Section 4.

Next Steps

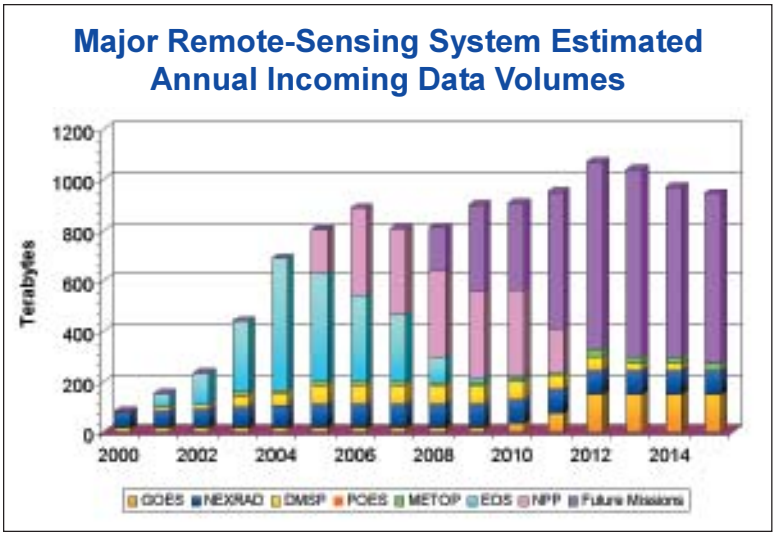
NOAA's constituents require multidisciplinary, *in situ*, and remotely-sensed environmental data and information that are fully integrated. This integration must span across political and geographic boundaries, scientific disciplines, agencies, projects, programs, instruments, and temporal and spatial scales. These requirements have been articulated in numerous NOAA Workshops; in reports by the President's Committee of Advisors on Science and Technology, *Teaming with Life: Investing in Science to Understand and Use America's Living Capital* (1998); and in the Committee on Environment and Natural Resources, *Integrated Science for Sustainable Ecosystems* (1998).

Over the next 15 years, current and planned remote sensing systems will produce volumes of environmental data on a scale that has not been seen before in this country or in the world. Many of these systems will be the operational responsibility of NOAA—all will produce data that will come to NOAA to be preserved and made available to support a myriad of users. However, effective systems do not yet exist within NOAA to address the data management issues that are associated with these massive volumes of data.

Increasing environmental stewardship responsibilities of NOAA and its many partners will lead to substantial growth in numerous specialized data sets, such as ecosystem data. These specialized data sets—although not of the tremendous size of those resulting from satellite or weather radar sources—are numerous, diverse in content, and often quite complex in structure.

Now is the time to develop and implement an effective data management plan for the environmental data and the associated data and information products that will result from these systems, as well as future environmental data-gathering programs. This data

Satellite data acquisition will require advanced technology to cope with data flow.



management plan must also cover the historical *in situ* data sets, providing improved accessibility to these data.

This plan must evolve with the data-gathering programs and extend beyond the life of these programs. In addition, this plan must address the entire life cycle of the data—from initial data capture to final disposition, implementing an end-to-end data management approach, as discussed in Section III of this report. The data management plan must consider the requirements of all of the following:

- Observing platform (i.e., satellites, land stations, ships, aircraft, buoys) and instrument/sensor characteristics
- Data processing at the sensor platform and on the ground
- Metadata creation
- Data capture and delivery systems
- Data validation and calibration procedures
- Archive procedures and media migration
- Data reprocessing to maintain the currency of the data set and to allow new uses of data
- Data uses and limitations
- Dissemination and user access, as appropriate for the data set.

10-meter NOAA discus buoy and U.S. Coast Guard Cutter.

The goals of this data management plan must be to improve the efficiency of NOAA's data management activities; minimize NOAA's overall cost of data management per data set; and improve user services by enhancing the usefulness, timeliness, and accessibility of environmental data for the benefit of the user community and our Nation. This data management plan must result in comprehensive, multidisciplinary data and information systems that will incorporate the following:

- Expanded environmental data holdings (e.g., EOS data archives) with more rapid delivery methods
- Enhanced access to distributed data sets, regardless of physical location
- Spatially and temporally integrated data
- Enhanced user services that include the ability to electronically access and deliver all service requests, with

NOAA must develop an architecture for an integrated, national, environmental data access and archive system that incorporates end-to-end data management functions.

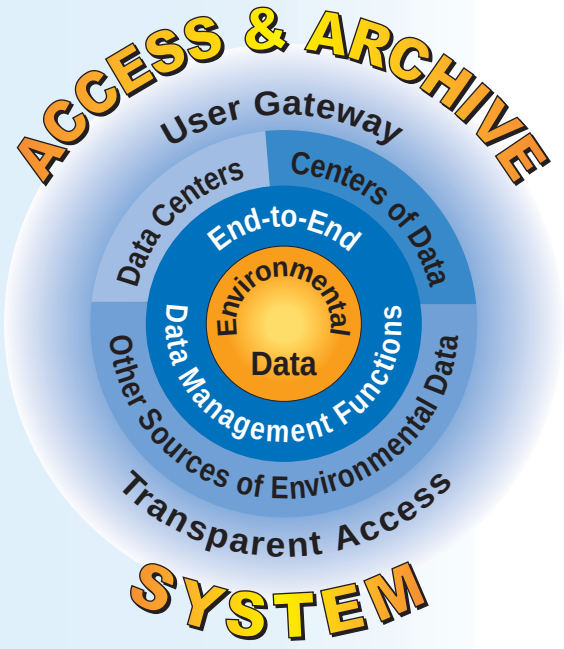
expert advice on the validity of the application of the data to user problems

- Support for new products and services that are developed in partnership with other Federal government agencies, academia, and the private sector
- Metadata based upon the appropriate standards, such as those promoted by NSDI and FGDC
- Real-time performance measures for each observing system based on data that users receive from the archives.

The Access and Archive System

The architecture of the new environmental data access and archive system must enable the ingest, archive, access, and long-term stewardship of the large-volume, remotely sensed data sets. This system must encompass the development of automated real-time or near-real-time systems to deliver data to NOAA, as well as automated ingest systems that integrate with those delivery systems to process the incoming data and prepare them for archive and dissemination.

Cataloging and validation processes must be automated where feasible, to improve efficiencies and reduce the growth of the resources required to complete these activities. Storage technologies must be optimized through the adoption of an ongoing technology refreshment program. New data access systems must be developed to ensure the usefulness and availability of archived data sets for both the current and future user communities, and to more effectively provide timely decision support capabilities. Migration and data rescue activities must be ongoing in order to prevent the unrecoverable loss of valuable data, ensuring the availability of these data for future generations of users. A real-time system must be implemented to provide information to observing system operators and data users on the "health" of instruments and observing systems, so this information can be applied to the development of new and improved instruments and observing systems.



The access and archive system should extend the concept of the NOAA Virtual Data System to all NOAA data collections. This will provide efficient access to the complex array of distributed data sets developed and managed by NOAA. The Web gateway will transparently direct users to data housed at the locations where the data are managed, regardless of the actual geographic location of the data set. This gateway will serve

both the large-volume array data sets and the smaller *in situ* land, coast, and ocean data sets. The system would enable access to those data sets that are continually updated, as well as appropriately controlling access to data sets where confidentiality requirements preclude open access.

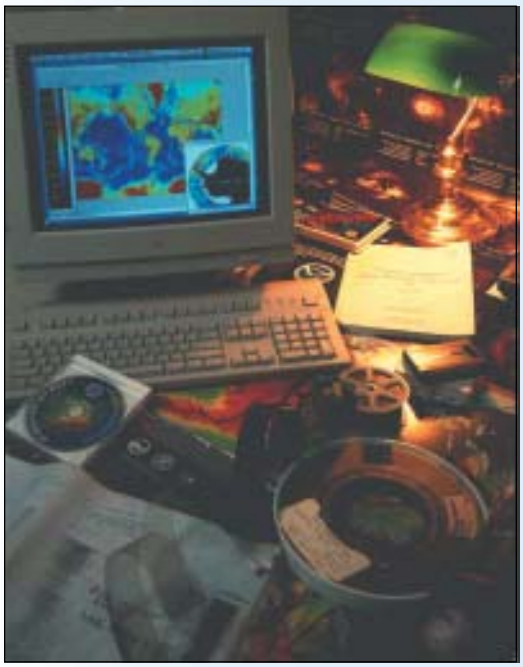
Technology will assist with many of these decisions, but policies, procedures, and schedules must be developed, reviewed, and modified on an ongoing basis. This policy infrastructure must include a diverse group of advisors to represent the broad spectrum of environmental data stakeholders—and they must be given the tools to assist them in their decision-making. The need is immediate. NOAA is committed to establishing these mechanisms for review.

There are three specific steps that must be undertaken immediately—*and in parallel*—in order to position NOAA for the onslaught of environmental data that is coming its way. These steps are:

Step 1

Build a Solid Foundation

The table on page 38, *Status of NOAA Environmental Data Management*, underscores NOAA's current inability to provide thorough, end-to-end data management. Until NOAA can ensure a healthy infrastructure that is capable of the ingest, archive, curation, and dissemination of those data sets it already maintains, NOAA cannot hope to do the same for the massive amounts of environmental data that will be produced by new observing systems over the next five years. The development of integrated architecture for this environmental data access and archive system must begin now.



For these large data sets, the system will provide a standardized archive system that is based upon an integrated distributed model. It will provide standardization in media, interfaces, formats, and processes for the ingest function, the metadata and cataloging function, the calibration and validation function, and the archive storage function. This standard platform will provide sufficient computing power to efficiently handle the archive processes of the very large data sets. Once implemented, all remotely-sensed satellite environmental data sets will be processed and archived by this system, regardless of the media/system used to deliver the data to NOAA Data Centers.

The standardized archive system will be integrated with a robust, large-volume, rapid-access data access system using an automated

The design and implementation of this integrated national environmental data access and archive system is critical to providing the most efficient organization and functional context in which NOAA's current data sets can be maintained.

data management system to manage on-line, near-line, and off-line data sets. The system must be capable of receiving a user's on-line data request, automatically processing the request, organizing the requested data, and providing the requested data on the media that are appropriate for the request. In cases where extremely large data sets are requested, the data may be made available on a super computing system to which the users would be allowed to connect to perform their research. By providing users with access to a super-computing system, NOAA will overcome several problems associated with the delivery of large data sets to users, as very large amounts of data are not suitable for on-line file transfer and NOAA may not have the specific media requested by the user. In addition, NOAA will realize a reduction in the types of media and mass storage subsystems it must maintain by providing users access to such a system.

This system will be built over time. Its design will be modular, allowing for the incremental integration of the data management functions of additional data sets, as needed. Building upon the NVDS, the system will initially link the polar satellite data with the active GOES and DMSP data, incorporating all of NOAA's end-to-end environmental data management functions for these data sets. In addition, the system will provide user access through the Web gateway to the diverse non-satellite data sets managed throughout NOAA, and will implement improved data management functions for these distributed data collections.

This integrated access and archive system will be a dynamic system—continually integrating improvements in technology and standards. Data migration activities will be ongoing—embracing the very technology that drives the size of NOAA's database upward—in order to prevent NOAA from being buried beneath advances in technology.

There is one additional requirement for a healthy foundation that is key to the success of the data management plan—the rejuvenation of our staff resources required to manage and utilize the new system, in order to process the current workload. Without these resources, NOAA cannot hope to keep pace with its current responsibilities, much less prepare adequately for its future responsibilities.

Step 2

Fix What's Broken

As the implementation of the new environmental data access and archive system begins, the network and data management functions for the remaining operational systems—those categorized as "historical" and "modernization" in the table, *Status of NOAA Environmental Data Management*, on page 38—must be rehabilitated into the technology and formats of the new system. This includes the following tasks:

- Revitalization of the observing, archive delivery, and ingest systems for the *in situ* ground systems
- Rescue of historical GOES, DMSP, and NEXRAD data
- Development of an electronic delivery and ingest system for NWS NEXRAD Doppler Radar data to provide near-real-time data delivery to the Data Centers, and reorganization of those data into a usable, accessible format
- Rescue of paper media, such as fishery logbooks, maps, hydrographic surveys, and photographic images
- Upgrade of the NWS Cooperative Network and Precipitation Network, and the development of a Climate Reference Network.

Substantial effort will be needed to improve compatibility among existing databases distributed throughout NOAA and to ensure that new data are also compatible. Presently, there are many data sets that can not easily be incorporated into a geographic information system (GIS) and layered with other relevant geospatial data sets. Also, a

GIS image of water temperature variation around the northern region of Channel Islands National Marine Sanctuary. The image was created with layers of satellite, topographic, and ecosystems data.

significant number of data sets require the development of adequate metadata to describe and identify the data.

Building the system that will bring NOAA's environmental data into the future includes preparing for the large amounts of data that will be acquired using new technologies, while also ensuring that the many valuable data sets acquired in the past are not lost. There is a critical need to have the best information on current conditions for our environment and living marine resources, as well as historical perspectives of conditions in the past.

Many important older data sets (e.g., commercial fishing logbooks, bathymetric surveys, water-level data, and shoreline manuscripts) are on paper media that are not only difficult to access, but are in serious danger of being lost due to their physical deterioration. Aerial photography, maps, lists of data from numerous sensors, and many other types of data are also in dire need of conversion into accessible data formats on accessible digital media. These older data are an irreplaceable component of NOAA's "Treasures at Risk."

These tasks must be performed in parallel with the implementation of the new environmental data access and archive system. In addition, adequate resources must be provided to enable the completion of these tasks, so that NOAA can reap the long-term benefits that will ensue from this revitalization of NOAA's historical data sets and their associated data capture and data delivery systems.

These data sets are vital to the creation of the long-term record needed for ongoing environmental change research.

Step 3

Prepare for the Future

Managing the massive volumes of environmental data that will come from the future data-gathering systems will build upon the foundation and rehabilitation of the existing systems, networks, and archived data sets. As the new access and archive system is being developed and implemented, the requirements for the Earth Observing System (EOS) data will be developed and tested for integration with the system. In this way, the EOS experience will be a pathfinder system—providing risk-reduction for the new information management systems that are planned to go on-line over the next 10 years (e.g., NPP and NPOESS).

Having a modular design, this new system will provide for the integration of the processes required for the ingest, archive, curation,

access, and dissemination of future data sets. Also, the system will accommodate the ongoing migration to new storage, access, and processing technologies for all data sets under its management. These ongoing integration and migration activities will guarantee the continued accessibility of these vital data sets.

Once a sound foundation has been ensured, NOAA will begin linking other non-NOAA data centers to its environmental decision support network, including States and local municipalities that are collecting coastal and fisheries data, marine laboratories, and the NASA Distributed Active Archives Centers.

Conclusion

With the implementation of a NOAA-wide planning process—plus the necessary additional capital investments for data rescue and management—NOAA will be able to develop the infrastructure required to keep pace with the immense environmental data and information management demands of the 21st Century. Backlogs of unprocessed and uncataloged data and information from the large volume remote-sensing data sets and historic paper records that should be incorporated into the archives will be eliminated.

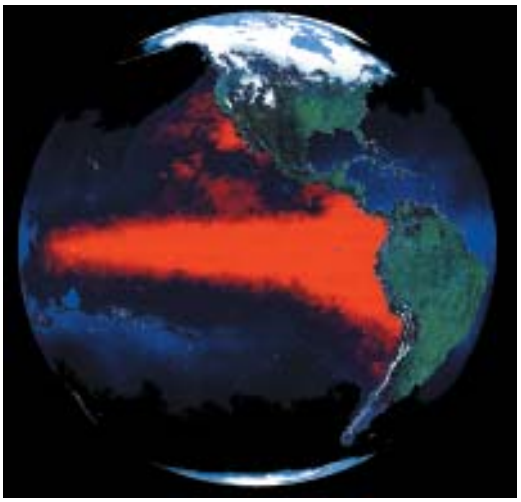
Once data-generating projects are concluded and Principal Investigators move on to new endeavors, the support systems for the data and information collected by those projects needs to ensure that the data are be saved, protected, and made available. Future researchers, industry, and the Nation will be able to subsequently use these data to solve new environmental problems and to address new environmental issues.

With the adoption of the goals and objectives described here, research scientists and managers will be assured of access to environmental data and information in useful forms. In addition, NOAA will develop the infrastructure and the resources required to preserve the massive streams of data and information generated by new *in situ* and remote sensing systems, for use in major Earth system and natural resource research and monitoring programs.

The public policy of the Nation has been to support an extensive national infrastructure of environmental observing system assets. The Nation has invested—and continues to invest—billions of dollars to develop systems that capture environmental data for

various vital uses. The trend toward needing more and better environmental information—and understanding the long-term implications for the climate in which we live—mandate that we take good care of the data that are provided by our multibillion-dollar observing systems. Implementation of this data management vision is time-critical.

By using a coordinated approach for managing environmental data throughout the data life cycle, NOAA's national treasure of environmental data will no longer be at risk.



Composite of NOAA's satellite views of the El Niño event, 1997–1998. Many parts of NOAA are involved in research, assessment, and monitoring of these phenomena. El Niño-related climate variations often have devastating impacts. Preservation and analysis of NOAA's environmental data will be crucial in predicting future events.

A change in our climate...is taking place very sensibly. Both heats and colds are becoming much more moderate within the memory of even the middle-aged. Snows are less frequent and less deep. They do not often lie, below the mountains, more than one, two, or three days, and very rarely a week. They are remembered to have been formerly frequent, deep, and of long continuance. The elderly inform me, the earth used to be covered with snow about three months in every year. The rivers, which then seldom failed to freeze over in the course of the winter, scarcely ever do so now...

The same opinion prevails as to Europe; and the facts gleaned from history give reason to believe that, since the time of Augustus Caesar, the climate of Italy, for example, has changed regularly, at the rate of 1 degree of Fahrenheit's thermometer for every century. May we not hope that the methods invented in later time for measuring with accuracy the degrees of heat and cold, and the observations which have been and will be made and preserved, will at length ascertain this curious fact in physical history?

Thomas Jefferson

Notes on Virginia, 1787