHTRASH	I	I1	N	Other Trash at Station
DATASHTS	I	I2	N	Number of Data Sheets
COMP_BY	A	A20	N	Event Data Sheets Completed By
REVIEWBY	A	A20	N	Event Data Sheets Reviewed By

10.7.2 Benthic Community Data

These results will be stored in 5 data bases which can be relationally joined by the SAMPLEID field.

10.7.2.1 BENGRN Benthos Sample Grain Sample

This table is a cross reference between the SAMPLEID and the grab number from which the sample was generated. This table will allow the automatic cross referencing of the grain size sample with the community analysis.

Variable	Type	Format	Index	Description
SAMPLEID	I	I9	Y1	Sample Id.
BENGRAB	I	I2	Y1	Benthic Sample Grab Number

10.7.2.2 BENINF Benthic Infauna Sample

This dataset will contain information on Benthic Infauna samples recorded by the field crews as samples are collected.

Variable	Type	Format	Index	Description
SAMPLEID	I	I9	Y1	Sample Id.
BENGRAB	I	I2	Y2	Benthic Sample Grab Number
GRABTIME	I	I4	N	Benthic Grab Time
BEN_RPD	I	R6.2	N	Benthic Sample RPD Depth
BEN_JARS	I	I1	N	Number of Jars Used

10.7.2.3 Benthic Community Results

The following data bases represent the results of the taxonomic analyses of the field samples. This design is based on the design currently used by VERSAR to store and analyze their data. Our structures differ from Versar's LTB files in that we do not include Station ID. and Date in our files. Instead, we use the SAMPLEID to identify the sample,

which can be linked to the Sample Log and Event Log files (LOGLOCK, LOCEVENT) to retreive the Station and Sampling Date. Also the Bottom Conditions File (BENTH_SD) does not include variables such as Botton Depth and Bottom DO that can not be derived from the samples analyzed at VERSAR.

10.7.2.3.1 BEN_SAMP - Benthic sample characteristic

Summarizes the characteristics of the core sample which will apply to all the species in the sample.

Variable	Type	Format	Index	Description
SAMPLEID	I	I6	Y1	Sample Id. Number
CONVFACT	R	F9.2	N	Converts #/Sample to #/m**2
SAMPVOL	R	F5.2	N	Sample Volume (liters)
NETMESH	R	I6	N	Screen Mesh Width (mm)

10.7.2.3.2 BENTH_TX - Benthos Species Abundance

This data set contains the actual species counts for each SAMPLEID. There will be a single entry for each species found for each SAMPLEID. This data set will be joinable with any other data set by SAMPLEID. This structure is modeled after the structure used by VERSAR (Rahnasinga, Personal Communication).

Variable	Type	Format	Index	Description
SAMPLEID	I	I6	Y1	Sample Id. Number
SPECCODE	A	A8	Y1	The species EMAP code
REP	R	I2	N	Sample Replicate #
CNT_05MM	R	F10.2	N	#/Sample (0.5mm Screen)
CNT_1MM	R	F10.2	N	#/Sample (1.0mm Screen)
DEN_05MM	R	F10.2	N	#/m**2 (0.5mm Screen)
DEN_1MM	R	F10.2	N	#/m**2 (1.0mm Screen)
N05_100L	R	F10.2	N	#/100L (0.5mm Screen)

10.7.2.3.3 BENTH_SD - Benthos Sample Bottom Condition data

This data set contains chemical and physical characteristics of the the sediment core taken out of the benthic community sample. These analyses come with the benthic community analysis.

Variable	Type	Format	Index	Description
SAMPLEID	I	I6	Y1	Sample Id. Number
CARBNATE	R	F5.1	N	Carbonate Content (%)
CARCHN	R	F5.1	N	Carbon Content: CHN Analyser (%)
CARIGN	R	F5.2	N	Carbon Content: Ignition (%)
CARWET	R	F5.2	N	Carbon Content: Wet Oxidation (%)
INTSAL	R	F4.1	N	Interstitial Salinity (ppt)
MEDDIAM	R	F5.2	N	Median Diameter (phi)
MOIST	R	F5.1	N	Sediment Moisture Content (%)
QUARTDEV	R	F5.2	N	Quartile Deviation
SAND	R	F5.1	N	Sand Content (%)
SILTCLAY	R	F5.1	N	Silt-Clay Content (%)

10.7.2.3.4 BENSPEC - Benthic Species Name Look-up Table

This data set allows the resolution of the species code and a cross reference to codes and naming schemes used in other programs.

Variable	Type	Format	Index	Description
SPECCODE	A	A5	Y1	EMAP Species Identification Code
BIOSCODE		A	A20 Y2	Bios Species Code
NODCCODE	A	A20	Y2	National Oceanographic Data Center Code (NOAA-NODC)
PHYLUM	A	A20	N	Phylum Taxonomic Name
CLASS	A	A20	N	Class Taxonomic Name
ORDER	A	A20	N	Order Taxonomic Name
FAMILY	A	A20	N	Family Taxonomic Name
GENUS	A	A20	N	Genus Taxonomic Name
SPECIES	A	A20	N	Species Taxonomic Name

10.7.3 Sediment Samples

Several different sediment samples will be extracted from one homogenized sediment "grab" at certain stations. The different samples will have different purposes: Sediment Chemistry analysis, Sediment Toxictity testing, QA reference samples, etc. Two SAS datasets will hold the field data collected for these samples: SEDCHM for Sediment Chemistry samples and SEDSMP for all other types of sediment samples. SEDCHM will contain data elements describing the "grabbing" process for the sampling event, as well as the volume of sediment saved specifically for Sediment Chemistry analysis. SEDSMP will contain only the volume of sediment saved for one particular sample. The two datasets can be linked together by linking records from each to the LOCEVENT dataset, and merging records on EVNTNUM.

10.7.3.1 SEDCHM Sediment Chemistry Sample Information

This data set contains a description of the sample and the sampling procedures used on the boat to get the sample.

Variable	Typ	Format	Index	Description
SAMPLEID	I	I9	Y1	Sample Id.
HOMGGRAB	I	I2	N	Homogenized Sed: no. of Grabs
HOMGFAIL	I	I2	N	Homogenized Sed: no. unsuccess. Grabs
HOMGCC	I	I2	N	Homogenized Sed: Volume
SEDCHMCC	I	I4	N	Sediment Chemistry Sample Volume

10.7.3.2 SEDSMP - Sediment Samples

The type of sample for records in this dataset will be saved in the Sample Tracking File LOCEVENT. Sediment samples of several types (Sediment Toxicity, QA Reference Samples) can thus be saved in this one file.

Variable	Type	Format	Index	Description
SAMPLEID	I	I9	Y1	Sample Id.
SEDCC	I	I4	N	Sediment Sample Volume

10.7.4 CTDRET - CTD Retreival Data

This dataset will contain data on the CTD retreivals, including the file name of the raw output file from the CTD. This dataset can be linked to Stations and Sampling dates by first linking to the Sample Tracking file on SAMPLEID and then the the Event Tracking file LOCEVENT on EVNTNUM, and to the STATION dataset by the variable STATION. This will allow all raw output files of CTD data to be listed for a certain station, by date.

Variable	Type	Format	Index	Description
SAMPLEID	I	I9	Y1	Sample Id.
CTDFILE	A	A20	N	CTD Raw Data File Name
CTD_ID	I	I6	Y2	Equipment Id. Number
CTD_SER	I	I9	N	Equipment Serial Number
BOT_DO	I	I1	N	CTD Bottom DO
BOT_SAL	R	R3.1	N	CTD Bottom Salinity

10.7.5 CHL_A - Chlorophyl Analysis Data

This dataset will store information on Chlorophyl Analysis samples. Records in this file will be linked to Sample Tracking records for these samples in the LOGLOCK datset, which can be linked to Events by EVNTNUM.

Variable	Type	Format	Index	Description
SAMPLEID	I	I9	Y1	Sample Id.

CHLA_ML	I	I9	N	Volume Filtered for Chloro. A
FLUORMET	I	R4.2	N	Fluorometer Reading

10.7.6 SUSPSO - Suspended Solids Data

This dataset will contain data collected in the field for Suspended Solids Analysis.

Variable	Type	Format	Index	Description
SAMPLEID	I	I9	Y1	Sample Id.
SUSPMETR	I	R8.7	N	Transmissometry Reading

10.7.7 Hydrolab Tracking and Reporting Data Base

These data bases will track the deployment, retrieval, maintenance and resulting data from the Hydrolab deployments.

10.7.7.1 HYDDEP - Hydrolab Deployment Data

This dataset contains data collected in the field whenever Hydrolabs are deployed.

Variable	Type	Format	Index	Description
SAMPLEID	I	I9	Y1	Sample Id.
HLD_ID	I	I6	N	HydroLab Deployment: Equip. Id.
HLDSER	I	I9	N	HydroLab Deployment: Eqip. Ser. Num.
HLDLATD	I	I2	N	HydroLab Deploy.: Latitude Degrees
HLDLATM	I	I2	N	HydroLab Deploy.: Latitude Minutes
HLDLATS	R	F5.2	N	HydroLab Deploy.: Latitude Seconds
HLDLNGD	I	I2	N	HydroLab Deploy.: Longitude Degrees
HLDLNGM	I	I2	N	HydroLab Deploy.: Longitude Minutes
HLDLNGS	R	F5.2	N	HydroLab Deploy.: Longitude Seconds

10.7.7.2 HYDRET - HydroLab Retreival Data

This dataset contains data collected in the field whenever Hydrolabs are retreived including the dataset name of the Raw Data File downloaded from the Hydrolab.

Variable	Type	Format	Index	Description
SAMPLEID	I	I9	Y1	Sample Id.
HLRFILE	A	A20	N	HydroLab Raw Data Filename
HLR_ID	I	I6	N	HydroLab Retreival: Equip. Id.
HLRSER	I	I9	N	HydroLab Retreival: Eqip. Ser. Num.
DO_DROP	I	I1	N	HydroLab Retr.: DO dropped to 0?
HLRFLAG	I	I1	N	HydroLab Retr.: Data Downloaded?
HLRLATD	I	I2	N	HydroLab Retr.: Latitude Degrees
HLRLATM	I	I2	N	HydroLab Retr.: Latitude Minutes
HLRLATS	R	F5.2	N	HydroLab Retr.: Latitude Seconds
HLRLNGD	I	I2	N	HydroLab Retr.: Longitude Degrees
HLRLNGM	I	I2	N	HydroLab Retr.: Longitude Minutes
HLRLNGS	R	F5.2	N	HydroLab Retr.: Longitude Seconds

10.7.7.3 Hydrolab Summaries

This data set contains a summary of the hydrolab units used at each staion and the duration for which the instrument was deployed. This data set will be created at the end of the field program. It will be created by analyzing the other data sets for the Hydrolab data.

Variable	Type	Format	Index	Description
STATION	A	A10	Y1	The station name
DEPDATE	D	MMDDYY	Y1	The date the hydrolab was deployed
RETDATE	D	MMDDYY	Y2	The date the unit was retrieved
HL_ID	I	I6	N	HydroLab Equipment Id Number.
HL_SER	I	I9	N	HydroLab Eqipment Serial Num.

10.7.8 EQUIPQC - Equipment QC data

The Equipment QC Dataset will contain QA/QC data recorded on data sheets for HydroLab deployments and retreivals, and for CTD retrievals. Records in this Dataset can be linked to any of those sampling procedures by joining the QC_ID in EQUIPQC to the Sampleid in CTDDEP, HYDDEP, or HYDRET. This will be a one to one join. Also, many QC records can be found for the equipment used in a sampling procedure by joining files on EQUIP_ID (a many to one join). This will allow us to scan several QC checks performed over a period of time on the equipment used at any sampling procedure.

Variable	Type	Format	Index	Description
QC_ID	I	I6	Y	Equipment QC Identifier
QC_TYPE	A	A3	Y2	Type of QC Check
EQUIP_ID	I	I6	Y2	Equipment Id. Number
EQUIPSER	I	I9	Y2	Equipment Serial Number
QC_DATE	I	YYMMDD	Y2	Date of QC Check
QC_TIME	I	I4	N	Time of QC Check
UNIT_PH	R	R4.2	N	Ph Reading from Unit
UNITSAL	R	R4.2	N	Salinity Reading from Unit
UNITTMP	R	R4.2	N	Temperature Reading from Unit
UNITSAT	R	R5.2	N	Saturation Reading from Unit
UNIT_OX	R	R4.2	N	Oxygen Reading from Unit
AMBI_PH	R	R4.2	N	Ambient Ph
AMBI_SAL	R	R4.2	N	Ambient Salinity
AMBI_TMP	R	R4.2	N	Ambient Temperature
AMBI_SAT	R	R5.2	N	Ambient Saturation
AMBI_OX1	R	R4.2	N	Ambient Oxygen: First Winkler
AMBI_OX2	R	R4.2	N	Ambient Oxygen: Second Winkler
DIFF_PH	R	R4.2	N	Diff. in Ph Readings
DIFF_SAL	R	R4.2	N	Diff. in Salinity Readings
DIFF_TMP	R	R4.2	N	Diff. in Temperature Readings
DIFF_SAT	R	R5.2	N	Diff. in Saturation Readings
DIFF_OX1	R	R4.2	N	Diff. in Ox. Readings (1st Winkler)
DIFF_OX2	R	R4.2	N	Diff. in Ox. Readings (2nd Winkler)

10.7.9 FTRAWL - Fish Trawl Data

This dataset will include all data collected in the field for fish trawls. There can be at the most two fish trawls per sampling event, so there can be up to two records per event in this data set. Records in this file can be linked to individual fish sample records in the FPATH and FLENGTH data sets by matching the TRAWLID in those files the the SAMPLEID of the record in this file. Multiple fish samples can be linked to the same Fish Trawl.

Variable	Type	Format	Index	Description
SAMPLEID	I	I9	Y1	Sample Id.
FTRLNUM	I	I1	Y1	Fish Trawl Number (1 or 2)
FNETID	I	I4	N	Fish Trawl Net Id.
FTRLDRIV	A	A20	N	Fish Trawl Pilot
FLINER	I	I1	N	Fish Trawl Liner Size
FBEG_DEP	I	I4	N	Fish Trawl Start Depth
FLINEOUT	I	I4	N	Fish Trawl Line out (in feet)
FTRLSTRT	I	I4	N	Fish Trawl Start Time
FDEGMAG	I	I3	N	Degree Magnetic
FBEGLATD	I	I2	N	F. Trawl Start: Latitude Degrees
FBEGLATM	I	I2	N	F. Trawl Start: Latitude Minutes
FBEGLATS	R	F5.2	N	F. Trawl Start: Latitude Seconds
FBEGLNGD	I	I3	N	F. Trawl Start: Longitude Degrees
FBEGLNGM	I	I2	N	F. Trawl Start: Longitude Minutes
FBEGLNGS	R	F5.2	N	F. Trawl Start: Longitude Seconds
FTRACK	I	I2	N	F. Trawl Track
FDEPMIN	I	I4	N	Fish Trawl Minimum Depth
FDEPMAX	I	I4	N	Fish Trawl Maximum Depth
FTRL_STW	I	I2	N	Fish Trawl Speed Through Water
FTRL_SOB	I	I2	N	Fish Trawl Speed Over Bottom
FTRL_END	I	I4	N	Fish Trawl End Time
FENDLATD	I	I2	N	F. Trawl End: Latitude Degrees
FENDLATM	I	I2	N	F. Trawl End: Latitude Minutes
FENDLATS	R	F5.2	N	F. Trawl End: Latitude Seconds
FENDLNGD	I	I2	N	F. Trawl End: Longitude Degrees
FENDLNGM	I	I2	N	F. Trawl End: Longitude Minutes
FENDLNGS	R	F5.2	N	F. Trawl End: Longitude Seconds
NUMTRLS	I	I2	N	Number or Trawls
FTRLCOMP	I	I1	N	Fish Trawl Completed Successfully, 0=no, 1 = yes
FTRASH	I	I1	N	Fish Trawl Trash
FWOOD	I	I1	N	Fish Trawl Trash: Wood

FPLASTIC	I	I1	N	Fish Trawl Trash: Plastic
FCANS	I	I1	N	Fish Trawl Trash: Metal Cans
FBOTTLES	I	I1	N	Fish Trawl Trash: Bottles
FTRSHOTH	I	I1	N	Fish Trawl Trash: Other

10.7.10 FPATH - Fish Pathology

This dataset will contain one record for each Fish Pathology sample collected. It can be linked with a Fish Trawl record on the variable FTRAWLID.

Variable	Type	Format	Index	Description
SAMPLEID	I	I6	Y1	Sample Id. Number
FTRAWLID	I	I6	Y2	Fish Trawl Id. Number
SPECIES	A	A20	Y2	Species Name
LENGTH	I	I3	N	Length of Fish
PATHPART	I	I2	N	Part Saved?
FPATH_1	A	A1	N	Skin Discolor
FPATH_2	A	A1	N	Raised Scales
FPATH_3	A	A1	N	Spots
FPATH_4	A	A1	N	Ulcers (Body Surface)
FPATH_5	A	A1	N	Fin Erosion
FPATH_6	A	A1	N	Lumps or Growths
FPATH_7	A	A1	N	Surface Parasites
FPATH_8	A	A1	N	Opercular Deformity
FPATH_9	A	A1	N	Other Body Surface Path #1
FPATH_11	A	A1	N	Other Body Surface Path #2
FPATH_12	A	A1	N	Skeletal Malformations
FPATH_13	A	A1	N	Eye Pathology
FPATH_14	A	A1	N	Gill Color
FPATH_15	A	A1	N	Gill Erosion
FPATH_16	A	A1	N	Gill Deformity
FPATH_17	A	A1	N	Gill Parasites
FPATH_18	A	A1	N	Gill Lumps or Growths
FPATH_19	A	A1	N	Buccal Cavity Hemorhage
FPATH_20	A	A1	N	Buccal Cavity Parasites
FPATH_21	A	A1	N	Buccal Cavity Lumps, Growths
FPATH_22	A	A1	N	Other Buccal Cavity Path
FPATH_23	A	A1	N	Other Pathologies

10.7.11 FLENGTHS - Fish Lengths

This dataset will contain one record for each Fish Chemistry sample collected. It can be linked with a Fish Trawl record on the variable FTRAWLID.

Variable	Type	Format	Index	Description
SAMPLEID	I	I6	Y	Sample Id. Number
FTRAWLID	I	I6	Y2	Fish Trawl Id. Number
SPECIES	A	A20	Y2	Species Name
LENGTH	I	I3	N	Length of Fish
SAMPTYPE	A	A10	Y2	The type of sample Fish Tissue or Taxonomy

10.7.12 BTRAWL - Bivalve Tow Data

This dataset will include all data collected in the field for bivalve tows. There can be at the most one bivalve tow per sampling event. Records in this file can be linked to individual bivalve chemistry records in the BIVCOUNT dataset by matching the TRAWLID in those files with the SAMPLEID of the record in this file. Multiple bivalve samples can be linked to the same Bivalve Tow.

Variable	Type	Format	Index	Description
SAMPLEID	I	I9	Y1	Sample Id.
BDREDGE	I	I4	N	Bivalve Tow Dredge Id.
BTOWDRIV	A	A20	N	Bivalve Tow Pilot
BSTRTDEP	I	I4	N	Bivalve Tow Start Depth
BLINEOUT	I	I4	N	Bivalve Tow Line out (M)
BTOWBEG	I	I4	N	Bivalve Tow Start Time
BDEGMAG	I	I3	N	Degree Magnetic
BBEGLATD	I	I2	N	Biv. Tow Start: Latitude Degrees
BBEGLATM	I	I2	N	Biv. Tow Start: Latitude Minutes
BBEGLATS	R	F5.2	N	Biv. Tow Start: Latitude Seconds
BBEGLNGD	I	I2	N	Biv. Tow Start: Longitude Degrees
BBEGLNGM	I	I2	N	Biv. Tow Start: Longitude Minutes
BBEGLNGS	R	F5.2	N	Biv. Tow Start: Longitude Seconds
BTRACK	I	I2	N	Biv. Tow Track
BDEPMIN	I	I4	N	Bivalve Tow Minimum Depth
BDEPMAX	I	I4	N	Bivalve Tow Maximum Depth

BTOW_SOB	I	I2	N	Bivalve Tow Speed Over Bottom
BTOW_END	I	I4	N	Bivalve Tow End Time
BE_LATD	I	I2	N	Biv. Tow End: Latitude Degrees
BE_LATM	I	I2	N	Biv. Tow End: Latitude Minutes
BE_LATS	R	F5.2	N	Biv. Tow End: Latitude Seconds
BE_LNGD	I	I2	N	Biv. Tow End: Longitude Degrees
BE_LNGM	I	I2	N	Biv. Tow End: Longitude Minutes
BE_LNGS	R	F5.2	N	Biv. Tow End: Longitude Seconds
BTOW_NUM	I	I2	N	Number or Tows
BTOWCOMP	I	I1	N	Bivalve Tow Completed Successfully?

10.7.13 Bivalve data

The bivalve data consists of data regarding the counts of a particular species, and of meassurements about individuals of each species at each station.

10.7.13.1 BIVCOUNT- Bivalve Count Data

This data set contains the number of bivalves found for each species.

Variable	Type	Format	Index	Description
SAMPLEID	I	I6	Y1	Sample Id. Number
SPECCODE	A	A8	Y1	The species code which can be resolved in the database SPECCODE.
BIV_CNT	I	I3	N	Total Bivalves Counted of ths species

10.7.13.2 BIVALVE - Bivalve Length Data

This data set contains the lengths of the bivalves measured.

Variable	Type	Format	Index	Description
SAMPLEID	I	I6	Y1	Sample Id. Number
SPECCODE	A	A8	Y1	The species code which can be resolved in the database SPECCODE.
BIVNUM	I	I2	Y1	Index number for each bivalve 1 -> 30
BIVLENTH	R	R4.1	N	Length of Specimen (CM)

10.7.14 Analytical Chemistry Results

The Following Results Files are based on the NOAA Status and Trends data system for storing chemistry results.

10.7.14.1 CHEMLOG

When an individual sample is given to the analytical chemists they are likely to partition the sample into a number of different subsamples for different analyses. Some of the subsamples will be fractionated via sample preparation extractions. Each subsample and each fraction will likely be given an individual ANALYTICAL CHEMISTRY ID NUMBER. The data will be reported with these CHEMID's rather than with the original analysis. In order to track the results and properly relate them to the original samples, a relational data base called CHEMLOG will be used to cross reference chemistry ID number with original SAMPLEID's.

Variable	Type	Format	Index	Description
SAMPLEID	I	I6	Y1	Sample Id. Number
CHEMID	A	A15	Y1	Chemistry identification number Supplied by the analytical chemsits.
FRACTION	A	A10	Y2	The fraction nubmer e.g. F4, etc.
DETECTID				
ANALTYPE	A	A10	N	The analysis type, ORGANIC, INORGANIC, etc.
QABATCH	A	A10	Y2	The batch number of the sample for cross referencing QA results

10.7.14.2 CHMRESLT

This data set contains the actual results of the chemical analysis. There is one record for each Analyte measured in each analytical sample.

Variable	Type	Format	Index	Description

CHEMID	I	I6	Y1	Sample Id. Number
REPL	I	I2	Y1	Replicate Number
ANALYTE	A	A8	Y1	Analyte Abbreviation
CONC	R		N	Measurements as required by ACG
QCCODE	A	A1	N	Quality Control Code
DETECTID	A	A17	Y2	The id number of the instrument used to do the analysis
SAMPLEQC	A	A1	N	Flag for Control Sample

10.7.14.3 DETECT

This data set contains information unique to the detector which measured results.

Variable	Type	Format	Index	Description
DETECTID	A	A17	Y1	The id number of the instrument used to do the analysis
QABATCH	A	A10	Y1	The batch number of the sample for cross referencing QA results
ANALYTE	A	A8	Y1	Analyte Abbreviation
CHEMID	A	A4	Y2	The chemistry identification number
MATRIX	A	A3	N	
LOD	R		N	Level of Detection
LOQ	R		N	Level of Quantition
LOR	R		N	Level of Reporting
METHOD	A	A8	N	Analytic Method Used

10.7.14.4 LIPID

This data set contains the lipid content of appropriate samples.

Variable	Type	Format	Index	Description
CHEMID	I	I6	Y	Sample Id. Number
QABATCH	A	A10	Y1	The batch number of the sample for cross referencing QA results
LIPID	R	I6	N	Lipid Concentration

10.7.14.5 SEDANC

This data set will contain the grain size data for particulate samples. These data can be keyed into the sample number instead of the CHEMID since the grain size characterizes a sample not a fraction of the sample.

Variable	Type	Format	Index	Description
CHEMID	I	I6	Y1	Sample Id. Number
REPL	I	I	N	Replicate Number
GRAVEL	R	F5.2	N	Percent Gravel Composition
SAND	R	F5.2	N	Percent Sand Composition
SILT	R	F5.2	N	Percent Silt Compostion
CLAY	R	F5.2	N	Percent Clay Coposition
FINE	R	F5.2	N	Percent fine clay compostion
TOC	R	F6.2	N	Total Organic Carbon
TOCQC	A	A1	N	
TIC	R		N	Total Inorganic Carbon
TICQC	A	A1	N	

10.7.14.8 ANALYTE

This data set is a cross reference of the codes used to identify the compounds which were analyzed. The cross reference will include common names, EMAP codes, and standard CAS numbers

Variable	Type	Format	Index	Description
EMAPCHEM	A	A12	Y1	EMAP CHEM CODE
ANALYTE	A	A8	N	Analyte Abbreviation
COMPOUND	A	A40	N	Full Chemical Name
UNITS	A	A20	N	Reporting Units
CASCODE	A	A20	Y2	The CAS number for this compound

10.8 Toxicity Data Bases

The detailed designs for the toxicity data sets are not available at the time of this writing. Conceptually, there will be a single data base for the results of the toxicity test. The results will be given as percentages

relative to the control values. The percentages will be for survival and fertilization. Each percentage will have a flag variable indicating whether or not the values are statistically different form the control values. Additionally each record of results will be given a BATC number which will relate it to a Quality control test. There will be one quality control test for approximately 15 toxicity tests.

10.9 Summary Data Bases and data integration

The data bases described above are intended to facilitate the field sampling, sample tracking, data analysis, QA/QC, and storage of raw data. As the data for each indicator are loaded into these data bases, and they pass validation and verification, the data will be summarized into a different set of data bases which will summarize the parameters needed to evaluate the indicators. It is clear that these summaries will be integrated by station at the first level and then at the station classification on a more consolidated level. Many schemes for post stratification of the data have been discussed. Any post stratification scheme used would potentially generate a summary data base which has indicator measurements summarized by the stratification variable(s). The specific structure of these data bases has not been clearly stated at the time of this writing.

11.0 VAX Directories and File Structures

Directories and files will be maintained by the Near Coastal Information Management Team (NCIMT) on three levels. Level one is on the field laptop computers. These are described below in the section on the field system. Level two is on the ERLN VAX Cluster. Level Three is on the RTP VAX CLuster. The directory structures for levels two is described in this section.

11.1 Access Codes

Two kinds of access codes are associated with directories and files below. The first is a general access which can be controlled using standard VAX file security codes. For these codes the user community is broken down into the following four groups:

1 - System - ADP professional staff with system privileges assigned to them.
2 - Owner - the owner of the file or the directory. In most cases for the EMAP data system the owner will be one of the NCIMT.
3 - Group - the members of the owner's user group. We will have a number of groups in NC EMAP. These groups will be assigned to particular users based on their role in the project.
4 - World - everyone else.

Access is listed as follows:

(SYStem,OWner,GRoup,WOrld) for each category the user can be granted right to Read (R), Write (W), Execute (E) and Delete (D) files. For example, if a file has the following privileges:

(RWED,RWE,R,R)

then system level people can read, write, execute or delete this file. The owner of the file can read, write, and execute the file, the owner's group and the rest of the world can only read the file.

To further control access to files, Access Control Lists (ACL's) can be attached to a directory or a file. The ACLs override the standard access codes applied using access as outlined above. These use the words READ+WRITE+EXECUTE+DELETE, but add a richer set of controls. We will use the additional controls CONTROL, PROPOGATE and NONE. CONTROL gives a particular user the same rights to the file they would have if they were listed as the owner. PROPOGATE is a flag which is placed on a directory to tell the computer to put a particular set of ACL's on every file which is created in a directory. This allows the programmer to

control the access to all files in a directory regardless of when they are made or by whom. The qualifier NONE restricts a particular user from any access to a file even if the regular VMS security gives them access to the file on the ERLN VAX.

11.2 The EMAP Logical Disk

A Virtual Disk will be created on the ERLN VAX Cluster:

DISK$EMAPINFO = EMAP$

ACCESS: (WRED,WRED,WRED,RE)
ACLS: NCPRIME R+W+E+D+C PROPOGATE
NCMASTER R+W+E PROPOGATE
NCLVL1 R+E PROPOGATE
FLDMASTER R+W+E PROPOGATE

11.3 Detailed directory structures for EMAP NC data

The following directories and their subdirectories will be created on the logical disk EMAP$. Their purpose, contents, and ownership and protection information will be further described below following the end of the figure.

The PRODUCTION directory and subdirectories will contain program files run on a routine basis to manage the database.

Directory	Subdirectories
production.....	.datamgt
	shells
	screens
	gis_extracts

doc
group

The Remote Nodes' directories will be set up as receiving areas for raw data as it is sent from field crews and analytic labs. Raw data from the field crews will be archived on the subdirectories

crew1.....boat
lab
ctd
hydro

crew2.....boat
lab
ctd
hydro

crew3.....boat
lab
ctd
hydro

cinci

gulfbr

vegas

toxcenter

The RAWDATA.PARSED subdirectories will hold parsed raw data files created from raw data by the parser. SAS programs will get their input files from these subdirectories. Data files of different types will migrate

to QA_PASS subdirectories as the pass QA/QC checks.

Directory	Subdirectory	subdirectories
rawdata	parsed	events
		qa_pass
		qa_fail
		shipments
		qa_pass
		qa_fail
		suspsol
		qa_pass
		qa_fail
		chloro
		qa_pass
		qa_fail
		grains
		qa_pass
		qa_fail
		benthos
		qa_pass
		qa_fail
		grainhom
		qa_pass
		qa_fail
		sedchem
		qa_pass
		qa_fail
		sedtox
		qa_pass
		qa_fail
		wtrtox
		qa_pass
		qa_fail

ambient.....qa_pass
. qa_fail
.
ctdqa.....qa_pass
. qa_fail
.
ctd.....qa_pass
. qa_fail
.
hydroqa.....qa_pass
. qa_fail
.
hydrodep.....qa_pass
. qa_fail
.
hydroret.....qa_pass
. qa_fail
.
ftrawl.....qa_pass
. qa_fail
.
btrawl.....qa_pass
. qa_fail
.
event_text_files

The SASDATA directory and its subdirectories will hold the SAS dataset libraries.

Directory	Subdirectory	subdirectory
sasdata	online	
	master	results
		lookup

. synthesis
.
.....temp_work

The GIS directories will hold the GIS data system.

Directory	Subdirectory
gis.....	.cov
	tools
	dev

A more detailed description of the above directories and subdirectories follows (note: if protection level, ACL, or owner is not defined for a subdirectory, it will match that of its parent directory.)

PRODUCTION: COM files necessary for routine data management.
Owner: HBUFFUM
Protection: (RWED,RWED,RE,)
ACL: None

.DATAMGT Routine update programs.

.SHELLS SAS Dataset definitions.

.SCREENS SAS routines for data entry.

.DOC Documentation of Database.
Protection: (RWED,RWE,RWE,R)

CREW1: Logon Direcory for Crew 1 for daily uploading of data.
Owner: CREW1
Protection: (RWED,RWED,RWE,)

.BOAT: Archive Area for Boat Logfile data
.LAB: Archive Area for Lab Data Entry Logfile data
.CTD: Archive Area for raw CTD data files
.HYDRO: Archive Area for raw HydroLab data files

CREW2: Logon Direcory for Crew 1 for daily uploading of data.
Owner: CREW2
Protection: (RWED,RWED,RWE,)

.BOAT: Archive Area for Boat Logfile data
.LAB: Archive Area for Lab Data Entry Logfile data
.CTD: Archive Area for raw CTD data files
.HYDRO: Archive Area for raw HydroLab data files

CREW3: Logon Direcory for Crew 1 for daily uploading of data.
Owner: CREW3
Protection: (RWED,RWED,RWE,)

.BOAT: Archive Area for Boat Logfile data
.LAB: Archive Area for Lab Data Entry Logfile data
.CTD: Archive Area for raw CTD data files
.HYDRO: Archive Area for raw HydroLab data files

CINNCI: Logon Direcory for Cinncinatti Lab for uploading of analytic results data.
Owner: CINNCI
Protection: (RWED,RWED,RWE,)

GULFBR: Logon Direcory for Gulf Breeze Lab for uploading of analytic results data.
Owner: GULFBR
Protection: (RWED,RWED,RWE,)

VEGAS: Logon Direcory for Las Vegas Lab for uploading of analytic results data.
Owner: VEGAS
Protection: (RWED,RWED,RWE,)

TOXCENTER: Logon Direcory for Toxicity Center for uploading of analytic results data.
Owner: TOXCENTER
Protection: (RWED,RWED,RWE,)

RAWDATA.PARSED Output data files from Parsing Programs.
Owner: HBUFFUM
Protection: (RWED,RWED,RWE,)
ACL: Read Access to NCMASTER and above

.BOATLOC: Parsed Boat Location data files waiting for QA/QC

.QA_PASS: Parsed Data Files that pass QA/QC

.QA_FAIL: Parsed Data Files that fail QA/QC

.SHIPMENTS through .BTRAWL: Subdirectories for all unique record formats written by the parser, with subdirectories each QA_FAIL, & QA_PASS

SASDATA: Database files in SAS format.
Owner: HBUFFUM
Protection: (RWED,RWED,RWE,)

.ONLINE "Online" Data files: SAS datasets that can be accessed by FSEDIT, and will be updated daily: Data Dictionary (during development), Sample Tracking (local,

unlocked, versions)

.MASTER Master Database files, updated in batch mode only, on weekly or as needed basis.
Owner: HBUFFUM
Protection: (RWED,RWED,R,R)
ACL: Read access to EMAPLVL2 and above.

.RESULTS Sample data: these datasets will grow in size as the data comes in.

.SYNTHESIS Aggregated, merged, synthesized datasets available to users

(More SYNTHESIS subdirectories defined as analysis develops)

.LOOKUP SAS formats (lookup tables) defining valid codes and ranges, with labels.

12.0 Near Coastal Geographic Information Systems (GIS)

A Geographic Information System (GIS) is being developed by ERL-N for the support of the EMAP Near Coastal Task Group. ERL-N's GIS support of EMAP will consist of:

1) Developing coverages pertinent to Near Coastal EMAP.
2) Producing daily/weekly maps of stations and the dates they were sampled.
3) Making Near Coastal EMAP data available on a daily basis for logistical support of EMAP field teams.
4) Developing a user interface which will allow the Near Coastal field coordinator to access EMAP data in support of the sampling program.

12.1 Hardware and Software

EMAP-NC has chosen to use the EPA Agency standard software ARC/INFO as its main GIS software. For the demonstration project this software will run on a VAX 3100 Workstation. The GIS system will have two 300 megabyte disk drives available for data storage. Other hardware available for this effort includes: a Calcomp 1044GT pen plotter, a Calcomp 5913 thermal plotter, a Calcomp digitizer, two Tektronix terminals and two electrostatic plotters (located at EMSL-LV and the University of Rhode Island.

12.2 INFO Data bases

Currently, several ARC/INFO databases have been developed for this project. EMSL-LV has provided ERL-N with a set of ARC/INFO coverages including:

- o The hydrography of the Virginia Province at 1:100,000 scale. This data will be used for a background coverage of maps produced for distribution purposes, for determining accurate locations of sampling stations, and for tracing areas of contaminations back to their source for demonstration purposes.

- o A complete coastline of the United States at 1:2,000,000 scale. This coverage will be used for maps produced on a daily/weekly basis for use by on-site EMAP staff and by EMAP field teams.

- o Several coverages from the Dunn and Bradstreet database. This includes the names and addresses of express shipping companies, marinas, hospitals, and dry ice companies. These databases will be used in the daily logistical operations of the field teams.

Coverages developed by ERL-N include:

- o The EMAP Virginia Province sampling stations including their locations and station types. Maps of this data have been produced and distributed to the EMAP field teams for use in identifying daily sampling sites.

- o A dynamic database, updated daily to weekly, of those stations sampled during the current sampling interval.

One of the 300 megabyte disks will store the ARC/INFO software and the 1:100,000 scale hydrography coverages. A directory will be created, DLG100K, to hold the 1:100,000 scale coverages. The area of the Virginia Province is broken into several coverages each composed of a 1:100,000 quad. All coverages for this area are edge matched.

All other EMAP data will be stored on the second disk. A main directory called NEARC has been created under which a subdirectory tree has been developed to hold all other coverages. The following is a list of all subdirectories and a description of the coverages located in each. Included is a list of variables and their formats for each coverage.

Definition of Formats: First number is the input format, second number is the output format and the third field is the type of variable where I=Integer, F= floating point, the width of the floating point number is given in the fourth (optional) field, C= character, N= Number with decimal places, B= Binary.

12.3 Directory DLG2M

This directory includes all coverages received from EMSL-LV at 1:2,000,000 resolution. This includes:

12.3.1 STATE-NC - the coastline of the Virginia Province projected with Albers,

VARIABLES	FORMAT	DESCRIPTION
STATE-NC-ID	4,5,B	ID for each arc in coverage.
FIPS	6,6,I	State code
REGION	2,2,I	EPA Region

12.3.2 **ROADS** - the roads of the Virginia Province,

VARIABLES	FORMAT	DESCRIPTION
LENGTH	4,12,F,3	Length of arc
ROADS-ID	4,5,B	ID for each arc.
MAJOR1-7	6,6,I	DLG codes for roads
MINOR1-7	6,6,I	DLG codes for roads

12.3.3 **RAIL** - the rail lines for the Virginia Province. Same variables as roads.

12.4 **Directory HEX** - includes EMAP hexagon coverages.

VARIABLES	FORMAT	DESCRIPTION
AREA	4.12.F.3	Size of hexagon
HEX-ID	4,5,b	ID of each hexagon

12.5 **Directory LOGISTICS** - includes:

EXPRESS - express shipping companies,
HOSPITALS - hospitals,
MARINAS - marinas,
ICE - ice companies.

All coverages in this area have the same variables.

VARIABLES	FORMAT	DESCRIPTION
STATE	2,2,C	State location
CNTY	3,3,I	County
TRACT	7,7,N,2	Census tract
ZIP	5,5,I	ZIP codes
LAT,LONG	8,8,N,4	Geographical position
DUNSID	9,9,I	Dunn and Bradstreet ID
NAME	30,30,C	Company name
STREET	25,25,C	Street address

PREDIR	2,2,C	Street direction
CITY	13,13,C	City
SIC,SIC2	6,6,C	Industrial class
GEOLEVEL	1,1,I	How position is determined

12.6 STATIONS - includes:

STATIONS - EMAP sampling stations, The data for this coverage was extracted from the EMAP NC data base and consists of a subset of the variables in that system:

VARIABLES	FORMAT	DESCRIPTION
TEAM	7,7,C	Team assigned to station
STATION	8,8,C	Station name
LATDEG/LNGDEG	10,10,N5	Station position
STA_SYS	5,5,C	Estuary contained in
STA_SIZE	8,8,C	Size of estuary
SCLASSES	20,20,C	Station sampling types

An additional data base called INT1 will be created which contains all sampling stations visited during interval 1. INT1 contains the same variables as STATIONS coverage.

12.7 PLOTS

This directory contains all plot files created for the Demonstration Project. These files can be queued to a plotter for immediate production without extensive preparation. The following plot files have been created:

DELAWARE.PLT - sampling stations in Delaware River system.

INT190 - stations sampled during interval 1 rotated 90 degrees to fit plotter.

10x10t1.plt - a 12 x 10 plot of stations sampled by team 1.

10x12t2.plt - a 10 x 12 plot of stations sampled by team 2.

10x12t3.plt - a 10 x 12 plot of stations sampled by team 3.

T1.PLT, T2.PLT, T3.PLT - plot of stations sampled by each team.

T12.PLT, T22.PLT, T32.PLT - stations to be sampledy each team for interval 2.

T12.PLT, T23.PLT, T33.PLT - stations to be sampled by each team for interval 3.

All plot files are stored in two formats. Those named with the extension .plt are formatted to be spooled to a plotter with the arc PLOT command. Plot files have also been saved in a format that can be moved to any system and sent to a CALCOMP plotter without the use of ARC/INFO. These files have the same names as listed above but with a cal prefix and a .dat extension.

12.8 Directory MAPCOMPS

This directory includes all map compositions created for the Demonstration Project. Map compositions are tools to help automate plotting and manipulation of data. They are used for plots and analyses which are likely to be repeated and in situations where small changes to an existing map or analysis may be needed.

12.9 GIS User Interface

A user interface to ARC/INFO will be developed by ERL-N GIS staff. This will allow individuals with no expertise in the use of ARC/INFO to access and analyze EMAP spatial data. The interface will allow users to make database queries and display the results of those queries on the terminal screen. The interface will also allow users to interactively produce maps which can then be output to a plotter or saved on line for

future use. The interface is a complex undertaking. Work done by the ERLN staff for region II will be used as a model for this interface. Until the interface can be developed all requests for maps will be produced by the Near Coastal GIS team.

As part of the interface, an index will be created which lists and describes all coverages and plot files created for this project. Included in the description will be the projection of the coverage, the origin of the data used for creating the coverage, the date of creation and a list of all variables contained in the coverage. Along with this general index, a file will reside in the PLOTS directory called README.TXT which will contain a current listing and description of all plot files and directions for access to these files.

All programming for the user interface will be done using ARC/INFO, C and Fortran.

13.0 Information and Data Base Security

Security will be the responsibility of the senior system analyst but he may delegate the every day maintenance of security to the senior programmer and/or the senior data base programmer.

13.1 Identifiers

A number of Identifiers will be granted to different accounts to control access to the directories, data bases and files. Access will be controled as for **R**ead, **W**rite, **E**xecute, **D**elete (**RWED**).

13.1.1 NCPRIME

The NCIMT staff at ERLN. These people will have RWED privileges on most files in the system. Some files will have Delete privilege removed from them to protect the NCPRIME group from themselves. Anyone of these people will be able to reset privileges on the protected files to

delete protected files, but they will have to do so actively and consciously.

13.1.2 NCMASTER

The principle facilitators of the Near Coastal EMAP program. These will Include the field coordinator, the logistics manager and the QAO. These people will not have access to data management code but they will have RWED privileges on all dat collected in the project. Again, some data files will have delete privileges removed for everyone so that these people will be able to delete files but only by actively changing the protection on the file first.

13.1.3. NCLVL1

The principle users and assessors of Near Coastal data. These people will have read access to all data on the system regardless of its level of QA/QC. They will not be able to delete any files on the system. They will not be able to change any files on the system.

13.1.4 FLDMASTER The Crew Chiefs and their seconds.

13.1.5 FLDCREW All other crew members.

13.1.6 EMAPLVL1

EMAP personnel outside of Near Coastal with high access privileges. Have access to some data files on the ERLN VAX and read access to all Near Coastal files on the Central EMAP system.

13.1.7 EMAPLVL2

Read access to all summary data bases on the EMAP central data archive.

13.1.8 WORLD

Read access to select summary data bases and reports on the EMAP Central system.

13.2 Data base of Access Rights

A data base will be developed which will indicate which users have what identifiers and rights.

13.3 Data Base and File access

The NCIMS will keep track of which data bases have been accessed and by whom.

References

Broutman, M.A. and D.L. Leonard. 1988. The quality of the shellfish growing waters in the Gulf of Mexico. NOAA, Rockville, MD.

Buhler, D.R. and D.E. Williams. 1988. The role of biotransformation in toxicity. Fish. Aquat. Toxicol. 11:303-311.

Capuzzo, J.M. , M.N. Moore, and J. Widdows. 1988. Effects of toxic chemicals in the marine environment: Predictions of impacts from laboratory studies. Aquat. Toxicol. 11:19-28.

The Conservation Foundation. 1988. Protecting America's Wetlands: An Action Agenda. The Conservation Foundation, Washington DC.

EMAP 1990. Environmental Monitoring and Assessment Program - Overview. EPA/600/9-90/001. Office of Modelling, Monitoring System, and Quality Assurance.

Fein, G.G., J.L. Jacobson, S.W. Jacobson, P.W. Schwartz, and J.K. Dowler. 1984. Prenatal exposure to polychlorinated biphenyls: Effects on birth size and gestational age. Pediatr. 105:315-320.

Food and Drug Administration, Shellfish Sanitation Branch. 1971. 1971 national shellfish register of classified estuarine waters. Northeast Technical Services Unit. Davisville, RI.

Food and Drug Administration, Shellfish Sanitation Branch, and National Oceanic and Atmospheric Administration, Strategic Assessment Branch and National Marine Fisheries Service. 1985. 1985 national shellfish register of classified estuarine waters. Rockville, MD.

Hargis, W.J., Jr. and D.S. Haven. 1988. Rehabilitation of the troubled oyster industry of the lower Chesapeake Bay. Journal of Shellfish

Research 7:271-279.

Jacobson, J.L., S.W. Jacobson, P.M. Schwartz, G.G. Fein, and J.K. Dowler. 1984. Prenatal exposure to an environmental toxin: A test of the multiple effects model. Dev. Psychol. 20:523-532.

Jacobson, S.W., J.L. Jacobson, P.M. Schwartz, and G.G. Feine. 1983. Intrauterine exposure of human newborns to PCBs: Measures of exposure, pp. 311-343. In: F. M. D'itri and M.A. Kamrin (eds.), PCBs: Human and Environmental Hazards. Butterworth Publishers, Boston.

Leonard, D.L., M.A. Broutman, and K.E. Harkness. 1989. The quality of shellfish growing areas on the east coast of the United States. National Oceanic and Atmospheric Administration OAD, Rockville, MD.

Malins, D.C. 1989. Sediment contamination and marine ecosystems: Potential risks to humans, pp. 155-164. In: Contaminated Marine Sediments -- Assessment and Remediation. Published by the National Academy of Sciences, National Academy Press, Washington DC.

Morganthau, T. 1988. Don't go near the water: Is it too late to save our dying coasts? Newsweek, August 1, 42-47.

NRC. 1989. Contaminated Marine Sediments -- Assessment and Remediation. National Academy Press, Washington, DC.

NRC. 1990a. Managing Troubled Waters: The Role of Marine Environmental Monitoring National Academy
Press, Washington, DC.

O'Connor, J.S., J.J. Ziskowski, and R.A. Murchelano. 1987. Index of pollutant-induced fish and shellfish disease. National Oceanic and Atmospheric Administration Special Report, NDS, Rockville, MD.

Office of Technology Assessment (OTA). 1987. Wastes in Marine

Environments. Washington, DC.

Officer, C.B., R.B. Biggs, J.L. Taft, L.E. Cornin, M.A. Tyler, and W.R. Boyn ton. 1984. Chesapeake Bay anoxia: Origin, development, and significance. Science 223:22-27.

Orth, R.J., and K.A. Moore. 1983. Chesapeake Bay: An unprecedented decline in submerged aquatic vegetation. Science 222:51-53.

Parker, C.A., J.E. O'Reilly, and R.B. Gerzoff. 1986. Oxygen depletion in Long Island Sound. Historical Trends Assessment Program, National Oceanic and Atmospheric Administration, OAD, Rockville, MD.

Pearl, H.W. 1988. Nuisance phytoplankton blooms in coastal, estuarine, and inland waters. Limnol. Oceanogr. 33:823-847.

Rabalais, N.N., M.J. Dagg, and D.F. Boesch. 1985. Nationwide review of oxygen depletion and eutrophication in estuarine and coastal waters: Gulf of Mexico. Louisiana Universities Marine Consortium, Data Report.

Sanders, J.E. 1989. PCB pollution in the upper Hudson River. pp. 365-400. In: Contaminated Marine Sediments -- Assessment and Remediation. Published by the National Academy of Sciences, National Academy Press, Washington DC.

Sinderman, C.J. 1979. Pollution-associated diseases and abnormalities of fish and shellfish: A review. Fish. Bull. 76:717-741.

Smart, T., E. Smith, T. Vogel, C. Brown, and K. Wolman. 1987. Troubled Waters Business Week, October 12, 1987, 88-104.

Toufexis, A. 1988. Our filthy seas: The world's oceans face a growing threat from manmade pollution. Time, August 1.

USEPA. 1984. Chesapeake Bay: A framework for action. Prepared for the

U.S. Congress by the Environmental Protection Agency, Region 3, Philadelphia,
PA.