REINAS: Real-Time Environmental Information Network and Analysis System: Concept Statement*

Darrell D.E. Long, Patrick E. Mantey *
Alex T. Pang, Glen G. Langdon, Jr.,
Robert A. Levinson, Harwood G. Kolsky,
Bruce R. Gritton (MBARI),
Carlyle H. Wash (NPS),
Leslie K. Rosenfeld (NPS/MBARI)

UCSC-CRL-93-05
January 4, 1993

Baskin Center for
Computer Engineering and Information Sciences
University of California, Santa Cruz
Santa Cruz, CA 95064 U.S.A.

Keywords: Real-time, Environmental, Sensor, Data management, Visualization, Seabreeze, Monterey Bay

---

*Supported by a Grant from the Office of Naval Research, No. N-00014-92-J-1807

Contents

0. ABSTRACT

REINAS is a research and development program with the goal of designing, developing and testing an operational prototype system for data acquisition, data management and visualization. This system is to support the real-time utilization of advanced instrumentation in environmental science where continuous time measurements and improved spatial resolution allow monitoring and understanding of environmental phenomena in much greater detail than has previously been possible. The system will also support the retrospective use of integrated environmental data sets. The scientific problem chosen to provide a focus for study is in the area of air-sea-land interaction, in particular, seabreeze development and the oceanic response to it in Monterey Bay.

The project is a multi-year effort of the Baskin Center for Computer Engineering and Information Sciences of the University of California, Santa Cruz, in cooperation with the environmental and marine scientists from the Naval Postgraduate School (NPS), Monterey Bay Aquarium Research Institute (MBARI), and the Center for Ocean Analysis and Prediction (NOAA/COAP).

This report is the Concept Statement for the REINAS project in which the initial system requirements and project plan for the system is described along with a preliminary architecture and its subsystems for data collection, data management, processing and visualization which are required for real-time operational applications in the marine environment.

The objective of this document is to provide project participants and reviewers a baseline description of the problem that REINAS is targeting, the technology to be applied, and a plan for meeting project goals. The concepts discussed are meant to provide clear and concise direction for project activities and clear evaluation criteria for products produced by the REINAS team.

Collected in the appendices of this report are details and documentation concerning the data sources, related projects, etc. that seem related to REINAS.

1. Executive Summary

The Baskin Center of the University of California, Santa Cruz (UCSC), in cooperation with environmental and marine scientists of the Monterey Bay region, from the Naval Postgraduate School (NPS), Monterey Bay Aquarium Research Institute (MBARI), and the Center for Ocean Analysis and Prediction (NOAA/COAP), are creating a real-time system for data acquisition, data management and visualization. The goal of "REINAS" (Real-Time Environmental Information Network and Analysis System) is to support the utilization of current and advanced instrumentation in environmental science where continuous time measurements and improved spatial resolution allow monitoring and understanding of environmental phenomena in much greater detail than previously possible.

The computer engineering research in REINAS is the architectural design, development, implementation, and optimization of a computer system to gather and integrate data from geographically dispersed sensors, to control individual programmable sensors, and to support the combined real-time and retrospective use of this data by scientists with workstations connected via a network to a data manager. REINAS has subsystems for data collection, data management, processing and visualization required for real-time operational applications in the marine environment.

The application focus for REINAS is on environmental problems of operational interest: those with small spatial scales and nearly continuous time measurements. A real context for this project is provided by scientific investigations of the interaction between air, land and sea in the development of the seabreeze and the resulting oceanic response. The unique support offered by MBARI, NPS and NOAA/COAP provides a growing array of instruments to support the experiments in Monterey Bay. The scientists from these organizations provide the computer systems researchers at UCSC with environmental science partners who will provide data, instruments, initial system requirements and the context for evaluation of the prototype implementation. The Cooperative Institute for Research in the Integrated Ocean Sciences (CIRIOS) also offers an organization for the continuing use of REINAS.

The goal of the project is to design, develop and implement an optimal (in terms of cost, performance, and flexibility) system architecture for data collection, data management, processing and visualization for real-time operational applications and retrospective research in meteorology and oceanography. The challenge in this applied research using real-time environmental data problems and applications in environmental science is to develop an extensible system architecture exploiting the latest developments in computer hardware: microprocessors, storage, and communications, and to apply and extend the latest developments in computer science and engineering in the key areas of real-time operating systems, networks, databases and electronic libraries, data compression, pattern recognition (e.g. signature analysis), and scientific visualization.

We proposed a four-phase approach to conduct this research:

- Determine system requirements via cooperation with marine scientists, and from characteristics of applications and of present and planned instrumentation;
- Develop a prototype system (real-time data collection, data management, processing and visualization) to support experimental use of available instruments;
- Based on evaluation of prototype in experimental use, modify and improve the system to support operational use of data from the available array of instruments;

- Transition into an operational mode following verification of the completed system.

System design starts with problem/application assessment, and the characteristics of available and planned instruments. Prototype system design will follow, including implementation sufficient to gather, on an experimental and occasional basis, data from all sensors simultaneously for examination and visualization. System functions, user requirements and prototype implementation will be the input for the design and development of an operational system.

Instruments presently available to the participating scientists in oceanography and meteorology have a variety of operational characteristics and data rates. Some can obtain profiles at an angle or direction or rate of choice. The system is to support control of the instruments so that optimal data can be gathered by that instrument relevant to the real-time development of the environmental phenomena being monitored. If successful this may involve some automated signature recognition by applying artificial intelligence techniques to the array of real-time measurements. The system design will be modular so that it can accept future planned and proposed instruments with additional data and data rates.

Data management is the key to the success of the project. Real-time measurements must be evaluated and processed both locally and in conjunction with data from other sensors, data from models, and past observations. This requires data management to support data retrieval and the exploration of data relationships via database query. It also requires distributed cooperative processing to support development of real-time signatures, and the measurements of means and fluctuations in essential flows. Because some of the instruments produce very large quantities of data, often in bursts, it is expected that collocating some processing at the instrument will be optimal, at least for supporting data compression. Based on the processing of measurements and the relationship of these to the immediate past, users are alerted to significant changes in the state of the environmental system under observation. This allows for rapid changes in operation of sensors with capabilities for steering or for changes in resolution.

Data quality assessment will be a critical to the operational value of a real-time system, and this will require dynamic utilization of a database of prior measurements, and support for correlation of data from sensors with that from static or dynamic models.

Visualization or data presentation is being emphasized. A primary objective of the real-time system is to provide investigators with a "picture" of dynamic developments on a meaningful time scale. Presentations must allow display and assimilation of both the individual measurements and the data developed by interpolation and or from models. This visualization support must convey the data to the observer in the geographic context, it must highlight dynamic changes, and must provide tools for rapid assessment of trends, and of the data quality and limitations.

The development of visualization tools is underway, working with environmental scientists at the participating institutions and with experimental (not real-time) data from available instruments. Visualization tools will be applied to real-time data when the prototype system implementation becomes available to support experimental evaluations. Continuing work on visualization will involve close cooperation between information scientists and environmental scientists, to develop data presentations that exploit any new understanding of the environmental phenomena, and changes in instrument characteristics, model developments, and characteristics of the operational system.

2. REINAS Problem Description and Requirements

2.1 Problem Statement and Special Characteristics

The major problem areas of REINAS come directly its goals: (1) to develop an architecture and system, to gather and integrate data from geographically dispersed sensors, and to control individual programmable sensors; (2) to store, manage, and provide secure distributed access to this data; and (3) to facilitate real-time operational use of this data by environmental scientists at geographically distributed locations.

2.1.1 Real-Time Tasks by People and by Instrumentation

In the REINAS project, a real-time goal of the system is to obtain and display to scientists the state of natural phenomena as they develop. This is important in order to provide feedback so experimenters can make corrections and reset the sensors and refine measurements in real-time. The term "real-time" was used, but not defined, in the original call for proposals. In REINAS it must be interpreted in the context of meteorology and oceanography. In this context "real-time" is primarily equivalent to Van Tilborg's "immediate" on-line data acquisition [vT91]. Today, many of the observations made in environmental science are at long intervals, where "long" is in hours. The processes going on at the air/sea interface may change more rapidly than those involved in deep sea processes. "Real-time" in this context is also constrained by how fast the data sensors collect data. For example, one of the faster sensors is the LIDAR which requires about 2 minutes per 180 degree scan and gathers 2400 data samples per second [Lid89].

In the context of systems, "real-time" is used where correct operation or control depends on "carefully orchestrated interactions with time" which is also related to some of the elements of this system. Some processes may be interdependent and time critical, such as the estimation of direction or locality of a developing phenomenon and the control or steering of an instrument to refine the data acquisition in that area. However, in our case there are no apparent issues of safety, as in aircraft control. The control is that of instruments, whose purpose is to refine the estimate of the state of the process, and not for the control of any physical phenomena.

In real-time the system must support recognition (or automatically recognize) the development of "interesting" phenomena and direct controllable sensors to acquire observations at higher time/space resolution. This characteristic comprises the closed-loop part of the REINAS system (and is one which differentiates it from other similar systems). Automatic recognition of phenomena, and prediction of sensor output, may permit the compression of the data flow from acquisition systems, enabling more intelligent data acquisition and management without compromising the resolution of the target phenomena.

2.1.2 Data Types and Rates

With the help of the environmental scientists at NOAA, NPS, and MBARI a detailed investigation is being made of the data rates and characteristics of sensors (such as those mentioned on page 8 and also on pages 73–76 of the proposal). Many of the sensors have variable data rates depending on the particular experiments being done. We are now developing detailed characterization of data sources: point observations, time series, transects, vertical profiles, grids, satellite images, RADAR output, etc.

These source data rates will serve as inputs to the system model to be developed, and will provide the basis for quantifying the data flow, and timing constraints, for this model. The system to collect, process, and present information will be sized to accommodate the data rates currently proposed. We fully expect these to evolve during the life of the project, and the system design is to accommodate this evolution.

Another characteristic of the observed data is that the noise content must be properly determined. This can be achieved, for example, by the correlation of measurements from different instruments. Some means must be developed to present the quality of the measurements in the visualization of the data. Visualization can then aid the oceanographer in understanding the marine interactions for models of the scale of the Monterey Bay area.

In general, data are not sampled at regular spatial locations; the processes being measured may have vastly different time scales; and the data transfer rates may vary by as much as five orders of magnitude. Complicating matters is the fact that there is no uniform physical data format nor logical data definitions.

The size of the physical data fields (wind velocity, temperature, etc.) may be modeled on a variety of scales. The transient behavior of the seabreeze is an important aspect of its development. The ability to "see" the physical parameters through visualization is a key tool. The rates of data from various instruments, and the rate of change of different phenomena under observation, translate into requirements for a storage hierarchy. In the REINAS system we expect data associated with currently developing phenomena will be kept in fast memory, while synoptic data in slower larger memory. We will explore algorithms for storage management to automatically move data down the hierarchy as new data arrives from the sensors.

Finally, the ability to visualize data in real-time as they come in will provide the motivation for dynamically controlling the data gathering process. This will involve instrumentation issues as well as artificial intelligence issues so that the detection of interesting phenomena and the subsequent steering of sensors can be effected.

2.1.3 Meteorology and Oceanography

One of the most unique problem characteristics that REINAS will address is the integration of data from the meteorological and oceanographic domains. The nature of the integration will not only provide single-channel access to data from both domains, but will assimilate them into a time/space continuum. Although REINAS will be designed to accommodate the study of phenomena on many time/space scales, the local seabreeze effect was selected as the demonstration phenomena for REINAS verification.

The seabreeze problem was selected after careful study as an ideal topic. It occurs regularly, i.e. it is not a rare phenomenon. It is not well understood in detail. It is of operational importance to the Navy. Monterey Bay (where our science partners have their operations) is a good site to make measurements.

The seabreeze is a critical component of coastal meteorology and oceanography. General descriptions of this phenomena are found in Atkinson [Atk81] and Hsu [Hsu88] while numerical modeling studies are summarized by Pielke [Pie84]. A survey of climatological studies of seabreezes by Atkinson [Atk81] illustrates mean surface characteristics of a seabreeze event. Although meteorologists have developed a basic understanding of the seabreeze mechanism, most of these investigations have used primarily surface data. In comparison there are fewer studies of the depth of the seabreezes, the return current and the interaction with the larger scale flow.

Most of the seabreeze studies have occurred over flat terrain. Many studies have been conducted along the Gulf of Mexico and Florida coasts and in the deep tropics. Many of the mid latitude studies have occurred along the Great Lakes. Johnson and O'Brien [JO73] studied a single event along the Oregon coast, and Schroeder et al. [SFCO67] studied coastal circulations of Southern California. Since these studies used only raobs and surface stations, only limited data on the vertical structure and temporal evolution of the seabreeze have been collected.

A popular information display method used in oceanography is contour plots of isolevel values of different variables. There is minimal use of color. For example, results from numerical simulations produce a volumetric data set. However, the data is not displayed in its entirety with volume rendering as in other areas of scientific data visualization. Instead, contour plots of selected orthogonal slices are used. Occasionally one may want to correlate simulation results with data gathered from a ship. Unfortunately, ships do not necessarily traverse in straight paths that lie on the same orthogonal slice. In this situations, one can take advantage of 3D reconstruction techniques used in medical imaging to obtain an oblique/curved cross-sectional view of the data set [RAT$^+$85].

Within the oceanography and meteorology group, a popular method for displaying vector fields, such as current flows, wind fields, mean transport terms, and turbulent fluxes, is with an array of arrows in varying directions and lengths. Again, there is minimal use of color to overlay other information on top of the arrows. The use of glyphs [Cox88] may also prove beneficial in this area. Tracer particles which are commonly used in Computational Fluid Dynamics experiments are just starting to be used in this area. Integrating data from both domains must account for differences in vertical locators, directional parameters, units for similar parameters, temporal latency of signal propagation between the atmosphere and the ocean, etc.

2.1.4 Visualizing Observations and Models

The requirement that real-time data must be captured, assimilated and displayed imposes a strong constraint on the choice of visualization techniques that can be considered. The real-time requirement is specially crucial for the area scale of study such as the Monterey Bay. Although most in-situ measurement devices are not remotely programmable to alter the frequency and area of data collection, there is no technological constraint for not doing so. The primary reason is the lack of motivation for providing this capability to the sensors. Being able to visualize real-time live data coming from the sensors will provide such a motivation. During the initial stages of this research, "hooks" will be provided to allow the scientists to view live data and to manually control the data collection process. A logical extension of this would be to implement smart, statistically based, analytical tools that would automatically detect and alert sensors to the onset of interesting occurrences.

Many different visualization techniques can be exploited in our current problem of visualizing fused atmospheric and oceanographic data. Oceanographic data in particular are usually sparse and visualization can help in creating a whole picture that provides more information and insight than the individual parts. In dealing with a wide array of data sources, types, rates and location, coupled with time constraints from continuously monitoring data sensors, straightforward application of known visualization techniques may not always be suitable.

In addition to data from different measurement devices (e.g. SODARS, CODARS, LIDARS, tide stations, buoys, satellite, etc.), the visualizer must be able to handle outputs from numerical simulations of models as well. Data from different sources must be intelligently fused together to provide relevant visualization and analytical results. One of the end results of this endeavor should be the fostering of communication and sharing of data between oceanographers and meteorologists. That is, both data sets must be correlated and registered so that the visualization environment can support simultaneous display of both sets.

Unlike data-rich Computational Fluid Dynamic simulations, observed oceanographic and atmospheric data are usually data-poor. Thus standard visualization techniques used in visualizing regular grids cannot be used. Because data sensors are expensive, proper judgement must be made in locating these sensors to maximize the quality of collected data for the experiment at hand. In general, the spatial distribution from which data are collected is sparse. On the other hand, data are continuously being collected at these points. Being data poor creates special challenges for visualization. In order to fill holes where data is not available, some form of generating data using models must be performed.

When dealing with visualization issues, one will have to address various types of data. For example, data can be raw sensor measurements, observations derived from measurements, observations derived from other observations, derived from analysis, or computed by simulation. Data can also be spatial, temporal or spectral. Ideally, the characteristic or classification of the data should be transparent to the user (but accessible by the user)

2.1.5 Integration of Current and Historical Data

Another unique problem that REINAS will address is the seamless integration of near real-time data and historical data. This integration will be required to support the visualization of current data against a historical perspective; to support the automated recognition of target phenomena based on past feature characteristics; and to support the use of the system for retrospective research. The architectural characteristics of a system to support real-time objectives may vary significantly from one to provide optimal support for retrospective research objectives. Therefore, REINAS may be designed as a bimodal system, which operates to optimize real-time performance during times of "interesting" phenomena development and to optimize response to ad hoc data access and display tasks during "slow" periods.

2.2 High Level Requirements List

This section gives an overview description of the major functional areas of REINAS. In general the goal of REINAS is to be a real-time system which can provide:

- Sensor control (and steering where possible)
- Real-time data acquisition (load path)
- Access to large archives of data in an intelligent fashion
- Visualization of oceanographic and meteorological phenomena

There are four major components of the engineering side of the REINAS project: instrumentation and networking; system architecture integration; intelligent data management, and visualization. The following hierarchical list gives requirements for the major functional areas. Each requirement describes a functional or performance need and includes the criticality and relative priority of the requirement. In addition, known quality attributes and development constraints are listed with each requirement.

The hierarchical format allows the specification of requirements at several levels of detail and supports the process of refinement as more is learned about the REINAS problem domain. This section of the document is expected to evolve throughout all phases of the project life-cycle. Requirements are defined for the three main areas of Data Acquisition and Delivery (DA), Data Management (DM), and Visualization (VS).

2.2.1 Data Acquisition and Delivery Requirements List

These requirements define the kind of acquisition devices and delivery environment that REINAS must accommodate and integrate into a seamless whole. Detailed requirements for interfacing to external data sources will be defined in the next phase of the project.

DA 1: Data Sources:

Data must be accepted from many diverse sources at greatly varying rates.

DA 2: Methods of Obtaining Data:

The methods of obtaining the data may vary greatly depending on the instrument. (FTP, shared file system, RPC, etc.)

DA 3: Data from Models:

Output from numerical models must also be accepted.

DA 4: Data accessible by Attributes:

Data must be easily accessible to the scientist, based on a wide variety of attributes which will not be fixed, but will expand over time.

DA 5: Data Compression:

The system should include support for lossless and lossy compression of data, as appropriate.

2.2.2 Data Management Requirements List

These requirements define what must be provided to accomodate the effective use, operation, and maintenance (monitoring, fixing, enhancing, and extending) of the data management component of the REINAS system. This function must accommodate the management of and access to data and information associated with the current, historical, and predicted geophysical environment within which REINAS operates. In addition, REINAS data must track the state of acquisition equipment and support the process of instrument reconfiguration. Data management will have the goal of making the storage transparent. In this sense Visualization is a client. The scientist should not have to know there is a storage hierarchy, or even where the data is stored.

DM 1: Data Management Usage:

To provide easy access to data and information services for REINAS clients. A range of clients are expected to interact with REINAS including automated instrument and computing processes, humans performing operational forecasting in the marine environment, and oceanographic and meteorological scientists performing retrospective research. Real-time clients will be given higher priority in the specification, design, and implementation of REINAS.

DM 2: Data Management Administration:

To provide inherent functionality, services, tools, and products to users who are responsible for the development, operation, and maintenance of the data management subsystem of REINAS. Database application development tools, performance monitors, and database design and evolution tools are included here.

DM 3: Data Management Operational Environment:

The data management function must operate in an environment that spans geographically dispersed system components that are implemented using diverse technological platforms (i.e. a distributed, heterogeneous system). In addition, REINAS will not have complete control over all system elements. Individual nodes in the REINAS Data Management Subsystem may operate to meet the mission and needs of local users as well as REINAS users. Therefore, REINAS must effectively integrate nodes with autonomous control.

In addition, basic characteristics of typical DBMS systems (integrity control, security enforcement, recovery, concurrency control, distribution of data, development tools, etc.), must be combined with characteristics of Real-Time systems (e.g. guaranteed response time). Other operational issues include generic data object types, physical storage and access mechanisms, and data exchange mechanisms. Within data exchange is included:

- Format and content: What is included - general semantics and meaning of items.
- Structure: How information items are related to reflect the general and specific semantics.
- Representation: How information items are mapped to real world values.

2.2.3 Visualization Requirements List

Research in visualization issues will be directed at supporting the overall mission of real-time environmental data management. Specific goals are to:

- Integrate presentation of data from wide array of distributed sensors
- Handle both real-time and historical data
- Provide a common visualization environment for meteorologists and oceanographers
- Provide an interface for remote access of the environmental database for other researchers
- Provide for progressive refinement
- Integrate data quality into visualization.

From a user's point of view, the visualization component is the main interface to the data and computations. From the system point of view, the visualization component is the main client of the data management system.

Research in visualization issues will be directed at supporting the overall mission of real-time environmental data management. Specific goals include:

VS 1: Support visualization needs of both oceanographers and meteorologists.

- Handle both ocean and atmospheric data sets.
- Provide coherent view of ongoing physical processes.
- Provide mechanism for selecting different data sets of interest.
- Provide user interface for remote access of environmental database.
- Provide the means for data exploration.

VS 2: Support multiple visualization stations, some may be on-board ships.

- Visualization stations, depending on their graphics capabilities and communication bandwidth, may be geographically dispersed.
- Minimal configuration should include keyboard, mouse and bit-map screen displays. Also need appropriate communication setup.

VS 3: Support bi-modal visualization operation.

- Visualization system may be used for viewing live data as they come in, or in retrospective mode in viewing historical data.
- Tradeoffs will have to be made in terms of graphics quality versus response time.
- Provide image storage/retrieval of visualization products.
- Provide progressive refinement option for high-volume data sets that require interactive speed interactions.

VS 4: Integrate data quality into visualization.

- Provide options for displaying data quality in various ways, e.g. to highlight confidence level of data to the users.

VS 5: Integrate presentation of data from wide array of distributed sensors.

- Provide data fusion tools that aid scientists in understanding the interactions among different data sets.
- Handle various data formats, types, characteristics.
- Generate appropriate graphics display for various data types.

3. REINAS Technical Direction

3.1 Description of Technical Approach

The REINAS project is a blend of computer science research, environmental science research, and computer engineering to create a prototype of a system for daily operational use in the Monterey Bay. Ultimately, the REINAS research prototype should provide a basis for the design and development of operational systems which can seamlessly integrate sensors and processing resources. Systems based on the REINAS research prototype should find applications in multiple organizations and in diverse operational environments and should exploit successive new generations of technology.

During system specification, a broad user community will be surveyed to identify, define, and document detailed requirements and constraints. In addition a period of critical prototyping will provide information to evaluate the current technology base against known requirements. On the basis of detailed system requirements and evaluation of critical technologies, a baseline system architecture will be refined as a basis for the development of the operational prototype.

In the design stage, alternatives will be evaluated in attempting to meet the identified requirements. One basic assumption is that the scientific users are computer literate but not necessarily expert programmers. Therefore, the system must require minimal amount of programming and generally be easy to use. Other design parameters are flexible architecture that permits change or adaptability, allows users to add new data sets or new functionality within the same software framework, or to generally enhance understanding of the data.

3.1.1 Use of Existing Technology

The REINAS development philosophy has three elements. First, the system will take advantage of existing off-the-shelf and mature laboratory technologies where appropriate. These will be combined with innovative developments in computer systems, data visualization, data management, data compression, and pattern recognition. Research on new storage technologies, communication networks, file systems and database management system technologies will be avoided to ensure maintainability in an operational environment as well as timely completion of the project.

3.1.2 Software Engineering Techniques

The second element of the development philosophy recognizes the need for some disciplined software engineering techniques. The initial REINAS prototype, or at least some of its components, is intended to be transferred to the environmental science community as operational tools. Therefore, the software developed should be readable, well documented and maintainable. Methods and tools will be used which support the processes of requirements analysis, design, and code walkthroughs.

3.1.3 Standards

The final element of our philosophy calls for the identification, adoption, and/or the development of standards for data acquisition, data management, and data visualization to support inter-operability in the distributed, heterogeneous, autonomous environment that REINAS must operate. Some areas where standards are expected include: high level access and control protocol for sensors/instruments to be linked to REINAS; a common information model for data, processing, and sensor resources available to REINAS; and a standard data exchange protocol for objects which must move between system elements.

3.2 System Quality Objectives

In order to meet operational requirements there are several system quality goals and associated objectives which must be met. REINAS must be portable, extendable, and modular. The following lists identify specific objectives for each quality goal.

3.2.1 Portability Objectives:

- The REINAS system must be transportable to different geographical contexts other than Monterey Bay.
- The system must be configurable to study phenomena on different time/space scales from the flagship seabreeze problem.
- The system must be able to work in different operational environments. For example, the system must be scalable, up or down, to work in ship and shore based operational scenarios.

3.2.2 Operational Objectives:

- The system must be maintainable
- The system must be used by people who have limited computer literacy.
- System operations must be guided by simple procedures.
- Users should be able to extend some of the products produced by the system without expert intervention (data queries, analysis procedures and algorithms, display options).
- System will operate in two modes, one to support real-time operations, the other to support retrospective research.

3.2.3 Extendability Objectives:

- Ability to add new functionality.
- Ability to add new data sources.
- Ability to add new users.
- Ability to add new data products.

3.2.4 Modularity Objectives:

- Major functional elements of the system should be reusable in alternate contexts.
- Major elements should accomplish one well-defined purpose (strong coherence)
- Modifications to individual elements should not cause major and/or undocumented side effects (loosely coupled)
- New elements must be able to be added to existing elements to achieve new or enhanced functionality.

3.3 Technical Components of REINAS

3.3.1 Real-time Data Management

The primary requirement for the REINAS project is to provide real-time data management with capabilities for viewing such data. It is not clear a priori what the best approach is. Some of the considerations include the fact that data is collected for processes that have different time scales. Depending on the relevant time scale, the scientists may be interested in different processes. Perhaps some kind of variable resolution data filter can preprocess the data. This conjures up a data-driven model where data from sensors are fed to data filters which in turn does some form of resampling, conversion, or more generally some data transformation, and are in turn piped into the data management system [BC92]. The visualization component then interacts with the data manager to extract relevant data to visualize through another data filter.

In situations where the data feed is too high, the scientist can interactively alter the resolution of the data filter either to a lower spatial resolution to selectively discard data. Alternatively, the spatial resolution can be maintained at the expense of the temporal frequency of the data, or the scientist may want to selectively window in or focus on a smaller area and still keep the same temporal frequency. Note that what is suggested here are just some considerations that we will examine. The design phase is an evolving and iterative process that must keep up with changing requirements and technological advances and must therefore be as flexible as possible.

3.3.2 Database

The database is the focal point of the engineering aspects of systems architecture, instrumentation and networking and visualization. Environmental data are usually sparsely distributed in space. However, each site may be a rich source of data. We plan to investigate optimizations in storage "hierarchy" by exploiting aspects such as spatial/temporal coherence of the data.

The visualization part must graphically present the data stored in the data base to the users. An easy-to-use graphical user interface with at least the capability of X-windows would assure remote access to the database by a wider group of scientific users.

Whether it is for data browsing or real-time visualization, an option for registering the data would facilitate interpretation. This can be inexpensively achieved by incorporating a static geographic database where data can be positioned.

3.3.3 Data Compression

We have identified three major areas where introducing compression into the system can provide significant benefits. These alternatives need to be considered and discussed with system users, to find where compression can do the most good.

A promising area of investigation concerns the compression algorithm for the archival copy of the data. Ideally, the compressed copy could also be the copy used for real-time storage on the on-line disk, with a decompression step employed prior to use of the data. The investigation includes studying the characteristics of the various types of data and studying trade-offs of compression gain versus commonality of algorithms versus real-time requirements for various types of data. Many of the data points are on an irregular grid, and the data location as well as the data values may both require compression.

Another interesting area of investigation is the compression of data "screens"' intended for scientific visualization of results. The processor that generates the image to be placed on a screen may be remote from the viewing station. Viewing scenarios and configurations suitable for the approach would be studied. A project in this area could involve the design of algorithms for transmittal to, and decompressing at, the user's workstation immediately prior to display. Successive views might be able to take advantage of data already at the workstation, and only differences transmitted. If animation is used, or data objects are rotated, then motion compensation and other moving picture compression techniques may reduce the overall traffic across a network.

Alternatively, if the workstation has storage space and can render the 3-D data for visualization then the larger body of data could be sent compressed from the file server, and decompressed at the workstation. Within this area, we also consider compressing the result of a scientific computation model applied to an experiment, for storage. Later, if the model is to be compared against other results, the decompression algorithm may recreate the data faster than recomputing it according to the model.

A third area concerns compression at the acquisition site, e.g., a wind profiler, that may be under computer control, prior to transmission of the data to a central location. The purpose of compression would be to reduce the communications bandwidth requirement of the data. For continuous measurement, this could mean recording data at a higher frequency than permitted by the bandwidth constraint and no compression.

3.3.4 Analysis and Modeling

Data analysis and modeling forms an integral part of the visualization process. Most of the analysis routines can be viewed as data transformation functions that convert data to other forms. It is then the responsibility of the visualization component to graphically present these derived data. Thus, in addition to acting as filters, data filters can be more general and may perform space time conversions, statistical interpolation and others. Specifying the required transformation is conceptually equivalent to designing operators or functional filters that operate directly on data sets. These filters can be generalized to functional networks where paths for data flow are specified as links, and nodes contain simpler filters. To accommodate different data transformation requirements, it is worthwhile to use a modular and hierarchical building block approach. These modules can then be combined to build more sophisticated filters using a functional network.

3.3.5 Pattern Recognition

It is easy to see how a real-time visualization of a simulation study can have a feedback loop into the model itself. This capability allows scientists to explore the effects of different conditions and parameters on their models, and to ultimately improve their models. Design of the visualization system must also provide the flexibility of allowing a feedback loop between visualization and the data collection process. Being able to dynamically view and analyze incoming data streams equips the scientists with information on interesting processes occurring in the environment. The human can then send control information to data sensors to modify their data gathering instructions.

From this point, it is a logical next step to automate this process so that the system can automatically recognize interesting features. With this option, the scientists will make more efficient use of the scarce data gathering equipment.

3.3.6 Visualization

It is envisioned that the visualization software will be X-based so as to run on numerous hardware platforms. With the graphics capability of Silicon Graphics (SGI) machines and the opening up of its graphics library (opengl), we are doing initial graphics development on the SGI platform.

The visualization component must handle various data types and formats; and must also accommodate for new sensor types as new technology are developed. As such, the software design dictates a modular, extensible approach. One existing technology on the SGI platform is the Explorer (currently at beta release 2.0) that meets these requirements. We will be developing our software around this existing technology.

Oceanographers and meteorologists may have different visualization needs; there may be situations when the scientists themselves do not know what they are looking for in their data set. Thus, we plan to support data exploratory type interactions. To accomplish this, different data filtering, transformation and rendering tools will be provided. These tools may also be tuned to seek out relationships among different data sets. The same tools can be used to handle data from different sensors. That is, data go through appropriate data filters and transform blocks before they are turned into graphics objects for rendering.

Existing visualization techniques such as surface, volume and flow visualization will be used. Novel methods will be developed as necessary. In particular, there is a need for a coherent metaphor or paradigm for the user to keep in mind when interacting with the visualization software. This is to avoid the illusion that the user is dealing with 100 separate programs behind the nice graphics interface.

Various ways of displaying noisy or missing data or data quality will be explored. Examples, include highlighting bad data above certain threshold; using fuzzy representation, so that good data are more opaque and bad data vanishes (graphically) as their confidence level drops; etc.

High volume data sets can be viewed at lower resolution to maintain interactive speeds. That is, progressive refinement of the image will be an option when image quality is more important than speed. Alternatively, exploratory type interactions to focus the area of interest within high volume data sets can significantly reduce the amount of data to be visualized and hence maintain interactive speeds.

3.4 Description of Applicable Technology and Research

3.4.1 Database Technology

Relational databases can be extended to include image data as well as logs of sensor data. Relational features can be used to reduce the total amount of data that need be transmitted to the user workstations.

The relational database will be used in conjunction with AI techniques to allow patterns/behavior in data to be quickly identified. This would enable the scientist to selectively collect data either at a desired rate (thereby remaining within the real-time parameters), desired direction or desired resolution.

An example of a readily available data set is the NOAA Seabeam bathymetry data of the Monterey Bay. This can be used together with a geographic information system (GIS) such as GRASS, MAP II or ARC/INFO, upon which data can be overlaid. We are evaluating Postgres (Sequoia uses it) [Sto91], and commercial database systems such as Oracle and Sybase.

There are many other database topics, such as data security, data quality, data correlation, spatial query, fault tolerance, and ownership of data, which are well-known and will be considered in the architecture of the REINAS system.

3.4.2 Data Compression

There are existing compression programs such as Unix compress which can be used in REINAS for experimentation. Tradeoffs need to be studied on the advantages, if any, of custom-designed algorithms specific to the data types.

If data acquisition is from smart instruments controlled by embedded microprocessors, an opportunity is presented for data compression at such user-programmable instruments or sensors. For slowly changing data values, an average value for a given time period may be sent, along with the highest and the lowest values seen in the period. The highest increase and lowest decrease in value between successive samples can also be stored. Data acquired where a computer is present could also be capable of signaling the occurrence of a significant event, such as when the sensed value crosses a limit or satisfies some other prescription requiring special treatment or recording of a time and event number.

3.4.3 Systems and Software

Commercially available hardware and software is evolving at a rapid rate. A key design consideration is to build a system which can be expanded to include new sensors or new computing equipment as they become available. Handling the timing requirements, desired "real-time" instrument control, as well as the diversity of data rates and variety of distributed processing, will most likely require some significant extensions to the operating systems chosen. This is an area for detailed exploration.

Commercial systems have several deficiencies. Real-time support, though it is available in some kernels, is not available for the kind of system that we envision. We know of no commercially available system that provides the required facilities for the task of monitoring and steering the instruments. Management of a multi-level store appropriate for handling the mass of data in this application, also is an apparent deficiency in commercial offerings.

Protection mechanisms for the sensors and data also appear inadequate. Applicable existing programming systems which will most likely be used in REINAS are:

- Standard Internet (for now TCP/IP) protocols.
- Remote procedure call (DCE most likely); perhaps a hint of object-orientation.
- Distributed file systems (AFS most likely: caching, replication).
- Capability-based security (Kerberos most likely). Perhaps PEM/MD5 for digital signatures, especially where there is high latency.
- Perhaps Isis for distributed applications, though concerned about the degree of connectivity.
- Multi-level storage hierarchy.

3.4.4 Operating System

We do not plan to write a new operating system because this would be a major venture, and beyond the scope of this project. Under consideration are a micro-kernel OS such as Mach or Chorus, which are UNIX compatible but which allow easy extension of their functionality. We plan to add to this micro-kernel our own storage hierarchy and protection mechanisms. We will also study using a tool, such as IDE for system design. OSF/1 is also a possibility.

We hope to use an existing database system, again developing any required extensions or "enabling" layer to match it to this problem context and to achieve required performance by such techniques as precompilation, complex object support, etc. The use of "object-oriented" methodologies will be stressed, and the language will probably be C++.

3.4.5 Visualization

In areas such as data modeling, compression, interpolation, and visualization, we also plan to evaluate, select and reuse as much of existing, appropriate software as possible. For example, visualization software like apE, AVS, Explorer, NCAR, JPL, offer excellent starting points. Among the things that need to be integrated is code to handle a combination of live sensor data, historical data and data from simulation models.

In recent years, there has been significant advances in visualization techniques. A major driving force is meeting the flow visualization requirements, usually in Computational Fluid Dynamics (CFD) experiments. Examples of techniques that came out of this effort include: particle traces along trajectories, streams and ribbons, [Hul90], [HH91] as well as volume visualization of field variables. These are usually incorporated in visualization packages such as: NASA Ames' Plot3d and Realtime Interactive Particle-tracer (RIP) [Lea87], [PCH$^+$90], NCSA's Research on Interactive Visual Environments (Rivers) [HM90], and University of Wisconsin's Man-computer Interactive Data Access System (McIDAS) [Hib89], [HS89] to name a few.

Another driving force in visualization is medical imaging. Non-invasive techniques for data gathering produce cross-sectional views that must be reconstructed to facilitate understanding of the underlying three dimensional structure. Techniques in viewing three dimensional data volumes have evolved from image processing and data conversion (e.g. extracting contour edges and surface tiling [FKU77]) to simplistic voxel based surface rendering [HL79], [LC87] and finally to volume visualization that simulates the amount of material traversed by a light ray in an opacity parameter [DCH88], [FLP89]. Some of these techniques have been incorporated to other application domains. For example, volume visualization has been applied to CFD data on curvilinear grids [WCA$^+$90]. Likewise, concepts from related areas may be fruitfully used in visualization e.g. caustic ray tracing [Wat90], ship wake visualization [Gos90], modeling of rainbows [Mus89] and SONAR range data display [SG89].

Since data come in different forms and from different disciplines, a convenient way of handling this inhomogeneity is through a data flow model. Once this abstraction is made, a user interface can provide a relatively easy framework for interacting with the data. Graphical user interfaces can then be built that offer a flexible, customizable solutions to a wide variety of applications. Two popular examples are the AVS system from Stardent [Uea89] and the apE system [Dye90] from the Ohio Supercomputer Center.

3.4.6 Artificial Intelligence Techniques

We intend to exploit patterns in the sensor data for multiple purposes:

- To enhance retrieval efficiency by improving indexing as in the self-organizing memory methods being used at UCSC.
- To improve existing knowledge of the data from an oceanographic perspective.
- As a feedback mechanism for the sensor control strategies.
- To achieve data compression.

Such patterns of activity can be discovered using traditional statistical pattern recognition techniques (such as clustering and factor analysis) and the symbolic techniques used in machine learning systems such as concept induction, explanation-based generalization (EBG) and even what is now called constructive induction which allows machines to create their own feature sets. EBG is a special form of goal regression which attempts to establish, through generalization, causal dependencies in the sensor data, that can then be exploited for creating expectations in the future.

Finally, the numeric technique of temporal-difference (TD) learning as used in neural net systems will allow feedback to be made about sensor decisions based on a time-oriented sequence of data. TD learning is one of the main techniques being used at UCSC in our adaptive-predictive search (APS) model which provides a mechanism by which decision systems can improve performance through experience. TD learning forms the basis for the reinforcement learning techniques that are being used successfully in problems of robot motion and planning.

What is important about these techniques is that they can proceed largely "unsupervised" without the burden of pre-classified training sets used in other forms of machine learning. Given the real-time nature of the decisions, it is crucial that the system be given as much responsibility as possible for both decision-making and learning to adapt based on the consequences of those decisions. For example, the AI branch at NASA Ames developed a program called AUTOCLASS that discovered several previously unknown classes of relations from the 5,425-record IRAS (Infra-Red Astronomical Satellite) low-resolution database [Den89].

In addition, statistical methods and multi-dimensional analysis tools such as the projection pursuit method and the parallel axes methods will be investigated. Though data may be spatially sparse, the frequency and the dimensionality of the data gathered may be very high. These techniques will aid in identifying dimensionality dependencies, as well as data trends, which can be used steer the data sensing functions.

3.5 Specification of Preliminary Architecture

The creation of the REINAS system is a challenge in computer systems architecture. The basic problem of system design is integration of a geographically dispersed sensor and user network, and combination of data into a data management, retrieval, analysis and distribution system. The objective is to provide a system that minimizes costs while maximizing flexibility and ease of use. Ease of use will require an interface to the real-time and archival data that makes this data appear local to the user's workstation, and not require access via complex protocols or sessions on multiple processors. Key parts of the system are the network of sensors and associated processors, called the "data delivery," and the information management system for storing present and past data, for control of access to data, and for presenting to users rapid access at their workstation of the relevant data.

Some of the design issues to be considered in this system are:

- Distribution of processing, data storage, data reductions and sensor control
- Making this distribution transparent to the users
- Providing appropriate access controls for data from participating scientists
- Networking and communication issues
- Designing a system that will scale to a large number of sensors and users
- Fault-tolerance

The single system image is a recent development in computer systems that provides great benefits to system users, because it hides the physical distribution of a system, and thus simplifies the development of user applications. Such systems have been described by Mullender [Mul89], Mullender et al [Mea87], Howard et al [HKM$^+$88], and Ousterhout et al [OCD$^+$88]. Existing technology makes it possible to distribute data processing so that with computing power local to the sensors, along with data archiving, analysis can take place either at sensor location or at the user's workstation as appropriate.

Different organizations cooperating in experiments may have data which they may not be willing to share with all possible users of the system, either for proprietary or security reasons. Techniques for limited access must therefore be implemented, as part of the data management component.

3.5.1 Design Alternatives and Decisions

Two issues which become important in a single integrated system are those of coordinating access to the sensors, and providing appropriate access protections to the data that these sensors provide (whether in raw or reduced form). Techniques will be developed and policies devised for equitably sharing sensor data. Further, support will be developed for remote controlled sensors, by a sensor-local microcomputer connected into the data network. This local processor can control sensor operation, perform some data processing, and provide local archival storage if appropriate.

In order to provide access to a large community of geographically dispersed scientists it is important that the system be built around a standard protocol suite. This will allow scientists who have access to the Internet [Com91] to access the data and (if they have the appropriate software) to control sensors. Transport-level communications will use a standard protocol stack such as TCP/IP [Com91] or OSI [Tan88] as deemed appropriate (e.g. by the funding agency). Building on these basic communication protocols, a set of high-level protocols will be created that will define the possible set of sensor-controlling processor interactions, as well as data transfer and access protocols. By using well-defined protocols in the system design, when new components become available they can be integrated into the system without affecting the other components.

Allowing for growth or scaling of the system is another important design consideration. Hierarchical structuring techniques will be used to ensure that the system will scale to as large a system as required, and to avoid inherent size limitations that can result when small and locally optimized systems are assembled into one large system. Important issues here are those of naming, object location, and access control [LPS81]. Another issue related to scalability is the data rates of the sensors. Though not unmanageably large now, sensor data flows will increase in the future as new sensors come on line and existing sensors are used more frequently. The system must be designed so there is are no data bottlenecks. Instead, the system will be composed of a set of communications paths such that when a path becomes saturated it can be replaced without affecting the rest of the system.

3.5.2 Architecture Description

The storage system for the data will be a multi-level hierarchy. As the data ages, it is used less frequently and so will move to slower (and significantly less expensive) storage. The benefits of caching are well-known [Smi85], [NWO88] and [GM90]. Studies have shown that by keeping frequently accessed data in a small, fast memory (with respect to the next level down), significant performance gains can be achieved. A possibility is a four level storage hierarchy. The first level would be memory of the computer (RAM), which has access times in the microsecond range. By making this memory sufficiently large, (and the changing economics of memory cost continue to be dramatic) most data access requests can be serviced directly. The second level in the storage hierarchy would be magnetic disks. These devices have access times in the millisecond range. Advances in storage devices such as disk arrays (RAID) [PGK88] are expected to vastly improve the performance of this level of the hierarchy. The third level of the hierarchy would be read-write optical disks, with access times which range from milliseconds (if the desired platter is mounted) to ten or more seconds (if it must be fetched by the platter robot). Recent advances in optical storage have greatly increased its density. The fourth level of the storage hierarchy is for archival data that is not expected to be accessed for long periods, and would be a high-density tape system, which has access times that range from a few seconds to a few minutes but virtually unlimited capacity.

Finally, fault-tolerance is a critical issue. Sensor data should not be lost if at all possible: the failure of a central data facility should not prevent data collection and processing. By constructing a distributed system where the data can be archived and partially processed locally, no single point of failure will exist. In this way, the system as a whole will be able to continue to operate even in the face of multiple failures (albeit in a degraded mode if sensors fail). The difficult problem to be addressed is to design the system in such a way

Figure 3.1: Sensors and their Specifications scheduled by MBARI

that a consistent view of the data is provided to the users [DGMS85], and that no data is lost or corrupted if at all possible.

3.5.3 Data Delivery

The objective of the data delivery subsystem will be to accommodate continuous, coordinated transmission of measurement data from a suite of sensors diverse in type and location, and to deliver the data to the storage and computation elements of the system. Sensors must also be able to receive and respond to sensor control commands from other system elements in order to support an intelligent sensing capability. The exact bandwidth requirements for sensor-to-shore transmission will not be completely determined until requirements analyses are completed during phase 1 of the project. However, an initial examination of the expected data sources in Figure 1 indicates that the overall throughput requirement can be easily handled with current communications technology. During times of high band-width burst transmission, especially from high rate sources like ADCP and LIDAR, caching data at the receiver and then using data compression techniques will be necessary.

The primary elements of the data delivery system will consist of sensors (described elsewhere), intelligent controllers which control one or more collocated sensors, ship based preprocessing and communications capabilities to integrate ship, ROV and other mobile subsurface sensor sources, preprocessing and routing processors located at each institution contributing to the system, and an inter-institution communications processor. This work is largely a contribution from ongoing MBARI programs, enhanced to support REINAS.

Presently MBARI has begun outfitting its Research Vessel Pt.Lobos with communications gear. Figure 2 presents a possible architecture for the data delivery subsystem. The actual architecture could change based on the overall system architecture defined by REINAS. As distributed database techniques are the key to providing a consistent view of the data, these techniques will be borrowed from the area of federated databases. The existing MBARI Scientific Data Management System provides a useful starting point.

3.6 Comparison with Existing Environmental Data Acquisition, Management, Analysis, and Visualization Systems

Any research and development effort should proceed within a well-defined context. The context should describe the problem to be solved and an approach which identifies basic principles for development and criteria for success. In addition, the effort should be compared and contrasted with previous and contemporary efforts to identify potential collaborations, reuse opportunities, potential pitfalls, and differentiating characteristics. Previous sections provided a problem description and technical approach. This section compares and contrasts the REINAS research and development effort with other system projects.

A system is composed of integrated parts which are designed to work together to perform a well-defined function. The system will perform the function within a context defined by the people, procedures and practices, and other systems with which it must interface. A complete comparison between system projects would include evaluations of problem contexts, development approaches, and system characteristics. At this stage of the REINAS project there is not a clear picture of the eventual structure of the system. Therefore, our analysis will focus on comparing the problem context and technical approach of REINAS with other system projects in the environmental data management domain.

The included figures identify the systems/projects that were analyzed, details of each comparison, and a map comparing the scope and specificity of the projects with REINAS. Projects and Systems compared are: AWIPS/DARE/PROFS (FSL PC Workstation) [USD91], STORM [NCA92], MADER [COA92], EOSDIS (GCDIS) [NAS92]. Also included are NEONS [TJ92] [Jur91], UNIDATA [MG92] and GLIS (GCDIS) [MG92]. SEQUOIA 2000 [Sto91], and MBARI Scientific Information System [GBD$^+$89] are also described in the appendix. Other systems which should also be compared are the Norman System, CGIP, and Copernicus.

Accurate placement of projects on a two-dimensional map was difficult due to the multidimensional nature of the comparison. However, the map should give the readers a rough idea of which systems could be components of REINAS and which could absorb REINAS as a component technology. Relative specificity depicts how easily the system or project could be moved as a whole into another context (e.g. different geographic region, to study different phenomena).

Figure 3.2: REINAS Program System Schematic

```
                 SYSTEM SPECIFICITY vs RELATIVE SCOPE
                 ------------------------------------

         -----------------------------------------------------------
        |                   |                   |                   |
        |                   | Sequoia           |                   |
GENERAL |                   |                   |                   |
        |                   |                UNIDATA                |
        |                   |     REINAS        |                   |
   S    |                   |                   |                   |
   P    |                   |                   |          NEONS    |
   E    |                   |                   |                   |
   C    |                   |                   |  AWIPS            |
   I    |                   |                   |                   |
   F    |                   |    MADER         MSIS                 |
   I    |___________________|___________________|___________________|
   C    |                   |                   |                   |
   I    |  GCDIS            |                   |                   |
   T    |                   |                   |                   |
   Y    |   EOSDIS     NWS  |       GLIS        |                   |
        |            Upgrade|                   |                   |
        |                   |                   |                   |
        |                   |                   |                   |
        |                   |                   |                   |
        |                   |                   |                   |
SPECIFIC|                   |                   |                   |
        |                 STORM                 |                   |
        |                   |                   |                   |
        |___________________|___________________|___________________|

              Superset            REINAS              Component
              Efforts           Isomorphic            Technology

                   RELATIVE SCOPE (Comprehensiveness)

                                       General -  Portable as a whole
                                                  out of context

                                       Specific - Not portable
                                                  out of context
```

Figure 3.3: Project/System Comparison Map

AWIPS/DARE/PROFS		REINAS
Meteorology - hydrology data	----->	Meteorology & Oceanography o Integration o Fusion
Real-time & historical data	----->	Real-time & Historical o Integrated access o Fusion o Automated comparison
Real-time tasks by people 0 -2 hr forecast of severe storms	----->	Scientists can adjust measurement instrumentation as they watch phenomena develop in real-time and/or in fast forward (prediction)
Integrated display/visualization of data sources - Images, overlays, 2D + time	----->	Visualization o Images o Overlays o 2D + time o 3D o Quality Information

Figure 3.4: AWIPS/DARE/PROFS - REINAS Differences

AWIPS/DARE/PROFS		REINAS (cont.)
Displays of historical data	----->	Historical Data o Display with current data o Automated comparison (pattern recognition) o Support real-time and retrospective use (?)
Technical Charateristic o difficult to extend	----->	o User will be able to add new data products o Will work in other geographical areas

Figure 3.5: AWIPS/DARE/PROFS - REINAS Differences (cont.)

STORM Project		REINAS
System development part of a 10 year program of field study	----->	Not embedded within a single sci field study
o Multiscale (time/space)		o Multiscale (time/space)
o Fixed geographic theater		o Portable to other areas
Intensive Observing Period	----->	Intensive observing period can be initiated by:
o Determined by analysis of science team		o Analysis by scientists
		o Automated recognition and response
Permanent Operations Center	----->	Permanent operations
o Real-time control		o Real-time control
o Operational control of experiments		o Operational control (?)
		o Retrospective mode (?)
Data sources more diverse	----->	No plans for hydrological data
o Includes stream flow		

Figure 3.6: STORM Project - REINAS Differences

```
       STORM Project                                REINAS      (cont.)
       ----------                                 -----------

Data Management Functions       ----->       Data management function
                                             also includes additional
   o Collection                              support

   o Processing                                o Data mining

   o Storage                                   o Visualization data
                                                 objects
   o Cataloging

   o Retrieval

   o Dissemination


Technical Approach              ----->       Technical Approach

   o Bottom-up integration                     o Top-down integration
     via collocation of                          to data acquisition and
     display/analysis capabilities               delivery subsystem -
                                                 bottom-up integration
       + Effective communication                 within data
         channel by scientists                   acquisition/delivery

   o Data management paradigm:                 o Data management paradigm:
     data set catalogs (metadata only)           open data set content
                                                 to clients


   o Data Communications        ----->       Data Communications

       o Not tightly integrated                o Tightly integrated on
         (e.g. file transfer                     top of internet protocols
           via OMNET)                            and well defined sensor
                                                 data delivery/control
                                                 protocols
```

Figure 3.7: STORM Project - REINAS Differences (cont.)

SEQUOIA 2000		REINAS
Focus on Global Scale	----->	Mesoscale/local scale focus
o Global Circulation Models		o Fusion with climate
Real-time	----->	Real-time
o Communication of "Fast Forward" data sets to remote users o Phase 2: real-time collaboration		o Visualize phenomena as it develops o Adjust sensor acquisition based on environmental activity o Real-time and asynchronous collaboration
Technology	----->	Technology
o Phase 1 - centralized data management, distributed display		o Distributed approach cooperative processing (Heterogeneous) - Data management - Data display - Data processing

Figure 3.8: SEQUOIA 2000 - REINAS Differences

SEQUOIA 2000		REINAS (cont.)
Technology	----->	Technology
o Hardware (Fast I/O system via RAID, compression)		o Not building I/O system
o OS (File system for large data sets to storage heirarchy to mask access latency)		o Building off DCE Framework (heterogeneous)
o DBMS (Extend data types indexing optimizer)		o Data management - Data set metadata - Data set content
o Network (Fast (T3) guaranteed delivery)		o Networks - T3, compression - Standardized sensor access/control protocols
o Applications (Electronic repository Visualization workbench)		o Visualization - Data Fusion - Quality display
o Phase 2 Massive distribution heterogeneity, data fusion (DBMS - VIS)		

Figure 3.9: SEQUOIA 2000 - REINAS Differences (cont.)

EOSDIS		REINAS
Embedded in large program	----->	Building a generalized capability
o Focus on supporting EOS		
- Global data sets		- Mesoscale/local data sets
- Satellite priority		- Diverse platforms
- Flight operations		- Experiment operations
Backbone system for EOS data	----->	Complete system movable to other environments
o Archive		- Data acq/delivery
o Distribution		- Data management
		- Visualization
o Satellite product generation		
o Link to other systems		
Technology	----->	Technology
o Network: Two components		- Network built on existing internet infrastructure
- Internal EOSDIS (controlled for timely and seamless operation)		- Distributed cooperative processing via DCE
- External - user access via standard infrastructure		- Data set and/or data set content
o Distributed processing via agreement		
o Approach		
- Interoperable data set catalog		

Figure 3.10: EOSDIS - REINAS Differences

EOSDIS		REINAS (cont.)
Practices	----->	Practices
o Standards		o Standards
- Data exchange = HDF		- "Sensor bus"
o Policy		o Policy
- No period of exclusive use		- (?)

Figure 3.11: EOSDIS - REINAS Differences (cont.)

UNIDATA		REINAS
Atmospheric Data	----->	Atmospheric & Oceanic
o Complete services for acquisition, display, analysis		o Not complete services
		o Not global
- Link (real-time)		
- Software		
- Consulting and support		
Technology	----->	Technology
o Approach		o Approach
- Not building on commercial technology (NETCDF)		- Extend/use available technologies and architectures (DBMS, DCE)
- Uncontrolled distribution		- Controlled distribution
- Data set management and random access to content		- Data set management
		+ Data set content

Figure 3.12: UNIDATA - REINAS Differences

MADER		REINAS
Model for NOAA Data System Modernization	----->	Integrated system
o Real-time archiving (built-in QC)		o Real-time visualization (built-in QC)
o Single Channel access to data sets (internal and external)		o Single channel access to REINAS data sets
o Clearinghouse to all related data		
Technology	----->	Technology
o Approach		o Approach
- Distributed database using existing data centers		- Distributed database
- Catalog interoperability		- Integrated catalog
- Limited metadata query event time location		- Query via all metadata and data fields
- Standards		- Standards
BUFR/GRIB NetCDF		translators

Figure 3.13: MADER - REINAS Differences

NEONS		REINAS
Generalized data management platform for geophysical data	----->	Complete system
o Assimilate new sources o Unify access protocols		o Real-time acquisition delivery and control of sensor network o Visualization o Built-in data management clients
Technology	----->	Technology
o Approach		o Approach
- Build from mature and active commercial DBMS technology		- Build on mature and laboratory technologies
- Data set management abstracted to generic types (LLT, IMAGE, GRID)		- Data set management data set content
- Standards BUFR/GRIB		- More extensive data types required, e.g. visualized features, events
o Distribution supported via underlying DBMS (EMPRESS)		o Distribution supported via DCE, DBMS and transaction managers

Figure 3.14: NEONS - REINAS Differences

```
                              SUMMARY
                           - - - - - - - - -

  Potential "Collaborations"

                                              ----->   DBMS
     o SEQUOIA (Broad Project Overlap)        ----->   Hardware
                                              <-----   Visualization
                                              <-----   Data mining
                                              <-----   Distribution
     o Copernicus (?)


 Potential Reuse

     o NEONS (Component technology/approach)

     o MSIS (Approach)

     o UNIDATA (Component technology)


Potential Gaps (or opportunities)

     o Real-time sensor control/access protocol

     o Bimodal system  to  Real-time/Retrospective

     o Access via metadata + data content/source transparency

     o Meteorology/Oceanography fusion (Data management/Visualization)

     o Real-time visualization (3D + quality information)

     o Complete system which is portbale, extensible, maintainable
```

Figure 3.15: Summary

3.7 Expected Academic Research

As a university project one of the main goals of the REINAS project is to provide a continuing source of Master's and Ph.D. dissertation-quality research problems in computer engineering/science. Already we can see several topics which are worthy of dissertations in computer engineering/science:

One interesting systems research topic is to investigate the use of the present models of real-time computing for design and prototyping of systems for real-time data acquisition. Most of the literature of which we are aware on the topic of modeling of real-time systems is in the areas of process control or in the database concurrency/locking/update area. We are not aware of applications of these methods in the design of real-time data acquisition systems, which occurs in a variety of areas in addition to the present context of meteorology and oceanography.

Related dissertation topics include the development of tools for performance prediction, and for the evaluation of design trade-offs. With the combination of sensors, models, and past data, there are a variety of interesting topics that can be developed in system extensibility and in fault tolerance. Studies in data compression can be a key ingredient.

Several opportunities have been identified in data compression. The existence of effective techniques for probability estimation and for efficient encoding and decoding provides opportunities to discover new statistical model structures for increased data compression, such as for 3-dimensional data. The discovery lies in identifying strong correlations in the data and using them to better compact the data. Since compaction implies a "short description length" of the data, the discovery of correlation between different data sets also aids the research of cause-and-effect between temporally separated data.

Possibly the richest area for dissertation topics is in the development of visualization techniques for this application. Examples include the tools for rapid comparison of immediate measurements and historical or predicted data; visualization of the quality of data when new measurements are inserted at isolated points; tools for visualizing sparse data and for sampling an irregular grid; the means for combining historical, real-time and assimilated data; and a variety of data analysis and data quality presentation techniques.

Protocols for control and communication, and definitions of interfaces, to "glue" such a large system together, are also an area of research. We will have many heterogeneous components interacting, and this must operate as seamlessly as possible. Clean interfaces and simple protocols are essential if new sensors are to be readily added to the system. This include communication protocols with performance guarantees (c.f. Anderson's SRP protocol [AHS90]), to the interfaces between the various components of the system (c.f. Cheriton's work on the V kernel [Che88]).

Research into appropriate protection and naming algorithms for a large heterogeneous system is also an important research topic related to this project. Research into coordinating of distributed processing activities at the sensors and the various control and data management sites raises important research questions in performance, optimality and fault tolerance.

Although REINAS is designed to SUPPORT scientific investigations in oceanography and meteorology, not to propose new topics in those fields, we do expect research topics in these fields will be suggested and facilitated by the availablity of REINAS data.

4. REINAS Project Plan - Concept, Requirements, Prototype, and Verification

4.1 Operational Goal

The REINAS system will consist of three primary subsystems; data acquisition and delivery, data management, and data visualization. The data acquisition and delivery subsystem must link sensors of diverse types and operational capabilities into a seamless network which supports real-time sensor control and measurement delivery for intelligent sensing. The data management subsystem must be able to accommodate the storage and access of large quantities of current and historical data for acquisition, visualization, and human clients. The data visualization subsystem must provide an effective human-computer interface to system functions as well as effective visualization options for multidimensional data sets.

The ultimate system will be required to operate or be configured to operate within multiple operational scenarios. The initial scenario and one which will receive highest priority is characterized as follows:

- Instrument network in Monterey Bay region. Will include conventional oceanographic and meteorological data sources as well as advanced remote sensing sources from land, water and air based platforms.
- Real-time tasks of the system will focus on timely control (automated and human driven) of the data acquisition parameters of the sensor network, and visualization of phenomena as they occur. The flagship phenomena for REINAS verification is the seabreeze phenomena.
- The priority end user profile is one of an operational forecaster, who must assimilate all available observations, analyses, and model predictions in order to perform well defined tasks (to be determined).
- Retrospective use of the system by researchers who are performing general process studies should also be supported, if possible.
- Support for the automated assimilation of data into mesoscale models should be considered a high priority.

Other scenarios that must be considered include:

- Instrument networks in other geographical regions.
- Geographical theaters which vary in scale.
- Phenomena with different time/space scales. (This requires us to consider how to verify the system in a different context.)
- Optimize the system to support the tasks of (modelers, researchers, other operational tasks?)

4.2 Project Phases

The project will proceed in four phases leading up to the experimental use of an operational prototype. Successful evaluation of the prototype will lead to two additional phases resulting in the development and deployment of a permanent operational version of REINAS in the Monterey Bay. The following subsections describe the goals, activities, and products of each phase.

4.2.1 Concept Formulation

The goal of concept formulation is to develop and document a baseline description of the problem that REINAS is targeting, the technology to be applied, and a plan for meeting project goals. During this phase, the project team will investigate, analyze, and document the problem to be addressed by the REINAS system, applicable technologies, similar system projects, critical issues and potential barriers to success, and early thinking on the technical direction of the project. The products of these activities include this concept document, an initial system architecture, and a hierarchical list of high level requirements which provides a baseline for the next phase of the project.

4.2.2 Detailed Requirements and Critical Prototyping

The goals of this phase require the project team to fully document the functional, performance, and quality requirements of the REINAS system and to perform prototyping of critical implementation technologies. The products of this phase include a system requirements specification and preliminary design document which will be presented to the user community at a system requirements review. Successful completion of this phase must support architectural verification, and the detailed design and implementation activities of the next phase.

4.2.3 Operational Prototype Development

This phase produces identification and design of basic functional units of the REINAS prototype, implementation and unit testing of each unit, and successful integration into a full system. Results of unit and integration testing should be reviewed with the user community prior to experimental testing.

4.2.4 Experimental Verification

During this phase the sensor network will be dedicated to the measurement and analysis of the REINAS flagship scientific problem, the seabreeze. The system will operate in the context of other experimental ground truth activities to provide all the information needed to verify system compliance to requirements and to validate correct results. The final activity in this phase will be a review of the evaluation of system performance during the experimental period.

4.2.5 Operational System Development

Upon successful completion of experimental verification, and contingent upon extension of project funding by ONR, the project team will initiate work on the development of the fully operational REINAS system. Activities will include refinement of architecture, design, and/or implementation of the prototype based on feedback from the previous phase. A complete verification and validation test suite will also be developed to confirm system compliance prior to entering fully operational mode. Appropriate reviews will be conducted to examine proposed changes to requirement priorities, system architecture and implementation.

4.2.6 Operational System Verification and Validation

The goal of the final phase is the successful transition of the REINAS system into a daily operational tool in the Monterey Bay. Activities will include execution of a full operational test suite and evaluation by researchers, operational forecasters, computer scientists, and other user groups. The final product is a working system with appropriate documentation including requirements specifications, design documents, maintainable code and hardware, operational procedures, and user guides.

4.3 Project Schedule

The first two phases of the REINAS project involve the development of a concept design, followed by the definition of detailed requirements. Activities include studying input from expected users, characterization of instruments and data, evaluation of technical alternatives, and development of a preliminary architecture.

Integral to the design will be a model of the system that captures both the data flow and the timing requirements. Models can provide framework for system design and validation, for estimating computational and bandwidth requirements of the system components, and for evaluation of engineering trade-offs.

The third phase will begin detailed system design and implementation of the prototype system. Evaluation of the prototype will take advantage of models describing its function and required performance. Preliminary field experiments will also take place during this period before the full-scale testing of the prototype in the fourth phase. There will be a number of project reviews scheduled corresponding to phase checkpoints, as well as reports documenting plans and progress.

If the project is successful and if additional funding becomes available the next phases to be undertaken would be Operational System Development and Operational System Verification and Validation, mentioned in the previous sections.

5/15/92: OFFICIAL START of REINAS PROJECT

10/01/92 Begin Phase I: CONCEPT DESIGN and DOCUMENTATION

- **Characterize: Instruments, Data, Users and Uses of REINAS**
- **Create Project Plan / Hire Staff**
- **Evaluate Technical Alternatives**
- **Develop Preliminary Architecture**

1/01/93 Begin Phase II: DETAILED REQUIREMENTS DEFINITION

- **Prototype Evaluations of Key Components**
- **Refine Architecture**
- **Develop Preliminary System Design**

7/01/ Begin Phase III: DETAILED SYSTEM DESIGN

-
-
- **Prototype Implementation**
-
-

1/01/95 Begin Phase IV: EXPERIMENTAL TESTING

-
-
- **Verification with Scientific Experiments**
-
-

5/15/95: SCHEDULED END OF 3-YEAR FUNDING

Figure 4.1: SCHEDULE of REINAS PROJECT

5. REINAS Issues and Action Items

5.0.1 Purchase of Computing/Network Hardware and Software

During the early phases of the project the equipment expenditures are quite modest. Part of the study will be to examine new software and hardware as it becomes available on the market. It is too early to make specific hardware or software selections at this time. When a detailed model of the system is developed, which includes the details of required control, timing, and data flow, then it will be used to evaluate and select between different hardware and software alternatives. A key design consideration is to build a system which can be expanded to include new sensors or new computing equipment as it becomes available.

5.0.2 Prototype Testing

Testing of the visualization components will be an ongoing process during the implementation stage. When the rest of the system components (instrumentation, networking and database management) are at a prototype stage, the visualization component will be integrated and tested. Prior to that, it can be evaluated separately to a certain extent. For example, the ability to handle real-time data can be simulated by feeding the visualization system with historical data at rates corresponding to real time data sensor rates. Historical data and data analyses visualization can certainly be tested independently.

5.0.3 Integration

After evaluating the prototype, any changes and further improvements can be incorporated during the integration stage. This stage will encompass porting the visualization environment to the prototype system, and continued monitoring and refinement of the visualization environment, expected to last through the completion of the project. The critical evaluation and testing during this stage is to let scientists run the system with real data and problems, get feedback from actual usage, and to guarantee a smooth transition to a system for production use.

5.0.4 Anticipated Critical Issues

- Issues caused by Data sets being from multiple sensors and numerical models: Diverse data sets are to be mapped to visual parameters, perhaps in sound too, to convey the information content to the scientists. These will require effective use of visualization/sonification techniques.
- Issues regarding special handling of data sets: Dealing with spatially sparse data sets. Representing data quality. High dimensionality of data sets. Real-time considerations and tradeoffs.
- Issues concerning initial project staffing (time-critical positions): Need people who are versed in visualization tools/techniques to do initial prototype. Also need people strong in mathematics to deal with interpolations and transforms. Need good graduate students - a challenge since “systems-sense” is rare.

- Issues concerning management: Need a "Chief Programmer" with management skills and systems-sense. He/she needs to understand software design and have a feel for performance issues.
- Issues raised by real-time requirements: Development of task scheduling scenarios. Is there a need to meet guaranteed timing constraints? How important are "correct" results in the real-time context? What about expected future real-time uses of REINAS other than seabreeze? Other scientific contexts may have different time/space scale phenomena. Will the needs be known at design time?
- Issues raised by portability: Does the portability requirement imply that the resource availability to meet real-time constraints will vary? Must REINAS account for dynamic resource profiles?
- Issue of Bimodal Design: Do we design REINAS to operate in a real-time and in retrospective mode, or is REINAS as only a real-time component that passes data off to a different retrospective system? (Note: Tools will work in either mode. They don't know or care whether the data is fresh or old.)
- It is not clear at this point whether the visualization component should be directly fed with data from sensors or go through the database first. This issue will be one of the questions we hope to address in our methodology.
- Are there are different interpretations of "operational" for the Navy, for science, and for both? What trade-offs will be necessary in REINAS?

References

[AHS90] D. P. Anderson, R. G. Herrtwich, and C. Schaefer. Srp: A resource reservation protocol for guaranteed-performance communication in the internet. Technical Report UCB/CSD 90/596, University of California, Berkeley, Sep 1990.

[Atk81] B. W. Atkinson. *Meso-scale Atmospheric Circulations.* Academic Press, 1981.

[BC92] N.J. Belkin and W. Bruce Croft. Information filtering and information retrieval. *Communications of the ACM*, pages 29–38, December 1992.

[Che88] D. R. Cheriton. The v distributed system. *Communications of the ACM*, 31(3):314–333, March 1988.

[COA92] NOAA COAP. Mader. In *NEONS Users Conference, Monterey California.* Naval Research Laboratory, Fleet Numerical Oceanography Center, 1992.

[Com91] Douglas E. Comer. *Internetworking with TCP/IP (2nd ed.).* Prentice Hall - Englewood Cliffs N. J., 1991.

[Cox88] D. Cox. Using the supercomputer to visualize higher dimensions: An artist's contribution to scientific visualization. *Leonardo*, 22:233–242, 1988.

[DCH88] Robert A. Drebin, Loren Carpenter, and Pat Hanrahan. Volume rendering. *Computer Graphics*, 22(4):65–74, July 1988.

[Den89] Denning. The science of computing. *American Scientist*, 77:216–217, May-June 1989.

[DGMS85] Susan B. Davidson, Hector Garcia-Molina, and Dale Skeen. Consistency in partitioned networks. *Computing Surveys*, 17(3):341–370, 1985.

[Dye90] D. S. Dyer. A data flow tool kit for visualization. *IEEE Computer Graphics and Applications*, 10(4):60–69, July 1990.

[FKU77] H. Fuchs, Z. M. Kedem, and S. P. Uselton. Optimal surface reconstruction from planar contours. *Communications of the ACM*, 20(10):693–702, October 1977.

[FLP89] H. Fuchs, M. Levoy, and S. M. Pizer. Interactive visualization of 3d medical data. *IEEE Computer*, pages 46–51, August 1989.

[GBD$^+$89] B. Gritton, D. Badal, D. Davis, K. Lashkari, G. Morris, A. Pearce, and H. Wright. Data management at mbari. *Oceans 89 - The Global Ocean*, pages 1681–1685, 1989.

[GM90] Henry M. Gladney and Patrick E. Mantey. Essential issues in the design of shared document/image libraries. *Proceedings Image Communications and Workstations*, d1258:54–65, February 1990.

[Gos90] M. E. Goss. A real time particle system for display of ship wakes. *IEEE Computer Graphics and Applications*, 10(3):30–35, May 1990.

[HH91] J. L. Helman and L. Hesselink. Visualizing vector field topology in fluid flows. *IEEE Computer Graphics and Applications*, 11(3):36–46, May 1991.

[Hib89] W. Hibbard. Meteorology applications. *ACM Siggraph Video Review, Issue 44*, 1989.

[HKM$^+$88] John H. Howard, Michael L. Kazar, Sherri G. Menees, David A. Nichols, M. Satyanarayanan, Robert N. Sidebotham, and Micheal J. West. Scale and performance of a distributed file system. *ACM Transactions on Computer Systems*, 6(1):51–81, 1988.

[HL79] G. T. Herman and H. K. Liu. Three-dimensional display of human organs from computer tomograms. *Computer Graphics and Image Processing*, 9(1):1–21, January 1979.

[HM90] R. B. Haber and David A. McNabb. Visualization idioms: A conceptual model for scientific visualization systems. In *Visualization in Scientific Computing*, pages 74–93. IEEE Computer Society Press, 1990.

[HS89] W. Hibbard and D. Santek. Visualizing large data sets in the earth sciences. *IEEE Computer*, pages 53–57, August 1989.

[Hsu88] S. A. Hsu. *Coastal Meteorology.* Academic Press, 1988.

[Hul90] J. Hultquist. Interactive numerical flow visualization using stream surfaces. *NASA Ames Research Center - RNR-90-009*, 1990.

[JO73] A. Jr. Johnson and J. J. O'Brien. A study of an oregon sea breeze event. *J. Appl. Meteorology*, 12:1267–1283, 1973.

[Jur91] A. Jurkevics. *Database Design Document for the Naval Environmental Operational Nowcasting System, Version 3.4.* US Navy, 1991.

[LC87] W. E. Lorensen and H. E. Cline. Marching cubes: A high-resolution 3d surface construction algorithm. *Computer Graphics*, 21(4):163–169, July 1987.

[Lea87] T. Lasinski and et al. Flow visualization of cfd using graphics workstations. *AIAA*, pages 87–1180, 1987.

[Lid89] Lidar. (to find). *(to find)*, 1(1):1–3, 1989.

[LPS81] B. W. Lampson, M. Paul, and H. J. Siegert. *Distributed Systems - Architecture and Implementation: An Advanced Course.* Springer-Verlag - Berlin, 1981.

[Mea87] Sape J. Mullender and et al. *The Amoeba Distributed Operating System: Selected Papers 1984-1987.* Centrum voor Wiskunde en Informatica - Amsterdam, 1987.

[MG92] P. Mantey and B. Gritton. *Site visit notes.* private communication, 1992.

[Mul89] Sape Mullender. *Distributed systems.* Addison-Wesley - New York N. Y., 1989.

[Mus89] F. K. Musgrave. Prisms and rainbows: A dispersion model for computer graphics. *Proceedings of the Graphics Interface '89 - Vision Interface '89*, June 1989.

[NAS92] NASA. *NASA, EOS Data and Information System (EOSDIS).* ESSO Document Resource Facility via NASA Headquarters, Earth Science and Applications Division(code SE), Wash, D.C. 20546, May 1992.

[NCA92] NCAR. *STORM Fronts Experiment Systems Test: STORM-FEST Operations Plan.* STORM Project Office, National Center for Atmospheric Research, 1992.

[NWO88] M. Nelson, B. Welch, and J. Ousterhout. Caching in the sprite network file system. *ACM Transactions on Computer Systems*, 6(1):134–154, 1988.

[OCD$^+$88] J. K. Ousterhout, A. R. Cherenson, F. Douglas, M. N. Nelson, and B. B. Welch. The sprite network operating system. *IEEE Computer*, pages 23–36, February 1988.

[PCH$^+$90] R. L. Phillips, B. Cabral, C. L. Hunter, R. B. Haber, G. V. Bancroft, T. Plessel, F. Merritt, P. P. Walatka, and L. J. Rosenblum. Scientific visualization at research laboratories. In B. Shriver G. M. Nielson and L. J. Rosenblum, editors, *Visualization in Scientific Computing*, pages 209–253. IEEE Computer Society Press, 1990.

[PGK88] D. Patterson, G. Gibson, and R. Katz. A case for redundant arrays of inexpensive disks (raid). *Proceedings of the ACM SIGMOD Conference*, pages 109–116, 1988.

[Pie84] R. A. Pielke. *Mesoscale Numerical Modeling.* Academic Press - New York, 1984.

[RAT$^+$85] M. L. Rhodes, Yu-Ming Azzawi, E. Tivattanasuk, A. Pang, K. Ly, H. Panicker, and R. Amador. Curved-surface digital image reformations in computed tomography. *Proceedings of SPIE, Medical Image Processing*, 593:89–95, December 1985.

[SFCO67] M. J. Schroeder, M. Fosberg, O. P. Cramer, and C. O'Dell. Marine air invasion of the pacific coast: A problem analysis. *Bull. Amer. Met. Soc.*, 48:802–808, 1967.

[SG89] A. Stettner and D. P. Greenberg. Computer graphics visualization for acoustic simulation. *Computer Graphics*, 23(3):195–206, July 1989.

[Smi85] Alan Jay Smith. Miss ratio analysis and design considerations. *ACM Transactions on Computer Systems*, 3(3):161–203, 1985.

[Sto91] M. Stonebraker. *An Overview of the Sequoia 2000 Project*, volume Report No. 91/5. 557 Evans Hall, UC Berkeley, 1991.

[Tan88] Andrew S. Tanenbaum. *Computer Networks (2nd ed.).* Prentice-Hall - Englewood Cliffs N. J., 1988.

[TJ92] T. Tsui and A. Jurkevics. A database management system design for meteorological and oceanographic applications. *MTS Journal*, 26(2):88–97, 1992.

[Uea89] C. Upson and et al. The application visualization system - a computational environment for scientific visualization. *IEEE Computer Graphics and Applications*, 9(4):30–42, July 1989.

[USD91] USDC. Awips/dare/profs. *Forecast Systems Laboratory 1990 Annual Report*, July 1991.

[vT91] A. van Tilborg. *Foundations of Real-Time Computing.* Kluwer Academic Publ., 1991.

[Wat90] M. Watt. Light-water interaction using backward beam tracing. *Computer Graphics*, 24(4):377–385, August 1990.

[WCA$^+$90] J. Wilhelms, J. Challinger, N. Alper, S. Ramamoorthy, and A. Vaziri. Direct volume rendering of curvilinear volumes. *Computer Graphics*, 24(5):41–47, November 1990.

Appendix A. CODAR

From: Madhukar N. Thakur [thakur@cse.ucsc.edu]

A.1 Introduction

A large number of man's ocean related activities take place near the ocean surface and close to the coast. An accurate snapshot of the ocean surface currents and a dynamic observation of these currents as they form and propagate is of immense help both to oceanographers studying such phenomena and for agencies coordinating other human activities close to the shore and surface. This provided the incentive for the development of the Coastal Ocean Dynamics Applications Radar (CODAR) system. This is a remote sensing system developed by the Codar Ocean Sensors, Ltd. that measures the ocean currents at the surface, within the half meter of the ocean.

Surface maps from CODAR derived velocities can be used to track and predict the movement of floating objects and suspended materials. Such monitoring activity has potential applications in oil slick tracking and containment, surface drifter studies and modeling biological material transportation, among other things.

A.2 Operation

The basic mechanism underlying the mechanism of the CODAR is that high frequency (HF) radar echoes scattered from the sea possess a unique "signature" due to motions of waves. These echoes can be related directly to certain ocean properties. At HF the radar wave selectively backscatters from two wavetrains moving directly toward and away from the radar. If surface currents are transporting these waves, the Doppler shift, and subsequent Fourier transformations, indicate the current velocity component in the direction of the radar; the processing software then determines the range and bearing, and hence forms a current map. Range from the observing station is determined by measuring the time delay of the echo from the transmission.

In the current implementations, the resolution is 2km. The bearing of the signal is derived from the signal intensities received by different elements of the systems antenna. It transmits only 200 watts of power and the total power consumption is 1kw and provides current maps up to 22km offshore in experiments at Monterey Bay.

A.3 Type of Data

National Oceanographic and Atmospheric Administration (NOAA) currently operates two CODAR sites alongside the Monterey Bay, one at Moss Landing and the other at Pacific Grove. A third site in Santa Cruz will be operational at the end of November 1992. The processed CODAR data consists of vector values, representing the velocities along two orthogonal directions, at points on a grid. Each CODAR point grid represents the center of a 2km × 2km box. The system collects data from an area of ocean over 550 square kms. The system operates by transmitting HF pulses for 26 minutes and surface velocities information is extracted from the raw data after post processing. The central processing is done at the company's office in Fremont. Before processing, data is obtained from the two sites using a slow (1200 baud) modem. (CODAR expects to start using 9600 baud modems by the end of November!) We were unable to extract the actual size of the raw date from the company's employees, though we learned that it takes about 1 hour to collect data from the two sites and to process it into useful data at the central processing site. Currently, users at the Naval Postgraduate School obtain data from the company's central processing site by a dial up modem every two hours.

The size of a file containing the final vectors (obtained every two hours) varies from 3K bytes to 8k bytes.

A.4 Scientific value of CODAR data

The CODAR data provides larger area of coverage than impossible with in situ measurement devices. The current values at some points in the grid are erratic, missing or improbable. This is due to atmospheric conditions, radio interference from passing ships or other natural sources. The effects of these are minimized using averaging techniques. The mean currents and low-pass filtered time series provide insights into the low frequency events (with periods greater than one day). For the high frequency events CODAR does not exactly compare with the data from the ADCP system operated by MBARI.

Real-time use of CODAR, as a stand alone system, is hampered by the presence of missing, erratic or improbable vectors. However other components of REINAS could provide alternate types of data that could be used to corroborate or reject some of CODAR's data points and make the system conducive to real-time use.

Appendix B. SEA PROFILE DATA

From: Eric Rosen [eric@cse.ucsc.edu]

B.1 Introduction

Sea profile data consists of a stream of samples taken approximately vertically through a column of water. Each stream consists of two major parts: a downcast and an upcast. During the downcast, instruments are lowered to the sea floor while attached to a long steel cable. A typical drop velocity is 60 meters per minute. Samples are taken continuously during the drop; the nominal sampling rate is 24 samples per second. This provides a typical raw resolution of 24 samples per meter of depth (or roughly one sample every 4.1 cm). During the upcast, the instruments are raised to the surface and used to trip water collection bottles at prescribed depths. Any drift in horizontal position is usually ignored. During adverse weather conditions, a lower drop velocity is usually used.

B.2 Data Samples

The downcast data is normally processed through bin-averaging to produce a resolution of one sample per decibar increase in pressure. A one decibar increase in pressure roughly corresponds to a one meter increase in depth. Each sample can contain salinity, temperature, oxygen content, and light transmissivity information at a specific depth; other parameters are possible. Each sample (or record) is typically no larger than 30 bytes.

The upcast data can be used to calibrate the downcast data by providing actual water samples at prescribed depths. Other biological data is also obtained from the upcast samples.

In addition to the upcast and downcast, each dataset will also include a small amount of header information indicating the latitude, longitude, time-stamp, and other identification information.

B.3 Datasets

Each sea profile dataset corresponds to a specific physical site (or station). Normally, a marine research vessel visits several sites in series, collecting sea profile data at each. The sites are *not* usually arranged in a regular pattern but rather depend on local terrain, weather and other arbitrary factors. There is no standard surface interval between profiles.

Assuming a 24 Hz sampling rate with a vertical drop velocity of 60 m/min, and a nominal depth of 500 m, the typical downcast will consist of 12000 samples. At 30 bytes/sample, a raw downcast dataset will have a typical size of 350 KB. In shallow areas, 100 m depths are common, and raw datasets may be as small as 70 KB. At canyon depths of 1000 m, 700 KB raw datasets are possible.

After processing, however, a smaller and less noisy dataset is produced. Under the previous assumption, a 350 KB dataset for a 500 m depth would yield a 15 KB dataset containing roughly 500 individual samples.

Sea profile datasets can be visualized individually as two-dimensional plots (eg: salinity versus depth) or in groups (eg: an iso-contour plot of temperature at 50 decibars over some spatial region of interest). Generating iso-contour plots over the scattered site locations can be challenging.

Qualitatively, the data is fairly continuous, both vertically (down a cast) and horizontally (in space). Changes are usually gradual. Early experiments on Antarctic profile data using primitive compression models suggest the data can be compressed to less than one-fifth of its original size. Better compression performance is expected in the future.

Appendix C. THE WIND PROFILER at FORT ORD

From: Chane L. Fullmer [chane@cse.ucsc.edu]

C.1 Radar Sampling

The Wind Profiler uses the Doppler shift of a 400MHz Radar to measure the velocities of wind movements. The system broadcasts three beams from which a vector is calculated to determine the actual velocity and direction of the wind. By varying the power level the system can measure winds at variable heights, generally from 550m to a maximum of 10km, with a resolution of 250m.

The system must be shut off for periods of 12 minutes when the polar orbiting satellites pass overhead because the operating frequency of the profiler is close to the rescue signal of downed air and sea craft. Sampling is done at 6 minute intervals - one minute of sampling in each beam direction. The data for each sample by itself is considered unreliable unless a consensus covering several samples, currently over one hour. The data file contains a header giving site location, time period for the data, and exact longitude and latitude coordinates. The data itself is given in rows containing the following:

C.2 Data Samples

`HGHT`	`The height of the sample in meters`
`DIR`	`The direction of the wind at this height in compass direction`
`SPD`	`The speed in meters/second`
`U,V,W`	`The radial velocity in m/s for each beam at this point`

`The following is saved with the data for verification purposes:`

`XPWR, YPWR, ZPWR`	`The signal power levels of the beams in db`
`XWIDTH,YWIDTH,ZWIDTH`	`The spectral width`
`XNL,YNL,ZNL`	`The noise levels`
`RES`	`The resolution beam (250m or 1km)`

There are about 30 Wind Profilers in the US, including 2 close by – at Fort Ord, and at Vandenburg AF base. The rest are scattered throughout the midwest and the east coast.

C.3 Additional Information on the Wind Profiler

From: Dick Lind [5064P@NAVPGS.BITNET@CORNELLC.cit.cornell.edu]

Somewhere in the little Blue book, is described the signal processing that is used to come up with a wind. Spectra of the In-phase and Quadrature channels are generated. Rather than send the 36 spectra across the phone line, local processing is done to glean the important data from each spectrum. There are four statistical properties that are kept: (1) Radial velocity, (2) Signal Power, (3) Spectral width and (4) Noise level.

Listed in the files I gave you are the last three items. PWR, Width and NL. They need to be saved because some quality control procedures can use them to objectively determine if the resultant wind can be believed. For example, if the wind is clearly erroneous, you are more than likely to find that one or more of these three values is vastly different that a neighboring wind that looks believable. Signal-to-noise ratio is commonly employed to do this. Since the PWR is the Signal power and NL is Noise power (both in dB), you can subtract the NL from PWR and add -10*alog10(256) to get the SNR. The 10log256 is due to the sampling–you can extract a weaker signal if you have more samples to work with.

Variable RES is the resolution of the beam. Since 1km and 250m pulses are used, some way of differentiating between them is required. The vendors use 999 instead of 1000 for the 1km pulse.

Regarding the surface data: at this time, it is not networked into our system, although I plan to integrate these data soon. At this time there are several formats, so I'll hold off on sending you a sample until I decide which is most appropriate and efficient. I also have my own consensus algorithm that incorporates some logic, rather than the dry statistical algorithm that is currently being used.

Appendix D. SEQUOIA

D.1 Introduction

From: Michael Stonebraker, UCB

The Sequoia 2000 project is a coordinated attack on the issues of Global Change on the planet Earth. It is supported by Digital Equipment Corp at UC Berkeley with participants from the Computer Sciences Division and the School of Library and Information Studies at Berkeley, the Computer Sciences Department at San Diego and the San Diego Supercomputer Center. Their mission is to carry out a collection of research studies and build prototypes.

The Global Change research groups wish to store approximately 100 Tbytes of data at reasonable cost and in a reasonable footprint. The goal is to build such a storage system, denoted BIGFOOT, capable of this level of capacity and with sufficient speed to support the traffic which will be directed to it.

The second system, denoted Jaquith, is a migrating enhancement which can be added to any file system. It assumes the existence of a disk-based file system for the machine being used. To this it adds a tertiary memory file system, oriented toward tape jukeboxes. The third file system is an innovative organization of database management system which discards unneeded pieces of the file system.

D.2 Postgres Database Management System

Postgres is a database research project under Prof. Michael Stonebraker at UCB. To facilitate research efforts, a software test-bed was created; this is the "Postgres" DBMS software. The Postgres DBMS is extended relational or object oriented, depending on the buzzword du jour.

The major purpose of this software is to provide a platform and a basis for the testing of implementations of new ideas in database research. Several graduate students, staff members, as well as undergraduate programmers have been working on the implementation of the Postgres software. Postgres is:

- Relational. One of the major goals of Postgres is to show that an essentially relational DBMS can be extended to handle complex objects, rules, and be highly extensible.
- Highly extensible. Postgres allows user-defined operators, user-defined objects, and user-defined functions.
- While Postgres is relational, object oriented ideas have been implemented in Postgres (inheritance, etc.).
- Numerous other features, such as query language procedures, rules, etc. which are beyond the scope of this discussion.

For more info on the Postgres research itself, you can get the Postgres technical documentation on postgres.berkeley.edu. A file called [pub/postgres-papers.tar.Z] contains troff and Postscript versions of several of the major Postgres technical reports. You can order hard-copies of technical reports individually; for more details, send e-mail or call Sarah Burke at (510) 642-3417. Her e-mail address is: [sarahb@postgres.berkeley.edu].].

D.3 Bigfoot Dataset Inventory

From: frew@postgres.Berkeley.EDU

Herewith a summary of our current data holdings at UCB. The purpose of this mailing is to let you-all know who's contributing what KINDS of data, so you can help us determine which datasets are of the greatest interest to the project as whole.

The descriptions are subdivided first by medium (i.e. how we received the data). The name following each dataset is that of the person who actually delivered the dataset to us; not necessarily the person responsible for generating/obtaining/maintaining the data.

I solicit corrections, additions, or suggestions regarding the following information, especially which datasets are YOU specifically interested in and what additional information do you need about particular datasets listed below.

```
--------------------------------------------------------------------------
magnetic disk (includes "Ninja" magneto-optical jukebox):

 - UCLA Coupled GCM Ocean model output                             Weibel
 - CIA World Database 2                                           Woodruff
 - DWR Intergraph "DGN" files for "The Delta Atlas"       Woodruff
 - DWR Delta hydrology model output                       Woodruff
 - DWR Delta hydromet telemetry, "SHEF" format                    Woodruff
 - DWR seismic waveforms (single event @ multi stn)       Woodruff
 - DWR tech reports, EIRs, etc.                                   Woodruff
 - UCB Computer Science technical reports                 Mosher
 - ISCCP B3 global GOES data                              Frew
 - TIGER/Line Census tract outlines                       Frew
 - GRASS Sequoia demo: SF Bay                             Gardels
 - data files for S2K benchmark                                   Frew
 - UCSB 64**3 fluid model time series                     Miley
 - NASA TM scene 042-034 for 18 Jan 1986, 05 May 1986     Rosenthal
 - NASA CZCS monthly composites, 1978..1986               Davidson
 - NOAA AVHRR biweekly composites 1990                            Frew
 - USGS hydrography DLGs (CA)                             Frew
 - USGS GIRAS landuse/landcover (CA)                      Frew
 - USGS GNIS names (CA)                                           Frew
 - NASA TOMS daily global ozone 1978..1991                Frew
 - NCDC CA/NV daily climatology 1877-1989                 Bainto
 - NOAA N hemisphere 700 mb height monthly 1946-1992      Bainto
 - NODC temperature/salinity profiles                     Hall
 - NASA TM scene 041-035 for 24 Jan 1985                  Davidson
 - COADS-derived global ocean atlas                       Waliser
 - COADS-derived Tropical Pacific climatol. 1957-1976     Waliser
 - UCLA atmospheric GCM 59 month simulation               Waliser
 - (Gautier) Global satellite-derived ocean products      Waliser
--------------------------------------------------------------------------
- 8mm tape:
```

```
 - CA TM resampled to 25m UTM (6 scenes)                  Bueno
 - ISCCP B3 1983-1986                                     Figel
 - NCDC HCN daily precip and minmax T 1950-1990                   Roads
 - NCAR global NMC 2.5 deg analyses 2x daily 1984-1991            Roads
 - ‘‘Willmott’’ 0.5 deg global precip and T monthly/annua   Roads
 - USAF 2.5 deg global snow depth                         Roads
 - NCDC US rawinsonde 1984-1990                                   Roads
 - NCDC first order weather daily summaries 1984-1990     Roads
 - NMC global medium range forecasts 1990-1991                    Roads
 - NOAA N Pacific 5 deg SSTs monthly 1947-1992                    Cayan
 - NOAA N hemisphere 700 mb height 2x daily 1947-1988     Cayan
 - NOAA N hemisphere sea-level pressure 2x daily 1947-1   Cayan
--------------------------------------------------------------------------
CD-ROM:

 - NASA CZCS West Coast Time Series vol 1 1979-1981       Smith
 - NASA FIFE satellite/aircraft/ground data May-Dec 1987  Frew
 - EPA global ecosystem database version 0.1 beta         Davis
 - USDA SCS / USGS NMD digital orthophoto demo (Dane Co. WI) Gardels
 - NASA SMMR N Polar radiances 25 Oct 1978 - 31 Dec 1982  Smith
 - NASA SMMR N hemisphere T grids 29 Oct 1978 - 31 Jan 1980 Smith
 - NASA SMMR N Polar radiances/sea ice 1 Jan 1980 - 30 Aug 1980   Smith
```

Appendix E. Great Lakes Forecasting System

From: Naim Alper [alper@cse.ucsc.edu]

I saw a reference that might be of interest to us. The relevant paragraph and the given reference appears below. I will go ahead and put in a request for an interlibrary loan and place the article in the file cabinet.

“The Lake Erie simulation is part of the Lake Erie Forecasting System (LEIFS), which is a prototype for a Great Lakes Forecasting System (reference below), using current satellite data and ground observations to make short term forecasts and provide images to interested groups associated with the Great Lakes. In the case of the commercial fishing industry , the images can be delivered by fax or over modems directly to ships and then displayed by on-board computers.”

```
Ref: Bedford, K.J., Yen, C.C., Kempf, J., Scwab, D., Marshall, R.
     and Kuan, C., "A 3D-Stereo Graphics Interface for Real-Time
     Great Lakes Forecasts," Proc. Estuarine and Coastal Models,
     ed. Lacolm Spaulding, American Soc. Civil Engineers, June
     1990.
```

Appendix F. Satellite Image Data

From: Eric Rosen [eric@cse.ucsc.edu]

Since 1978, NOAA has maintained a collection of Earth observing satellites in polar orbit, generally referred to as the NOAA or TIROS family of satellites. The primary instrument of interest on these satellites is the Advanced Very High Resolution Radiometer (AVHRR), a cross-track scanning system providing images in four or five spectral bands ranging from visible light to infrared.

The AVHRR images are broadcast continually by the satellites and can be received by ground stations that are within line-of-sight. The maximum amount of data is obtained when the satellite passes directly over the ground station antenna, and is then visible for between 15.5 to 16.0 minutes (under ideal circumstances). The ground track width of an AVHRR image received as an orbiter passes directly over Monterey Bay, California is about 1330 km. This translates to a resolution of about 1.5 km/pixel. From such a pass, roughly 975 pixels can be expected to cover the 2500 square kilometer area containing Monterey Bay. About four satellite passes per day will contain data pertinent to Monterey Bay, equivalent to about 15 MB.[1]

Each scanline consists of 2048 pixels, each with 16 bit resolution. The individual scanlines are time-stamped and transmitted without any position information. Knowledge of the satellites' ephemeris is required to determine exactly what geographical location has been imaged by the scanline. Orbital elements can be obtained via electronic means.

In addition to the obvious value of the image of the planet from low Earth orbit, various valuable parameters can be derived from the images. Of these, surface temperature, and surface and cloud albedo are among the most important.

Applying image-processing to the problem of automatically detecting environmental features from these satellite images is a potential area for research and would combine well with the focus on developing a system which can detect interesting environmental phenomena. Such a capability would also be valuable in providing a means of searching a historical database of satellite images for specific features.

In addition to the NOAA polar orbiters, a GEOS satellite in geostationary orbit continuously transmits images containing Monterey Bay to Earth. However, the resolution of these images is significantly less than that provided by the polar orbiters.

```
Ref:  "NOAA Polar Orbiter Data User's Guide"
       Compiled and Edited by Katherine B. Kidwell
       NOAA Satellite Data Services Division, July 1991.
```

From: Chuck Wash [WASH@LADY.MET.NPS.NAVY.MIL]

You are welcome to any of the data we have here or we can collect from other sources. The confusion about the AVHRR satellite data concerns its applicability to the seabreeze problem and the processing required to convert the raw data into geophysical units.

---

[1] These parameters are derived from information given in *NOAA Polar Orbiter Data User's Guide*.

The AVHRR data covers the Monterey Bay region once every 6 hours. Its use in describing the seabreeze, particularly over 1/2 hour geostationary satellite data, is not clear at this point. The data is visible and IR radiances and they need to be earth located, calibrated and converted to surface or cloud temperatures and reflectances. This is about an hour of processing on a SUN SPARC 2.

Since Kurt has 200 passes of data from the Azores he is working on, he correctly wonders who will process the Monterey Bay AVHRR data and even if it the highest priority data to be working on. The answer to the first question is the environmental data post-doc and engineer as well as the visualization engineer are the folks who would learn how to process these data. We all agree we need these folks as soon as possible.

I repeat my offer to Alex that we need to have some discussion sessions on what data is the most valuable and prepare a plan of data processing to support this effort. All of us need to form a collective vision of REINAS data processing and display in 6 months, 1 year, 18 months, etc.

Appendix G. Korea-China Yellow Sea Experiment

From: Bruce R. Montague [brucem@cse.ucsc.edu]

I have come across a very interesting paper by Pat Wilde, of ONR Asian Office (ex Berkeley Ocean Engineering Chairman, head of Lawrence Berkeley Marine Sciences Group).

ONR maintains an Asian office in Tokyo that monitors Japanese and other Asian research. They put out a quarterly collection of reports called "Office of Naval Research Asian Office Scientific Information Bulletins," dedicated to "providing items of interest well in advance of the usual scientific publications" (they go to a lot of meetings and tell what's going on). These are not "fluff" articles, they are written by scientists for scientists.

The current issue, TR ONRASIA Vol 17, No 3, pp. 231-236, contains an article "First Workshop on the Yellow Sea Experiment (YESEX-1)", by ONRASIA staff member Pat Wilde. Dr. Wilde joined ONRASI in 1991 as ocean science liaison. He has been chairman of Ocean Engineering at UC Berkeley (1968-1975) and head of the Marine Sciences Group at Lawrence Berkeley Labs (1977-1982). In this article he introduces the YESEX-1 project and reports on a conference he attended, including summaries of many of the talks presented. The conference was conducted at Seoul National University in April of 1992.

The Korea Ocean Research and Development Institute (KORDI) is proposing expanding a current project to an international joint effort with the Chinese and US. My reading of this project description is that it has many REINAS similarities. I will put a copy of this paper in the top draw of the filing cabinet in room 350 (under W, for "Wilde"). Some excerpts:

G.1 Introduction

"KORDI" has proposed the Yellow Sea become an international full-size test laboratory as an outgrowth of the Korean program of real-time coastal monitoring and prediction initiated in 1991... development and testing of both equipment and various numerical models."

G.2 Keynote Session (Dr.Dong-Young Lee, KORDI)

"Importance of regional studies of this nature to both the global and local environmental scenes.... major United Nations effort." Outline of "development of the Integrated Coastal Monitoring Network and the Yellow Sea Experiment."

"combination of... in-place and at-a-station instrumentation ... also wide area real-time measurements such as wave properties from radar... 175 fixed stations already monitored by the Korean Fisheries Agency to obtain some historical record... 350 light buoys... ships of opportunity..." Dr.P.K.Park (US NOAA)... international disaster, ocean survey, satellite databases... parameters on the 'synoptic scale'.

G.3 Marine Environment of the Yellow Sea (Prof. Young-Ho Seung)

"...lack of systematic and long time series records... Since 1986 formal cooperation among China, South Korea, US.. marine pollution.. (At this point chemical parameters of marine pollution, heavy metal concentrations, etc.), sedimentation..."

G.4 Research for Coastal Development

Ocean engineering... Russian activities in wave monitoring (A.Rabinovich from the Russian Academy of Science active in tsunami prediction at Kamchatka presented here). Tidal land reclamation project... field information... sea dike.. 2-D numerical models... port and harbor engineering... shore protection structures...

G.5 Coastal Prediction Models

Tidal computation... 2-D model... 100,000 grid points in future models... actual field measurements are scarce...need for integrated tidal station network... have a 2-km gridded model of the area near Pusan... Development of a Mixed Spectral-Finite Difference Model for Computation of Wind Induced Currents in the West Coast of Korea (Dr. Jae Kwi So, KORDI)... model...3-D Davies type...

G.6 System Synopsis

Kordi "described plans for the development of a real-time coastal prediction system. The goal is a real-time user input system, probably PC computer based, where a series of oceanographic/ocean engineering questions can be answered. Ideally the user gives the time/s of observation, location, and parameters/s requested and receives the answer through a computer search of the appropriate model or tables stored in the database... maritime accidents... search and rescue... drift predictions..."

G.7 Coastal Environmental Studies

Remote sensing... atmospheric interference... 'in-water algorithm'... dust... NASA standard clear water radiance method... Emphasizes the importance of site specific algorithm evaluation and the danger of reliance on some 'universal' formula for satellite interpretation... marine oil spill.

G.8 Technologies for Coastal Monitoring System

Real-time wave monitoring system... 8 basic stations.. via radio... field study of wave spectra... wave gauges... Maximum Entropy Method... satellite data from NOAA... radar remote sensing... new technologies of ocean data measurement using optical fibers... chemical concentration... might be monitored in real-time using new optical fiber techniques.

G.9 Summary (Dr. Wilde's)

"Showcase for Korean scientists..." all the nations bordering... formative stages... proper evolutionary phase...

The politically astute, intimate involvement of the government, academic, and private sector groups in all facets of the planning leads both to concrete results and general long-term programmatic stability, which are critical to the success of a project as vast and as complex as the Yellow Sea Experiment... a full-scale test facility to examine various oceanographic problems and models should attract workers from the international community not specifically interested... in the Yellow Sea.

References:

Dr. Dong-Young Lee of KORDI (principal workshop organizer) was apparently one time of U of Florida at Gainesville:

```
1. Lee, Dong-Young, 1949-
     Wave spectral transformation in shallow water / by Dong-Young Lee and
     Hsiang Wang.  Gainesville, Fla.  :  Coastal and Oceanographic
     Engineering Dept., University of Florida, 1984.
     Series title:  Technical report (University of Florida. Coastal and
     Oceanographic Engineering Dept.) ; 53.
       UCB   WRCA      G229 XU2-2 no.53
```

Kordi has apparently been in business since at least 1978.

```
1. Han'guk Haeyang Kaebal Yon'guso.
     Oceanographic atlas of Korean waters.  [Seoul] :  Korea Ocean
     Research and Development Institute, 1987- Scales differ.
       UCSD  Scripps   GC791 .O5 1987 Atlas Shelves

3. Oceanographic atlas of Korean waters / edited by Sangbok D. Hahn.
     Seoul:  Korea Ocean Research and Development Institute, KIST, 1978.
     UCSD Scripps C2331.C7 O34 Atlas Shelves
```

Melvyl has information on at least 22 papers by Dr.Wilde. Some that seemed of interest:

```
20. Wilde, P. (Pat) 1935-
    Physical oceanology / Pat Wilde (Instructor), University of
    California, Berkeley.  Berkeley, CA:  Committee on Ocean Engineering,
    College of Engineering, University of California, [1972?].
        UCB   WRCA      96 K2-1 v.1

11. Wilde, P. (Pat) 1935-
    Geological oceanology / Pat Wilde (Instructor), University of
    California, Berkeley.  Berkeley, CA:  Committee on Ocean Engineering,
    College of Engineering, University of California, [1973].
        UCB   WRCA      96 K2-1 v.3

4. MAP
   Chase, Thomas E.
     Oceanographic data of the Monterey Deep Sea Fan / by T.E. Chase,
    W.R. Normark and P. Wilde ; assisted by L. Hydock [et al].  1st ed.
    June 1975.  [La Jolla, Calif.] :  Geologic Data Center, Scripps
    Institution of Oceanography, 1975.
     Scale 1:898,524 at 350 Latitude ;.
```

Series title: IMR technical report series ; TR-58.
UCB WRCA 96.1 XU2-4 no.58

5. MAP
Geographic, bathymetric, geologic, and physical oceanographic data of
potential OTEC sites / Prepared by T.E. Chase, P. Wilde, [et al.];
prepared in cooperation with the Marine Science Group, Department...
[Reston, Va.] : U.S. Geological Survey, 1986.
Scales differ ;.
Series title: Open-file report (Geological Survey (U.S.)); 86-333-B.
UCB WRCA m27.9 M6

8. Lee, J.
Recent sediments of the central California continental shelf, Pigeon
Point to Sand Hills Bluff / by J. Lee, T. Yancey and P. Wilde.
Berkeley, Calif.: University of California, Hydraulic Engineering
Laboratory, 1970-. Series title: Technical report (University of
California, Berkeley. Hydraulic Engineering Laboratory); HEL-2-28...
UCSC Science GC398.L44 1970 Library has: v. A-B.

10. Wilde, P. (Pat) 1935-
Chemical oceanology / Pat Wilde (Instructor), University of
California, Berkeley. Berkeley, Calif.: Committee on Ocean
Engineering, College of Engineering, University of California,
[1973?].
UCB WRCA 96 K2-1 v.2

12. MAP
Wilde, P. (Pat) 1935-
Oceanographic data off central California: 37 to 40 North including
the Delgada deep sea fan / by P. Wilde, W.R. Normark, and T.E.
Chase; assisted by L. Hydock [et al]. 1st ed., April 1976.
Berkeley: Lawrence Berkeley Laboratory, Energy and Environment
Division, University of California, 1976.
Scale 1:864,518 at 38 latitude ;.
Series title: LBL publication ; 92.
UCSC McHenry G9236.C7 1976.W5 Maps

Appendix H. Florida State University Movies at UCSC

From: Alex Pang [pang@cse.ucsc.edu]

There are 43Megs worth of movies on shiva in pang/fsu. Use "lookat filename" to view the data sets. Here is a brief description of what they are:

```
elsstssh.anom - eq. Pacific SST and SSH anomalies for early 1980's

sst.pac.mov - Pacific SST climatology

wind.pac.mov - FSU Pacific wind climatology

tuna-1.mov - ) two years of FSU NE Pacific ocean model with tuna data
tuna-2.mov - )

layer1.mov - FSU Indian ocean model of Bay of Bengal, upper layer only

pacbuoy40.mov - Pacific ocean drifting buoy movie
```

Appendix I. DATA SETS on NASA CD-ROM

From: Cheng Tang [cheng@cse.ucsc.edu]

Disk 1 of the CDs from NASA is mounted on [willow:/mnt] for those need access the data. I also installed some data and mainly the documentation and software for accessing the data on oak [/onr/data/nasa]. There is a README in this directory to help you get started. Currently there are five subdirectories and there is a README in each of them. General info is provided below for those who just want to know what's going on:

I.1 The Greenhouse Effect Detection Experiment 1992 Update

These CD-ROM disks have been produced by NASA's Climate Data System (NCDS) staff at Goddard Space Flight Center. NCDS has recently been integrated into Goddard's Distributed Active Archive Center (DAAC) for the Earth Observing System Data Information System (EOSDIS). The GEDEX program is a part of NASA's involvement in the International Space Year activities.

The data contained on these disks were selected at the GEDEX Workshop in July 1991 as being of primary relevance to investigations of global warming. They are all in a uniform format (Common Data Format, or CDF) widely used in the climate research community and reside on two disks. The first disk holds temperature, solar irradiance, cloud and radiation budget data. Atmospheric constituent data reside on the second disk. Several data sets have been updated since the initial release in March 1992.

Full software installation instructions are given in the file, SOFTWARE.DOC, in the DOCUMENT subdirectory.

The data sets from NASA CD-ROM include surface, upper air, and satellite derived measurements of temperature, solar irradiance, clouds, greenhouse gas, fluxes, albedo, aerosols, ozone, and water vapor. Many of the data sets provide global coverage. The spacial resolutions vary from zonal to 2.5 degree grids. Temporal coverage also varies. Some surface station data sets cover more than 100 years, while most of the satellite data sets covert only the most 12 years. Temporal resolution, for most data sets, is monthly.

A summary of all data sets and the instruments used to obtain the data is in *.det files on oak; [/onr/data/nasa/detailed].

I.2 Summary of Satellite Sensors in NEONS Database

Information describing satellites and their sensors from the NEONS database is in [cheng/info.chuck]. For LLT (lattitude longitude time - or observation report) data, we don't store information about the sensors that collect the data. Instead we store some information that describes the station that collected the data like where it is. We store information describing the parameters and levels, not which sensor collected it.

For more information contact Chuck Stein at NRL Monterey [stein@nrlmry.navy.mil], (408) 647-4798, or at Mirror Imaging [mirror@cats.ucsc.edu], (408) 458-2731.

Index