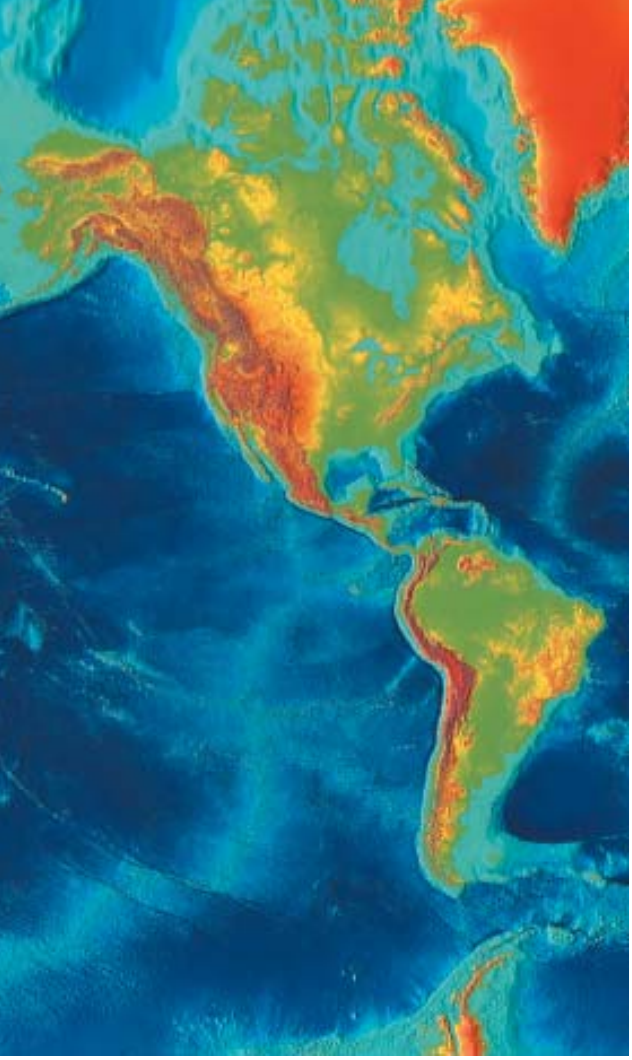


Portion of NOAA's computer-generated, color-shaded relief map, Surface of the Earth *(2000).*

Executive SUMMARY

The quality of our lives and the health of our environment will be determined by the choices we make today. There are clear connections among the environment, the economy, and society. Economic growth, maintenance of environmental quality, and wise use of resources must go hand-in-hand to ensure a rising standard of living for all Americans.

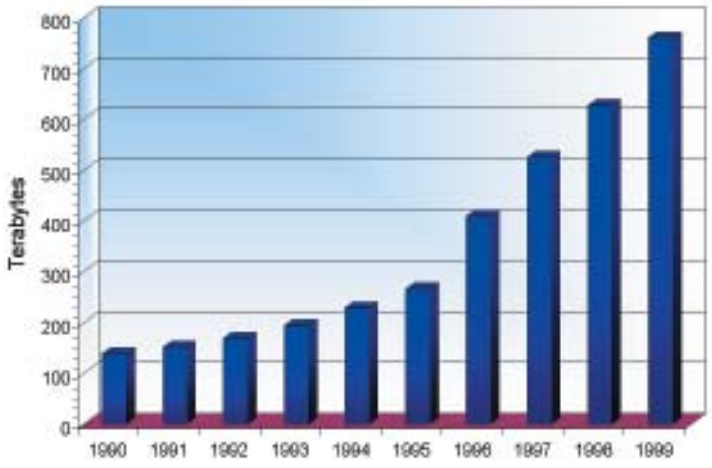
The National Oceanic and Atmospheric Administration (NOAA) has the unique mission within the Federal Government of promoting global environmental stewardship by describing, assessing, monitoring, and predicting Earth's environment. Management and information services involving the acquisition, archive, access, integration, and dissemination of environmental data are critical to the success of this mission.

NOAA spends almost a billion dollars each year collecting environmental data from around the world and from Space in support of its mission. In the 21st Century, NOAA envisions its data and information products to be available to the Nation as part of a national decision support system for:

- Saving lives and protecting property
- Promulgating public policy for environmental stewardship
- Managing and conserving living marine resources
- Enhancing the economic prosperity and quality of life in the United States.

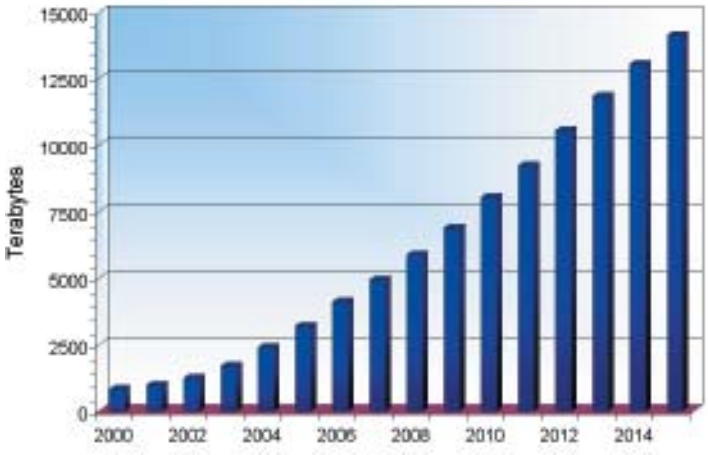
NOAA's vast data observations and measurements are collected and stored in many different distributed facilities across the country, some of which are responsible for the perpetual stewardship, archiving, and dissemination of environmental data.

The NOAA National Data Centers (NNDC) have the unique responsibility for the long-term management and stewardship of the bulk of NOAA's data—in addition to environmental data collected by other Federal agencies, countries, and research programs—for use in resolving today's and tomorrow's environmental issues. Also, there are numerous distributed Centers of Data with data collections located throughout NOAA's line offices, and programs that are responsible for the management of data sets developed in the process of fulfilling their particular environmental missions and operational responsibilities.



Archive Growth During the 1990s

Historical archive growth will be dwarfed by projected growth.



15 Year Projected Archive Growth

MORE DATA to Manage

New knowledge, new technologies, and new issues are driving increases in environmental data-gathering efforts. Programs that address stresses on the Nation's coastal areas, the recovery of populations of endangered species and depleted fish stocks, and the mitigation of the effects of natural hazards and human activities have become increasingly important. With the recognition that phenomena that occur globally may affect us locally, data for our world as a whole acquired through technological advances—particularly in the development of remote sensing systems—have gained particular significance.

Throughout the 1990s, the NOAA archives have experienced a steady growth in digital data, more than quadrupling in volume by 1999 to more than 760 terabytes. By 2005, NOAA's holdings will be nearly 5 times this amount, growing to almost 20 times the 1999 volume over the next 15 years. This massive influx of environmental data and information far exceeds NOAA's archive and access capabilities.

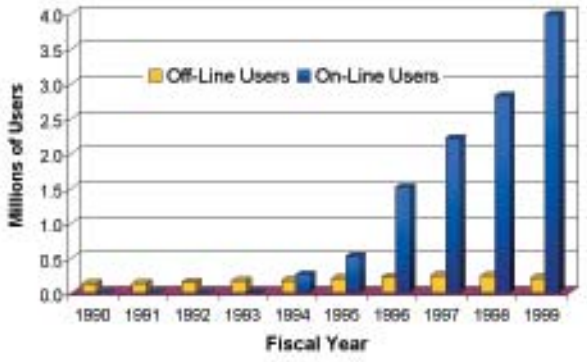
Accurate climate information, including temperature and precipitation data, is critical for developing weather-related decision support systems tailored to energy business need.

Bill Coley
President,
Duke Power Company

MORE DEMAND for Data Products

Satellite operators are using new and exciting techniques to image large sections of Earth and its environment. The techniques generate huge amounts of data that are difficult to interpret and understand. However, products derived from these images are extremely informative both to scientists and the public. For example, satellite-generated microwave images can be used to compute the rate of rainfall over remote ocean and land areas. These products also produce huge volumes of archived data.

In the past, the National Weather Service (NWS) was one of the few NOAA programs which provided the public with real-time access to NOAA environmental data on demand. Today there are user demands for near-real-time information in virtually *all* of NOAA's programs. These demands include timely information about natural and man-made disasters such as coastal hazardous material incidents, harmful algal blooms, wildfires, and volcanic eruptions.



Requests for Data and Information

MORE USERS to Support

The advent of the Internet brought an enormous increase in the numbers of users requesting and accessing NOAA environmental data and information. Off-line data requests doubled from 1990 to 1997 and then decreased by a third, while the number of on-line users has increased exponentially since 1995.

NOAA's customers prefer on-line ordering.

Users are demanding on-line ordering and on-line search and browse capabilities with electronic file transfer for data delivery. Furthermore, users are no longer content to wait days or weeks for their data or information requests to be processed and delivered. Increasingly, users want information rather than data, as information and products derived from observations are frequently more useful to business and industry than the original data.

New user groups require access to data within minutes, hours, or days of the recorded environmental observations. For example, the global financial community is using new financial instruments to manage the risk of losses due to variabilities in weather and climate. Their trades are based on information about historical trends obtained from data acquired in the past, coupled with predictions of the future, in addition to information from the current observing systems.

User Spectrum Expanding

- ***Attorneys*** *using meteorology in court cases.*
- ***Insurance*** *using hazards data for rate determination.*
- ***Agribusiness*** *using ecosystem data for crop management.*
- ***Navigators*** *using nautical charts for safety.*
- ***Fishermen*** *using oceanographic data for stock locations.*

MORE ACCESS to NOAA's Data

Environmental scientists and advisors have a critical need for a long environmental record of historical and recent global data. These data are used to assess long-term trends, evaluate current status, and predict future conditions and events. The public, industry, and policy makers ask: Is global climate change really occurring? Are the oceans getting warmer? What causes harmful algal blooms? Why did this fishery collapse? What is happening to the Nation's ecosystems? How do El Niño and La Niña affect the weather? How bad is this event compared to earlier events? How long will the drought last? How severe will winter be this year?

Issues such as climate change need global-scale data in addition to a long environmental record in order to identify the trends. The timeliness of access to these data and the completeness of the available environmental record are crucial to the formulation of answers to these and other pressing questions.

NOAA made great strides during the 1990s in streamlining operations and in using technology to support the ongoing operations of its distributed data facilities. The development of the interactive data access systems such as the NOAAServer and the NNDC Virtual Data System (NVDS) enabled many data sets to be made available to users via the Internet, and enabled more user requests to be handled by a dwindling staff. However, vast quantities of data are available only via labor-intensive, manual search and retrieval activities.

Many archived data sets that were in danger of being lost due to aging and deteriorating storage media have been "rescued" through migration to modern media. However, many more NOAA data sets need to be rescued. Several of NOAA's most valuable "modern" data sets—including Geostationary Operational Environmental Satellite (GOES) data sets—are inaccessible due to obsolete storage media and reader technology. In addition, many important older data sets such as commercial fishing logbooks, bathymetric surveys, 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.

In the past two decades, NOAA's archives have increased 38-fold and user requests have grown 44-fold, while NNDC staffing has decreased by 45%.

	1979	1999
Archives	20 TB	760 TB
User Requests (Accesses)	95,400	4,200,530
FTE Staff	582	321

All of these data sets are in dire need of conversion into accessible data formats. The data sets need to be transferred onto modern, accessible media. These data sets are important in monitoring the frequency and severity of tornadoes, floods, hurricanes, and droughts; in managing our coastal zones; and in managing and conserving stocks of fish and other living marine resources. The increasing volume of incoming data *also* affects NOAA's data-rescue activities. Unfortunately, as NOAA's ingest and storage capabilities fall farther and farther behind, so do its rescue and migration capabilities.

Inflation has taken its toll on the NOAA data and information budgets, and demands on staff have reached a point of over-saturation. Although spending on environmental research systems has increased over the years, data management and archive responsibilities for these programs have been undertaken without comparable increases in funding for basic operations—salaries and infrastructure. The result of this disparity is that NOAA is struggling to provide these functions for massive amounts of data. Consequently, the data are not available in a timely manner for continuing research.

Steps to 21st Century Environmental Data Management

Over the next 15 years, current and planned remote sensing observing systems will produce volumes of environmental data on a scale that has not been seen before. Data from these systems will come to NOAA to be preserved and made available to support a myriad of users.

Effective systems must be in place to address data management issues that are associated with massive volumes of data.

Effective systems must be in place to address data management issues that are associated with massive volumes of data. Therefore, NOAA must implement an architecture for an integrated, national environmental data access and archive system to support the ingest, archive, access, and long-term stewardship requirements of the large array, remotely-sensed data sets.

In addition, NOAA must extend the concept of the NNDC Virtual Data System (NVDS) to all NOAA data collections, providing efficient access to the complex array of distributed data sets developed and managed by NOAA.

NOAA must develop a handling process for these massive volumes of environmental data. Decisions must be made about which data sets will be stored on-line, on robotic systems, off-line, or in deep-archive. Schedules must be developed for data set migration and reanalysis (e.g., reprocessing). Technology will assist with many of these decisions, but policies, procedures, and schedules need to be developed, reviewed, and modified on an ongoing basis.

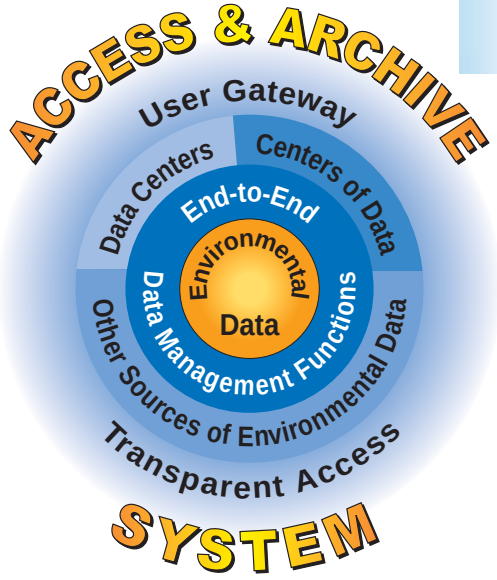
This policy infrastructure should include a diverse group of advisors to represent the broad spectrum of environmental data stakeholders. They must be given the tools to assist them in their decision-making. The need is immediate, and NOAA is committed to establishing these mechanisms for review.

There are several specific steps that should be undertaken to position NOAA for the onslaught of environmental data that is coming its way.

Step 1

Build a Solid Foundation

NOAA must ensure a healthy infrastructure that is capable of the ingest, archive, curation, access, and dissemination of data sets it already maintains. With the new access and archive architecture, NOAA must adopt a comprehensive data management strategy to accomplish two goals: (1) replace the assortment of aging and inefficient systems it currently operates, and (2) incorporate all of the data management functions listed in the table (next page).



NOAA needs to attain compliance with the appropriate policies, standards, and procedures—including those of the National Spatial Data Infrastructure (NSDI)—to ensure a common framework and minimize duplication. Building upon systems already in place, this strategy must include an Internet-based virtual "gateway" data access, regardless of the geographical storage location of the data.

The new environmental data management system must include the development and implementation of standardized archive methods, which will be integrated with a robust, large-volume, rapid-access storage and retrieval system that is capable of receiving a user's on-line data request, automatically processing that request, and providing the requested data via the most appropriate media. This system will provide standardization in media, interfaces, formats, and processes for the very large data sets produced by satellites and radars.

While many of NOAA's distributed data sets have lower requirements in terms of data volume, the new access and archive system must be capable of handling a very broad array of data types and formats for these data sets. Moreover, the system must provide additional access controls for these data, as some of the data sets contain proprietary information or are used to support navigational safety. The Internet-based virtual gateway of the new system will provide users with access to the smaller and varied data sets. For example, these capabilities will enable researchers to study and manage coastal and marine ecosystems and the organisms that live within them.

Status of NOAA Environmental Data Management

	Data Sets and Observations	End-to-End Environmental Data Management Functions							
		Planning	Collect or Rescue	Ingest	Metadata & Cataloging	Calibrate & Validate	Store	Access	Migrate
HISTORICAL	In Situ - Centers of Data	√	×	√	×	×	⊕	⊕	⊕
	NOAA National Data Centers	√	√	√	√	√	√	⊕	⊕
	COOP / USHCN	√	√	√	√	√	√	⊕	⊕
	GHCN	√	√	√	√	⊕	⊕	⊕	⊕
	CARDS / COADS	√	√	√	√	⊕	⊕	⊕	⊕
« MODERNIZATION »	DMSP	√	√	√	√	⊕	⊕	⊕	×
	POES	√	√	√	⊕	⊕	⊕	×	×
	ASOS	√	⊕	⊕	⊕	⊕	⊕	×	×
	NEXRAD	√	⊕	⊕	⊕	×	×	×	×
	GOES	√	⊕	⊕	⊕	×	×	×	×
FUTURE	New In Situ Land & Ocean Observing Systems	√	×	×	×	×	×	×	×
	NPP	⊕	×	×	×	×	×	×	×
	NPOESS	⊕	×	×	×	×	×	×	×
	EOS	⊕	×	×	×	×	×	×	×

√ = Can Do With Current Resources

⊕ = Need Incremental Resources

× = Requires Substantial Additional Resources

Note: See Table D-1 or Appendix E for details.

Step 2

Fix What's Broken

The existing operational systems must be rehabilitated. This includes:

- Revitalization of the observing networks, the archive delivery systems, and the ingest systems for the *in situ* ground systems
- Rescue of historical GOES, DMSP, and NEXRAD data
- Rescue of paper media, such as fishery logbooks, maps, hydrographic surveys, and photographic images
- Development of an electronic delivery and ingest system for the National Weather Service (NWS) NEXRAD Doppler Radar data, plus the organization of information into a usable, accessible format
- Upgrade of the NWS Cooperative Network and Precipitation Network and development of a Climate Reference Network

NOAA must improve data management to support seasonal forecasts. 1998 was one of the costliest natural hazard years for our Nation. The U.S. sustained seven disasters, each costing more than $1 billion.

Substantial effort will be needed to improve compatibility among existing databases, to improve data compatibility with geographic information systems, and to develop adequate metadata documentation to describe and identify the data sets and their contents.

Step 3

Prepare for the Future

As the new access and archive system is developed and implemented, the requirements for the National Aeronautics and Space Administration's (NASA) Earth Observing System (EOS) data will be developed and tested for integration with the new NOAA system. This will provide risk-reduction for the new systems that are planned to go on-line over the next 10 years.

Many more fish stocks will require assessment, decisions will need to be made in an ecosystem-based context, and more extensive data will be generated and used in managing the endangered species and essential fish habitat. Also, support for navigational safety is in the forefront as our Nation's maritime commerce will continue to grow.

With the major increases in management responsibilities related to living marine resources and coastal stewardship, data management and access requirements for NOAA's distributed data collections will necessitate a considerable investment in resources and skilled personnel as well.

Protecting NOAA's National Treasures

If additional resources became available, NOAA will be able to build the foundation required to keep pace with the immense data and information management demands of the 21st Century. The Nation has already invested billions of dollars in collecting these environmental treasures. By using this coordinated approach for managing environmental data throughout the data life cycle, this national treasure will no longer be at risk.

The challenge is to develop data management and archiving infrastructure and procedures that can handle the rapid increases in the volumes of scientific data, and at the same time maintain older archived data in any easily accessible, usable form. An important part of this challenge is to persuade policy makers that scientific data are indeed a precious resource that should be preserved and used broadly to advance science and to benefit society.

National Research Council, *Bits of Power: Issues in Global Access to Scientific Data.* National Academy Press, Washington, DC, 1997, p. 62.