10.5 THE *NEW* NATIONAL COASTAL DATA DEVELOPMENT CENTER: A STATUS REPORT

Joe Stinus, Director
NOAA/National Coastal Data Development Center, Stennis Space Center, MS

Fred Klein and Amy Sheridan,
Mitretek Systems, McLean, VA

1. INTRODUCTION

Observations and services along the coast has been recognized as an important national priority. The complexity of the coastal regime and the need to predict natural and human environmental impacts and their consequences underscore the critical importance of our understanding of coastal processes. The higher resolution measurements of the dynamic coastal environment generate numerous data types of a magnitude requiring a rigorous and efficient storage capability and access satisfied only by a state-of-the-art data center. A strong link between the data collectors and the users is a primary focus. New tools and rigorous documentation are the two hallmarks of coastal data service.

In response to this challenge, Congress provided funding to establish the National Coastal Data Development Center (NCDDC) at Stennis Space Center, Mississippi. Nearly ready to open, the NCDDC will be one of NOAA's environmental data centers and provide for the archive of and access to long-term coastal data records. It will organizationally be under the National Environmental Satellite, Data, and Information Service (NESDIS), and NCDDC will work closely with many of the other Federal/state/local agencies, academic institutions, non-profit organizations, and the private sector to create a unified, long-term database of coastal data sets. Initially, the NCDDC will focus on improving the quality and access to physical and geophysical data sets, such as water levels, bathymetry, winds, waves, and SST. Over time, chemical (i.e., nutrients) and biological (i.e., harmful algal blooms), as well as many other data types will also be addressed. NCDDC staff will use established and emerging technologies to access and integrate data stored in geographically distributed repositories in heterogeneous formats. Data and information accessibility will increase as NCDDC implements a virtual data system linking to the NOAA data centers and other repositories of coastal environmental data. This will provide data for applications that will bring about interoperability.

Interoperability has been a desired objective of agencies involved with environmental and other data management functions and responsibilities. Architectural designs and considerations to enhance and achieve interoperability have evolved and progressed commensurate with the evolution of technology. Government, Industry and Academia have recognized that a multiple tiered architecture, with a data broker middleware as the central tier, is the solution to interoperability. Current technology, specifically object oriented programming techniques and languages, support the design and implementation of the desired multiple tiered, broker architecture for a prototype design and implementation of advanced data management techniques for NCDDC. Certification as a FGDC node will be achieved in late FY2001 and the center is expected to open at the end of March 2002.

2. MISSION AND FUNCTIONS

To meet the above mission goals the NCDDC will initially concentrate its efforts in four main areas:

1. Discovery/Cataloging of Coastal Data
2. Data Quality Assurance through data integration and fusion
3. Improved access to coastal data
4. New product generation

2.1 *Discovery/Cataloging of Coastal Data*

The efforts towards locating and preparing complex coastal data for access and potential storage is the main function of four regional liaison officers (LOs) assigned to the NCDDC yet supporting other data centers as well. The LOs serve as principal liaisons with constituents on each coast, East, West, Gulf coasts, and Pacific Islands. They identify and acquire coastal data for submission to the appropriate data center. Through *Project Accelerated Coastal Community Environmental Science Service* (ACCESS) [1,2] workshops, conducted throughout each geographic region, they identify local sources of data that require processing by NOAA, and develop a plan to acquire the data and integrate it into the NCDDC catalog. They also promote partnerships for specific projects such as data exchange, and product development. These partnerships also promote excellence in use of coastal data within community-based groups and educational institutions.

All of the expected results from partnerships, and data discovery will bring out a multitude of complex data types that must be cataloged into the NOAA data centers. Unique data types such as those that support risk assessment or monitoring efforts on coral reefs, Harmful Algal Blooms (HAB), and coastal storms will necessitate unique data cataloging requirements. More and more data will need to be placed into electronic catalogs to

* *Corresponding author address*:
Fred Klein and Amy Sheridan, Mitretek Systems,
McLean, Virginia 22102-7400
e-mail: fred.klein@mitretek.org.;
amy.sheridan@mitretek.org

ensure ease of access.

The NCDDC development will take advantage of and integrate distributed object and other new technologies to establish a "virtual catalog". This virtual catalog will provide electronic access to data via the Internet allowing data to reside almost anywhere. This will expose a larger realm of data, both inside and outside of NOAA, using a modular and layered approach pointing the way for the customer. The virtual catalog will provide directory-type services, in that it will point to other systems that catalog coastal environmental data or specific locations of data. This catalog may serve not only as directory and "normal" catalog services but also will provide access to other sources of data.

The results of a data discovery effort will be the basis for the directory and virtual catalog entries. Metadata will be evaluated for completeness and additional metadata added as needed. If no metadata exists in the data set, NCDDC personnel will work with the data set owner to develop metadata. NCDDC catalog browse products may include simple images, data plots, and other methods of summarizing data set contents. The cataloging and access process will integrate existing catalogs from other NOAA data providers.

2.2 *Data Quality Assurance (QA)*

In addition to routine statistical analyses, one of the main data processing innovations that will be implemented at the NCDDC is the use of numerical models to improve the quality of coastal data, to enhance the spatial and temporal resolution of observed data, and to allow the useful integration of satellite and disparate *in situ* data sets. Remote sensing provides spatially synoptic "snap shots" of surface properties. The strength of *in situ* sensing is in its capacity to provide high-resolution time series in three dimensions. Numerical models are based on human understanding of the processes controlling spatial and temporal variations in the physical variables in the coastal ocean. Data assimilation methods are used to merge observations with numerical models. The observational data are used to derive initial conditions for the models and to produce dynamically consistent interpolations and extrapolations of the observed coastal variables.

Models provide a dynamical basis for quality assessment of observational data as was demonstrated in the National Ocean Service's Long Island Sound (LIS) Project. During this project, a numerical model detected an inconsistency between observational data in the LIS. Model simulations established that the mean transport of water in the East River must be about 450 m^3/s. The historical observations of mean water level elevation differences between water level stations at the ends of the East River showed no difference in water levels. It was established using theoretical calculations that there must be a 9cm difference in the mean elevations to support the calculated current velocities. When the observational data were rechecked, the new datum calculations exactly corresponded to the elevation gradient predicted by the model. The model was used effectively to establish an integrated picture of the hydrodynamics in the East River, and was instrumental in establishing the correct environmental management plan to alleviate the influence of nutrient transports on hypoxia in Western LIS. NCDDC will employ model-based data fusion QA methodologies where possible to check coastal observations for consistency and accuracy.

2.3 *Improved Data Access*

The NCDDC will also employ advanced object oriented computing techniques to improve access to coastal marine data. Presently, there is not a general consensus within the marine data user community on hardware platforms, operating systems, network protocols, or data formats. Consequently, acquisition of a variety of marine data from multiple data sources frequently requires users to write application specific interfaces to each of the various data sources in order to use these data.

By utilizing an advanced "middleware" approach that employs distributed object computing techniques coupled with spatial data models, the NCDDC will leverage efforts of Data Exchange Infrastructure (DEI) research into a process to access distributed data stored in heterogeneous formats in different locations. The NCDDC will be able to acquire data for users from a variety of sources and provide these data in several fundamental views using conventional data formats and evolving formats: regularly spaced grids (e.g., rectilinear grids), imagery (e.g., synthetic aperture radar, or WSR-88D weather radar), irregularly spaced grids (e.g., point and line observations, or curvilinear grids), and vector representation (e.g., a data model based scheme to represent common values of an irregular shaped feature as a contiguous line such as shorelines or eddies). The user will then only have to write application specific interfaces for these general data "types" rather than for multiple specific data sources/formats. Also, changes in the source databases (e.g., format, new observation types) will be transparent to the user because the middleware will still provide these new or modified data in the common view (i.e., regular and irregular grids, and imagery) as noted above. The successful implementation and proof-of-concept testing of this new middleware approach at NCDDC will allow the National Oceanographic Data Center (NODC)/NESDIS to virtually link all of NOAA's data assets, as well as the data holdings of many other marine data organizations (e.g., Navy, US Geological Survey, Environmental Protection Agency, and the US Army Corps of Engineers).

2.4 *New Product Generation*

NCDDC products will be stored and managed at NCDDC even though data and other products will come from other data centers. Data will be supplied via numerous products such as observations, Geospatial Information System (GIS) applications, trendlines, documents for text data, and other products as requested. Many new data sets will be discovered by the liaison officers to try to piece together the environmental picture of coastal regions.

For example, under the Coastal Risk Atlas (CRA) pilot

project, NCDDC is working with the NOAA Coastal Services Center (CSC) to produce products using the North Carolina Community Vulnerability Assessment Tool (CVAT). Recently, the CSC developed a community vulnerability assessment methodology for use in identifying key vulnerabilities facing coastal communities. The CVAT is an informational aid designed to assist communities in their efforts to reduce hazard(s) vulnerability through strategies relating to awareness, education and mitigation. This CD-ROM based product contains a methodology that helps State and local governments determine and prioritize their localities' vulnerabilities to coastal hazards using data sets of infrastructure and environmental parameters. Physical factors such as the location of critical facilities and infrastructure relative to high-risk areas, the distribution of vulnerable populations such as the elderly, poor and under-insured, significant environmental resources and the vulnerability of primary economic sectors are all included as issues for consideration. This methodology will be used for initial guidance and as a template in the risk atlas for information searches, providing modern data access tools, and developing application oriented results.

The Coastal Risk Atlas is one example of the NCDDC pilot projects.

3. PILOT PROJECTS

NCDDC is participating in four pilot projects; Coastal Risk Atlas (CRA), Harmful Algal Bloom (HAB) Support, Coral Reef Data, and Coast Watch Hypoxia Warning. In each project, the NCDDC will be helping to discover and provide access to coastal data. The projects will continue into FY2002, pending FY02 funding.

The CRA has been planned as a NCDDC pilot project to be performed jointly with the National Oceanic and Atmospheric Administration (NOAA) Coastal Services Center (CSC), FEMA, and USGS. The project is to provide identification and access to coastal data so that community vulnerability assessments can be performed using techniques developed and successfully applied to particular locations by CSC. Hazard vulnerability analyses and resultant mitigation measures will have significant economic, environmental, and quality of life benefits.

HABs are a principle source of marine biotoxins and a major cause of marine mortalities and increased human health risk. The NCDDC is actively involved in two HAB projects. The first is HABSOS (Harmful Algal BloomS Observing System), a partnership between government agencies (state and federal) and coastal research laboratories (state, federal and academic). NCDDC will house the Regional Information Center and the HABSOS Portal. It is expected that this regional project will contribute to the development of a sustained (in perpetuity) and integrated (measuring physical, biological, and chemical variables synoptically in time and space) Coastal Ocean Observing System. In the second HAB task, NCDDC is assisting the National Oceanographic Data Center (NODC) in developing a system to synthesize data from monitoring and research programs nationwide to assist in HAB management and research. The HAB-Data Management System (HAB-DMS) will provide access to physical, chemical, and biological data acquired from many disparate sources, through an integrated system, for HAB decision support and model development.

Coral reefs cover less than 1% of the earth's surface, yet reefs and their associated mangrove, seagrass, and other habitats are the world's most biologically diverse marine ecosystems. Scientific information from coral reef research and monitoring help resource managers make informed decisions to reverse reef degradation. NCDDC will provide state-of-the-art web based user access to comprehensive and up-to-date diverse coral reef data that reside in numerous locations. Utilizing NCDDC's developing Internet data identification and access technologies will benefit this pilot project. An NCDDC goal is providing data for use with modern analysis tools, such as GIS, that may be important project elements. Publication of digital data sets and integrated publications with text, data and images are anticipated as new products and services.

Scientific studies show that the hypoxic (low dissolved oxygen) zone on the floor of the Louisiana Inner Shelf has increased in frequency, duration, and geographical extent since the 1950's. Through this pilot project, NCDDC will be able to quality control the data as it is gathered onboard ships. The near real time view of data will be beneficial to the fisheries along Texas and Louisiana. NCDDC will also rescue 15 years of data that needs to be processed with metadata and quality checked.

4. TECHNOLOGY

The NCDDC will employ a multiple tiered, data management architecture. However, architectural design decisions and progressive technology were applied to physically separate the tiers to accommodate the natural infringement of tier identity. The fundamental architectural consideration was a middleware data broker that physically separated the remaining tiers. This design decision accounted for the lack of clear tier definition and allowed each tier to assume the identity of a data provider or a data user, as desired or required. Distributed object computing techniques coupled with the power of a spatial data model were the technologies that were applied to the middleware development. Fig. 1 illustrates a simplified example of brokered architecture.

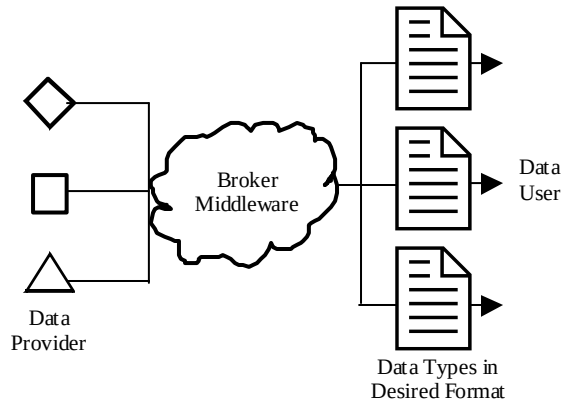

Fig. 1. Simplified illustration of a brokered architecture.

Through a NOAA approved and compliant portal, the user will gain access to the NCDDC web page. This web-based utility will provide the means for the user to identify the geographic region, data type, time and other search parameter such as key words to accurately define the desired search criteria. When the search criteria is completed, a query is launched and discovered data, registered within the middleware by Metadata, will be written to the NCDDC catalog. The user can then select a desired data set, review the accompanying Metadata, launch a request to retrieve the data, and finally view or download the requested data for further use.

The simplicity of the above user interactions is intended to demonstrate the power of the NCDDC prototype architecture. Technology issues such as programming languages, interface development processes, data model schemes are not and should not be issues of concern to a user. Simplicity is the key. The user makes a request for data. The data are delivered in a consistent and expected manner.

This is accomplished within the NCDDC primarily through the adopted concept of managing data. By implementing the architecture design criteria of a brokered middleware as the data delivery mechanism, the NCDDC has successfully separated the data provider from the data user thereby at least minimizing, and in most case eliminating, the constraints associated with coupled data delivery processes.

5. SUMMARY AND CONCLUSIONS

Unique, numerous, and high resolution environmental data from the dynamic coastal environment will require a rigorous and efficient access capability satisfied only by a state-of-the-art data center. The NCDDC will satisfy this need by establishing itself as an electronic data center and a gateway to coastal data holdings across the U.S. An initial operating capability, open to the public is planned for March 2002.

As a data center using new data management technology, the NCDDC will discover and catalog data, provide quality assurance through data integration and modeling, and make new products available. Many of the cataloged data sets will be derived from the four pilot projects. These products will incorporate modeled information, new trends and time series analysis, and GIS to allow users to develop tools and conduct the research necessary to improve and maintain the state of our coastal environment.

ACKNOWLEDGMENTS

The authors would like to express their appreciation to John Ellis and Susan Starke of the Naval Research Lab, Stennis Space Center, and Jim Etro and Andy Lomax of Anteon Inc. It is their diligent efforts to bring about a simple yet robust method of data retrieval through object-oriented techniques that will significantly improve the operations of this center.

ENDNOTES AND REFERENCES

[1] A Project for Community Outreach, an Overview, Project ACCESS, NOAA Atlantic Oceanographic and Meteorological Laboratory, March 2000.

[2] Report of the Workshop, Coastal Storms, NOAA National Coastal Data Development Center, March 2001.