Section 2.

The Accelerating Pace of Change

> *The fact that observational data are unique and not reproducible lends to the conclusion that they should be preserved as part of the historical record of the dynamic behavior of Earth and its inhabitants.*
>
> National Research Council, "Bits of Power: Issues in Global Access to Scientific Data." National Academy Press, Washington, DC, 1997.

NOAA spends approximately one billion dollars each year on observing systems that collect environmental data from all over the world. These data are used for a wide range of environmental prediction programs—from severe weather forecasting in which the data must be used within minutes of collection, to climate prediction programs that use data from the past 100 years to project the climate over the next 100 years. These data are also used to predict water levels for safety of navigation, predict fishery yields, and for the recovery of endangered species.

Within the Federal government, NOAA is responsible for the dissemination and archiving of environmental data. The agency has implemented two complementary approaches for its data-stewardship mission:

- NOAA National Data Centers, which are centralized repositories of data generated and used by NOAA and our many collaborators in the USA and around the world, and
- Distributed Centers of Data, which are maintained by various components of NOAA with specific responsibilities, such as monitoring tides and currents or managing fisheries.

This dual philosophy of data stewardship within NOAA has evolved because no one entity can address all of NOAA's data stewardship needs. The data sets are too diverse, and the expertise is too specialized and dispersed around the agency. As the demands for NOAA's data continue to expand, these different data-management entities must improve their collaboration for providing users with seamless access to their data sets, regardless of physical location, as well as in developing new data and information products.

The NOAA National Data Centers (NNDC) have the responsibility within NOAA to provide the perpetual stewardship of, and to be the primary repositories for, the greater part of NOAA's environmental data. The NNDC consists of three National Data Centers:

- **National Climatic Data Center (NCDC).** NCDC is the world's largest active archive of weather and climate data. The Geostationary Satellite Archive System (GSAS), operated through the Cooperative Institute for Meteorological Satellite Studies (CIMSS) with the University of Wisconsin, and the NOAA (Polar) Satellite Active Archive (SAA) are associated with NCDC.

- **National Oceanographic Data Center (NODC).** NODC is the world's largest active archive of publicly available oceanographic data. The award winning NOAA Library and the new National Coastal Data Development Center (in Mississippi) are associated with NODC.

- **National Geophysical Data Center (NGDC).** NGDC manages one of the world's largest and most diverse active archives of geophysical, space weather, and paleoclimatological data. The Snow and Ice Data Center, operated through the Cooperative Institute for Research in Environmental Science (CIRES) at the University of Colorado, is associated with NGDC.

These Data Centers manage environmental data collected over hundreds of years, providing thousands of years of Earth's environmental record. The archives (as of December 1999) contain data in many forms, including paper manuscripts, index cards, posters, and maps; rolls of 16mm and 35mm film; microfiche; and more than 760 terabytes of digital data. Digital data are stored on a variety of magnetic tapes, magnetic disks, and optical disk technologies, some of which date back more than 25 years.

NOAA's environmental data may be stored in a variety of formats.

The responsibilities and activities of NNDC are complemented by those of numerous distributed data collections throughout NOAA, operated by the various NOAA line offices. These distributed collections are located in geographically dispersed facilities around the country, which are usually related to the locations in which the data are actually collected, generated, and used to support ongoing operational responsibilities. Examples of NOAA's distributed Centers of Data include:

- **National Geodetic Survey (NGS) Center of Data.** NGS is responsible for maintaining, distributing, and archiving data and information related to the shape, size, and gravitational attraction of the Earth's surface. Data include horizontal and vertical reference station locations for the entire U.S. and the maintenance and monitoring of the Global Positioning System Continuously Operating Reference Stations (CORS).

- **National Marine Fisheries Service (NMFS) Centers of Data.** NMFS has five regional science centers located around the country. These centers generate, manage, and archive much of NOAA's biological data. These data are used to manage fisheries, with emphasis on protecting marine mammals and the recovery of endangered species.

- **Center for Operational Oceanographic Products and Services (CO-OPS) Center of Data.** This center acquires, analyzes, archives, and disseminates data on tides and currents.

It also operates a national quality-controlled system for real-time tide and current data and maintains the databases for the U.S. vertical benchmark system.

More detailed information on NOAA's distributed data collections is provided in Appendix B.

NOAA's earliest data collection efforts pre-date computers, when archiving was mostly of manuscript and analog records and disseminating archived data was to a relatively small scientific user community. New responsibilities have evolved with the growth of computer technology. The most pervasive changes affecting NOAA data management efforts in recent years have included the following:

- Enormous growth of the Internet and the tremendous demand for instant access via the Internet

- Advances in the use and number of remote sensing systems that provide immense amounts of data

- Implementation of a data sales policy in response to the Office of Management and Budget (OMB) Circular A-130, *Management of Federal Information Resources*

- Increased interest in understanding the environment, predicting environmental change, conserving living marine resources and mitigating potential environmental hazards.

The increasing use of electronic means for data collection, storage, manipulation, and dissemination is one of a number of broad and interrelated trends that have significant implications for access to data in the natural sciences. These trends include the following:

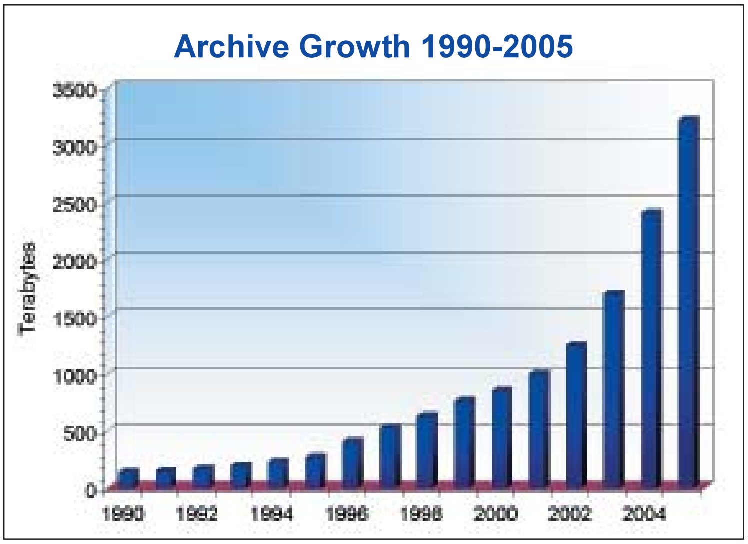
- √ *Rapid growth of the body of scientific data*
- √ *Development of large international research programs*
- √ *Insufficient funding for data management and preservation activities worldwide*
- √ *Decentralization of data management and distribution*
- √ *Electronic publication*
- √ *Increasing use of simulations and animations as scientific data*

National Research Council, "Bits of Power: Issues in Global Access to Scientific Data." National Academy Press, 1997, p. 57.

In the early 1990s, NOAA responded to these changes by developing a modernization plan for its data management responsibilities that included providing on-line access to digital databases and use of the Internet. In the mid-1990s, NOAA began implementing this plan, and today its implementation is well underway. However, change is inevitable and—even as these changes are being implemented—users, circumstances, and technology are redefining NOAA's data management mission. For example:

- The volume of digital data coming into NOAA is increasing exponentially and in complexity. The size of the digital archives has more than quadrupled during the 1990s and is expected to expand to almost five times that of the 1999 archives over the next five years. This massive influx of data is choking NOAA's ability to archive new data. This is also creating a backlog in NOAA's ability to archive and keep old data accessible via new media.

- Global environmental records need to scale to the local level to feed operational forecasts and hazard mitigation models.



Archive growth rates are accelerating, as new satellite systems come on line.

Related data sets of both historical and current data must be integrated (reprocessed) with incoming data into databases that share a common format so that they are usable and accessible. The goal is to create a continuous data record from the past to the present.

- New mandates are greatly expanding NOAA's responsibilities for managing and conserving living marine resources. In response to recent National Research Council recommendations, NOAA is working to increase the number of fish stocks assessed by 180%, and to conduct these stock assessments within a vastly more complex ecosystem context. There is also a rapid increase in the data and information needed to support NOAA activities under the Endangered Species Act and the Marine Mammal Protection Act.

- New user groups (such as those needing weather risk information) are evolving, and the size of the user community is increasing due to the usage of the Internet, the increased interest in environmental change, and the desire to mitigate environmental hazards. Growth of the user community is resulting in a corresponding increase in requests for data and information.

- The spectrum of users is much broader than was the case earlier in the decade. As the "interested public" and business user communities grow, the average level of sophistication and scientific expertise of the overall user community is decreasing. These phenomena result in NOAA being asked to provide increased information and different data products.

- Users are demanding seamless access to NOAA's data relevant to a particular issue, regardless of the physical location or the format of the 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 real-time or near-real-time information in virtually *all* of NOAA's programs.

These changes are severely straining the capabilities of NOAA's current resources to fulfill NOAA's data archive and dissemination mission.

NOAA is moving forward with the development of new observing systems and initiatives in response to the needs of the Nation.

Researchers use a hydraulic drill to collect a core from coral. Corals often live hundreds of years and are useful in documenting climatic change.

New Programs and Initiatives

Changes in NOAA's organization are reflecting the landscape of how its data management systems are viewed—not focusing specifically on the data, but on how the data support specific fundamental program capabilities. NOAA is moving forward with the development of new observing systems and initiatives in response to the needs of the Nation. These systems and initiatives are producing new data and driving the need to develop and maintain new, accessible data sets to be used in answering important questions and solving real problems. More and more, NOAA and the users of NOAA data are using geographic information systems (GIS) as a basis for overlaying economic and social information for local decision-makers.

The modernization of the NWS has resulted in new observing systems and the automation of surface observing systems. The Climate Prediction Center's new ability to issue seasonal to interannual forecasts—and the database on which it is founded—sets the stage for future climate initiatives.

The Climate Database Modernization Program, managed by the National Climatic Data Center (NCDC), is making major climate databases available via the World Wide Web, thus increasing the utilization of this national resource. To accomplish this, three objectives have been established. First, modernize the archives by creating digital databases containing complete and accurate climate data and metadata; second, update databases in a cost efficient and timely matter; and third, provide the infrastructure required so that access to these data is quick and convenient.

NCDC has begun digitizing key 19th and early 20th Century data including hourly precipitation values, daily weather observations, and climate summaries/extremes. Paper records, containing hourly weather observations, are being imaged with some 20 million pages currently converted to digital images.

Plans are now in place to have all incoming climate records received by NCDC to be rapidly and efficiently digitized. This will insure timely accessibility and also increase the quality of these data. The on-line NCDC serial publication database will eventually extend back to the 1890s. In addition, systems have been installed by NCDC that will allow for instant access via the Web for the newly digitized data. During 2001, NCDC's on-line data will exceed 1.5 terabytes.

The Coastal Services Center (CSC), part of the National Ocean Service (NOS), was created in response to a need for emphasis on coastal issues such as environmental pollution, land use, and coastal erosion. Many coastal issues involve the entire watershed area on the land side, coastal seas including estuaries and bays on the ocean side, and the wetlands between them. Understanding this land-water interface is becoming increasingly important area due to its sensitive ecology coupled with its vulnerability to hazards and rapid population growth. NOAA holds legacy data both from the land side and the ocean side, and seeks to increase the acquisition and management of new data to fill the gaps in these data sets.

The National Coastal Data Development Center (NCDDC), located at the Stennis Space Center, Mississippi, and operated by the National Oceanographic Data Center, was established in FY2000 to serve as a National Center that provides stewardship of the long-term coastal data record. NCDDC will use established and emerging technologies to access and integrate data stored in geographically distributed repositories in heterogeneous formats. Its products are intended to bring together the scientist and coastal manager. Working closely with NOAA's Coastal Services Center, NCDDC will supply the coastal manager with environmental data upon which to base management policies, plans, and decisions.

Programs addressing coastal issues are now among the highest priorities within NOAA. This emphasis is driving the need to develop accessible, integrated databases to support these programs. With the many recent advances in technology, significantly larger data sets will be acquired during the performance of hydrographic surveying to support nautical charting efforts. In addition, NOAA's large-volume, multibeam sonar data sets are providing much-needed information to the scientific community. NOAA's digital side scan sonar data sets are of interest to the Department of Defense, fisheries researchers, geologists, and many other users.

As the U.S. population becomes more concentrated in coastal regions, it becomes increasingly important to understand how this population increase is affecting—and is affected by—coastal ocean processes and conditions.

NOAA expects to see an increasing emphasis on the real-time acquisition and application of data during storms, tsunamis, and hazardous material spill events. In addition, NOAA's Physical Oceanography Real-Time System (PORTS) Program—a program providing ship masters and pilots with accurate real-time information required to avoid ship groundings and collisions—is rapidly expanding and is expected to grow by an order of magnitude over the next several years.

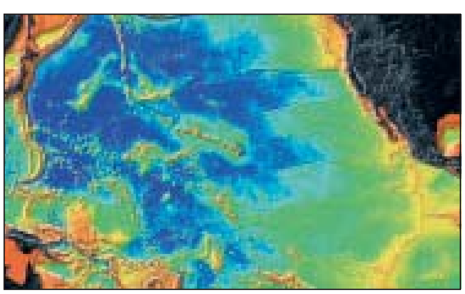
NOAA's National Marine Fisheries Service is implementing a plan to double the number of days spent at conducting living marine resource surveys. Within five years, the annual number of days at sea spent by fisheries research vessels is planned to reach approximately 20 ship years. Moreover, these surveys will rely increasingly on remote-sensing technologies, which generate large volumes of data.

Changes such as these are having a sizable impact on NOAA's data management activities, particularly on its archive and dissemination mission. These changes are also impacting the working relationships among the various Line Offices, which manage much of NOAA's data.

NOAA Data Produces New Ocean-Mapping Techniques; Saves Billions

Sea life is abundant where continents, ridges, and sea mounts deflect nutrient-rich water toward the surface. The general features of the ocean floor have been mapped, but detailed ocean mapping has only been accomplished for about 3 percent of the ocean because of the extreme expense involved—a complete survey is estimated to cost more than $1 billion. NOAA's Laboratory for Satellite Altimetry has developed a method for using NOAA satellite data to produce maps of the ocean floor in globally uniform detail.

These maps, although not as rich as those based upon detailed surveys, have brought significant savings to industry and government. Ocean-bottom maps have revealed previously unknown volcanic mountain ranges, which are now new fishing grounds. The maps also allow scientists to more accurately select areas to survey, resulting in greater efficiency and cost-savings in ship-time and research dollars.



New Technologies

During the 1990s, NOAA has taken the lead in adopting new technology to improve data management and data availability through activities that include the following:

- Adapting or developing innovative search and retrieval capabilities, and in broadly implementing on-line access to inventories and data
- Migrating data access from traditional off-line retrieval to on-line access
- Converting many aging paper and film records to digital form, thereby preserving the data and making those data more accessible for use by science, industry, and the general public.

Although geographically separate, NOAA's distributed data locations are now tied together by high-speed networks and the Internet, allowing on-line data exchange. For example:

- The NOAAServer System is used to meet customer on-line access requirements for a consistent "look and feel" for NOAA data sources.

- The NVDS provides a prototype for a seamless data access and delivery system that allows users to locate, access, browse, and order data and information products without regard for the physical location of the data.

Through the full implementation of these concepts, NOAA strives to provide customers with an integrated, consistent on-line experience. No longer should they need to contact separate data facilities, make separate data requests, and pay (if appropriate) each NOAA component separately for the data and information they obtain.

> *The grain market is controlled on a worldwide basis by local weather and climate conditions. Comprehensive records of timing and intensity of rain capture much of the essential relationship between local grain production and local weather...Most grain producers are important consumers of weather and climate information.*
>
> Darius W. Gaskins
> High Street Associates

Increasing Global Data Requirements

NOAA is required by public law to collect, archive, and distribute environmental data, and to use these data to describe the state of the environment. In recent years, it has become apparent that Earth's environment is a fully coupled, complementary system. Therefore, to adequately understand what is occurring in the U.S. and our adjacent waters, we must be aware of, and understand, the phenomena that are occurring globally.

Regional or national environmental data are no longer sufficient to fulfill NOAA's mission. In order to meet the needs of modern environmental forecasters, NOAA must acquire near-real-time global observations for local, regional, national, and international analysis and prediction.

As more and more nations understand that environmental phenomena are not purely local in cause and effect, there are greater demands for cooperation in collecting and sharing environmental data, on-line data exchange via the Internet, and the creation of global databases that can be searched via the Internet.

Increasing National and International Partnerships for Satellite and *In Situ* Data

Satellites have become key to the success of many components of NOAA's mission. However, no one nation—let alone one satellite operator—can provide complete global observations from space. Therefore, it has become critical for NOAA to negotiate and acquire space-based data from a growing number of national and international missions. NOAA must partner with other Federal government agencies and international sources of environmental data and information. However, the increasing numbers of these missions are placing ever-

NOAA GOES-8 satellite.

increasing demands on NOAA's data management structure—which must capture, describe, process, calibrate, access, retrieve, utilize, maintain, and exchange all of the various data sets. NOAA's satellite data responsibilities include the following:

- NOAA owns and operates the U.S. civil operational environmental satellites—Polar-Orbiting Operational Environmental Satellites (POES) and Geostationary Operational Environmental Satellites (GOES)—to collect near-real-time environmental observations.
- NOAA has assumed responsibility for the long-term archiving and user access of environmental satellite data from the Department of Defense's (DoD) Defense Meteorological Satellite Program (DMSP).
- NOAA is responsible for providing data access to non-NOAA and non-U.S. environmental satellite missions when these missions are able to support NOAA's operational requirements for global space-based observations.

Each of these sources reports data in its own data format, complicating the processes involved in archiving the data and making the data usable and available to the user community.

NOAA and NASA have a Memorandum of Understanding that NOAA will provide the long-term archival stewardship of the large-volume, remotely sensed atmospheric and oceanic data from the Earth Observing System (EOS). In addition, NASA is a partner in the pilot National Polar-Orbiting Operational Environmental Satellite System (NPOESS) Preparatory Program (NPP), for which NOAA will provide data archive and dissemination services.

NASA estimates that the EOS and NPP programs will generate a continuous data set of nearly 15 years of data, which will serve as a baseline for future global climate variability. The size of this data set is expected to be more than 5,000 terabytes—larger than all current geophysical satellite data archives combined. NASA and other Earth science research communities plan to utilize the NPOESS-series data, which NOAA will archive, for climate research.

All of these factors make NASA a natural partner for NOAA, with NOAA's mandate to archive and disseminate environmental data. Exchange of "lessons learned" from strategic planning, operations, and the needs of users will be key to solving the national archive problem. Shared utilization of existing assets will also be explored to maximize efficiencies and available expertise. Partnerships will also eliminate redundant processing and archiving activities, leading to savings on expenditures.

Polar-Orbiting Operational Environmental Satellites (POES)

NOAA currently has two POES satellites that provide global observations of imagery, temperature, and humidity at several levels in the atmosphere, ozone, and space environment. The observations are used in numerical weather prediction models for analysis and forecasting. These are also used to develop products related to the oceans' surface temperature, the health of vegetation around the world, the wetness of soils, snow cover, sea ice extent, detection of fires, and the size of the ozone hole.

By 2008–2010, the polar constellation will consist of three satellites—two U.S. and one European. With new instruments providing significantly more detail, each of these satellites will produce daily data volumes estimated to be more than 10 times those of a current POES satellite.

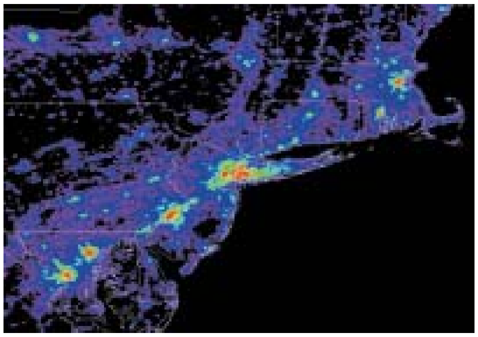
Several NOAA programs rely on the U.S. Geological Survey (USGS) for critical data needs. These data relationships are often formalized in official agreements. For example:

- NOAA and the USGS have had a Memorandum of Understanding since 1992 in an international effort with NASA, the European Space Agency (ESA), and others, to develop a Global Land 1 Kilometer Advanced Very High Resolution Radiometer (AVHRR) data set in cooperation with the International Geosphere-Biosphere Programme.

- In 1992, the Earth Resources Observation Systems (EROS) Data Center (EDC) was designated as the National Satellite Land Remote Sensing Data Archive. The holdings at the EDC include more than 120 terabytes of Landsat data and more than 12 terabytes of NOAA AVHRR data.

- The EDC has also been designated as one of the Distributed Active Archive Centers within NASA's Earth Observing System Data and Information System.

- The USGS collects and archives the national water data records, including surface and ground water quantity and quality. The USGS provides the bulk of river height and streamflow data used by the National Weather Service to produce flood warnings, river height, and water supply forecasts.



Radiance calibrated DMSP-OLS data for a portion of the northeastern United States. DMSP data are used nationally and internationally for a variety of applications such as analysis of population distribution, energy use patterns, and the impacts of urban sprawl on agricultural production. NOAA scientists have participated in fire monitoring projects using DMSP data for Brazil, Southeast Asia, and Mexico.

NOAA participates extensively in programs that focus on developing standards for metadata, exchange formats, and common search and display applications to reduce the effort required to make both archived data and incoming data useful. For example:

- NOAA is implementing elements of Executive Order 12906—Coordinating Geographic Data Acquisition and Access, the National Spatial Data Infrastructure (NSDI). The NSDI promotes the use of a consistent methodology for the sharing of geographic information between computer and Internet users. NOAA's efforts include: (1) developing data collection and content standards for several types of geospatial data; (2) implementing the NSDI metadata standard for all of NOAA's data holdings; (3) building databases for GIS systems; (4) implementing the Spatial Data Transfer Standard; and, (5) maintaining NOAA Data Directory Nodes on the NSDI Federal Geographic Data Committee (FGDC) Clearinghouse.

- NOAA is an active partner on the Committee on Earth Observation Satellites (CEOS), which includes membership from the countries that own and operate satellites and which cooperate in the exchange of Earth observations. CEOS is the major coordinating body for the satellite operator community. It also coordinates data management and services with a primary objective of encouraging complementary approaches to limit redundancies.

- NOAA is seeking opportunities to collaborate with Space agencies in other nations by examining the possibility of placing instruments on each others' platforms and by exchange of Earth observation data, from polar-orbiting satellites such as EUMETSAT's METOP and Japan's ADEOS and GCOM series of satellites.

- NOAA is engaged in the transition of the Global Observation Information Network (GOIN) from a bilateral initiative under the U.S.-Japan Common Agenda to a multilateral framework under CEOS. GOIN explores the use of high-performance research and education networks to exchange and utilize environmental and other scientific data and information among agencies and institutions around the world.

Operation of World Data Centers

NOAA maintains global databases that provide U.S. science and industry with convenient access to a wealth of foreign data to which they might not otherwise have access.

As part of international data exchange agreements with 160 countries and under the auspices of the International Council of Scientific Unions, NOAA operates seven of the 13 World Data Centers (WDC) located in the United States: Meteorology, Oceanography, Glaciology, Marine Geology and Geophysics, Paleoclimatology, Solid Earth Geophysics, and Solar Terrestrial Physics. As World Data Centers, NOAA maintains global databases that provide U.S. science and industry with convenient access to a wealth of foreign data to which they might not otherwise have access. Even in cases where the U.S. does not have close diplomatic relations with a country, the WDC System often provides a means of sharing data and information between various scientific organizations within those countries and the U.S.

Many of the changes affecting the WDC System in recent years are similar to those changes affecting NOAA data facilities. For example:

- Resources have shifted from the archive of paper and film to reformatting or documenting digital data, and to developing computer systems to receive and distribute digital data.

- Cost recovery policies of funding agencies are resulting in more and more data that are under copyright, and therefore not available to the WDC System. This necessitates an increasing efforts toward locating and maintaining paths for obtaining data that can be distributed without copyright restriction.

- On-line file transfer is becoming the preferable means for both contributing data and related information to the WDC, and delivering data and information requests to the user community.
- The WDCs are being pressured to develop and maintain integrated, global databases from both historical and current data sets, as well as information for the user community.
- In some instances, the data are no longer specifically held at one WDC, but rather the WDC is linked via the Internet to digital systems in many WDCs and other National centers.

Growing Archives

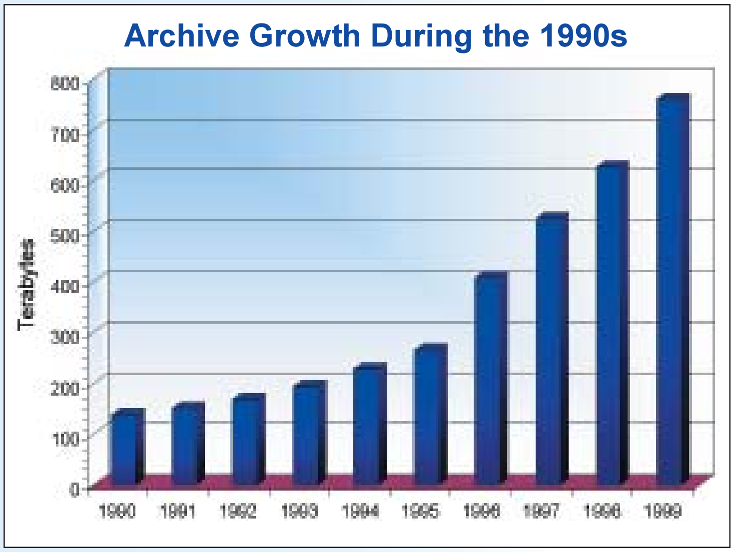
No previous decade has seen the magnitude of changes in the volume of data coming into NOAA for archive as those experienced in the 1990s. In the 12-year period between 1978 and 1990, the NOAA digital archive increased by 130 terabytes. In the next 5 years, the archive grew by another 130 terabytes, followed by a 1-year growth of the same amount in 1996. The NOAA archives have experienced a steady growth in digital data, increasing at an average rate of nearly 20 percent per year—more than quadrupling in volume by 1999.

- The geophysical digital archive essentially doubled each year between 1990 and 1994. It more than doubled again between 1995 and 1998—due to the influx of Defense Meteorological Satellite Program data for the new Space weather archives.
- The oceanographic digital archive increased in volume 27-fold between 1990 and 1999, largely due to significant increases in the amount of satellite sea-surface temperature data provided for archive.
- The weather and climate digital archive doubled in volume, between 1995 and 1997, due to dramatic increases in NEXRAD Doppler radar (the latest generation of weather radar) data and the continuing influx of satellite data.

The staggering growth of the NOAA digital archive has resulted primarily from the proliferation and use of remote sensing systems—typically radar and satellites—to provide environmental observations.

Data Exchange with Foreign Country Nets Tangible Benefit for U.S.

Data were provided to the Italian Istituto Idrografico della Marina at a cost to NOAA of $130 for labor and materials. In exchange, NOAA received data from 1,415 ocean stations. Figures provided by the National Science Foundation indicate that oceanographic research ship operations cost roughly $10,000 per day. During normal operations, about 2.3 stations per day can be occupied. Therefore, this one small data set would have cost the U.S. roughly $5.6 million dollars to collect.

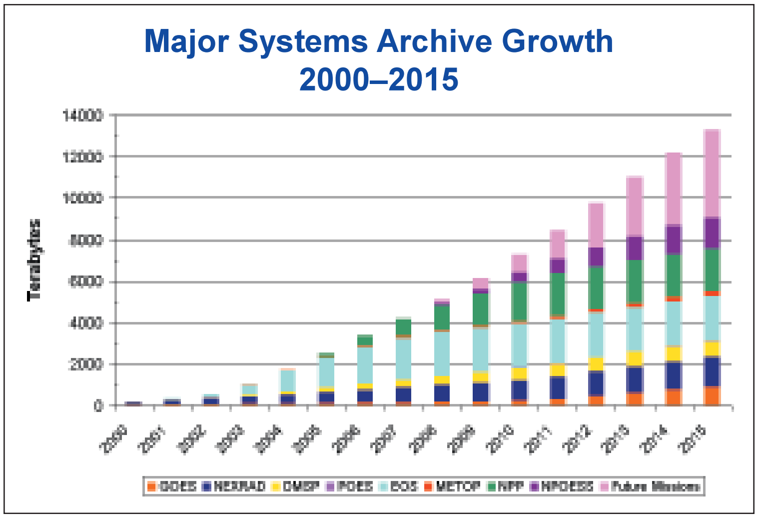


Throughout the decade of the 1990s, the NOAA archives experienced a steady growth in digital data, increasing at an average rate of nearly 20 percent per year—more than quadrupling in volume by 1999.

Traditional (*in situ*) observations involve individual pieces of data collected from individual locations on the ground, in the ocean, or in the air. A rain gauge at a commercial airport, a ship report of water temperature or wave height, geomagnetic observatories, continuously operating Global Positioning System (GPS) stations, or flight reconnaissance report of the wind speed and barometric pressure in a hurricane, are examples of *in situ* environmental observations. These types of data sets can be stored by recording the data manually or by electronic means, and tend not to produce massively large data files.

In contrast, remote sensing systems observe the environment from a distance over wide swaths of Earth and its atmosphere rather than from each individual location. These swaths can encompass millions of square miles in area, and they can provide data readings for each square mile within the area observed, providing data readings for millions of locations. These large array data sets necessitate very large archive storage capacities, such as the following:

- The 164 operational NEXRAD WSR-88D (Next Generation Radar, the Weather Surveillance Radar-88 Doppler Program) sites are now providing approximately 100 terabytes of data each year—an annual amount equal to about 10 percent of the entire NOAA digital archive at the end of 1999.
- Polar-Orbiting Operational Environmental Satellites (POES) data are increasing by almost 2 terabytes per year. POES data provide global weather information. These satellites also provide data for assessing and tracking atmospheric ozone distribution, and the effects of solar activity on communications and electric power systems.
- The Geostationary Operational Environmental Satellites (GOES) archive is growing at a rate of 17 terabytes per year. Data from these satellites are used to compute hourly temperature and humidity values for the U.S. and adjacent ocean areas. Wind speeds and directions are inferred from cloud movement captured in GOES imagery.
- Scientific instruments on DMSP satellites and space environment monitors on GOES and POES satellites produce nine gigabytes per day or three terabytes per year. These data rate numbers will increase by a factor of four in two years and 12 in another four years. These data are useful for monitoring surface light sources.



Archives are expected to grow at approximately 1.2 petabytes per year. NOAA's total archive, entering the 21st Century, was less than 1.2 petabytes.

Even as the current reporting systems continue to provide data for archive, new satellite systems such as EOS, NPP, and NPOESS will be going into operation. These systems will provide massive amounts of new data, which will present formidable challenges for NOAA in archiving and providing data access.

Although the volume of data will be less from fisheries and hydrographic missions, these data sets are growing rapidly and are extremely diverse in their content and structure. While the data sets will remain predominantly *in-situ*, a growing component of remote sensing data will also begin accumulating, and the amount of data to be managed will grow at an increasing rate.

There are massive amounts of new coastal environmental data that need to be managed by NOAA due to improvements in the use of remote sensing systems. Until recently, remote sensing systems for coastal areas were confined mainly to estimating sea-surface temperatures, which were made available in near-real-time through NOAA's CoastWatch Program. Within the next several years, however, both the number of observation points and the resolution of these observations will be substantially increased.

In addition, new data—such as ocean color and sea-surface salinity—will be included. NOAA is also increasing its focus on high-impact issues at the land/sea interface, such as harmful algal blooms, fish abundance and the preservation of coral reefs.

NOAA's digital archives are expected to grow to almost 20 times their 1999 volume over the next 15 years. Once EOS becomes fully operational, that system alone is expected to provide more than 500 terabytes of new data to be archived *each* year. By the Year 2003, NOAA will be expected to ingest and archive as much data in one year as was contained in the total digital archive in 1999.

Massive amounts of new data will present formidable challenges for NOAA in archiving and providing data access.

Ensuring Continuity of the Nation's Environmental Record

A critical function performed by NOAA is to ensure continuity of the Nation's environmental record across heterogeneous observing systems, through changes in sensor technologies over time, through re-calibration of instruments, through changes in instrument site, etc. With the modernization of observing systems and the increasing deployment of new observing systems, the performance of this function is becoming even more critical to solutions to today's climate change issues.

For example, the National Weather Service has historically relied on manual observations made by personnel at Weather Service Offices. With the deployment of the Automated Surface Observing System (ASOS) during this decade, observations of air temperature, wind velocity, visibility, etc., are taken automatically and with much greater spatial resolution and temporal frequency than before. In many locales, this change in the observing system coupled with a change in the site of the observations has led to apparent "changes" in the local climatological record. Through the aggregation of observations and metadata from many different sources, NOAA scientists can separate "real" environmental changes from such artifacts of the observing technology.

Automated Surface Observing System (ASOS).

The capability to create an accurate record has enormous consequences from an economic standpoint. For example, capacity planning by local power utilities is heavily dependent upon average temperature conditions derived from the data in NOAA's archives. A decision on building a new power plant at the potential cost of hundreds of millions of dollars may hinge on a fraction of a degree difference in the climatological average. Therefore, a smooth and continuous record reflecting true conditions—and not artifacts of the observation process—is critical in making such decisions.

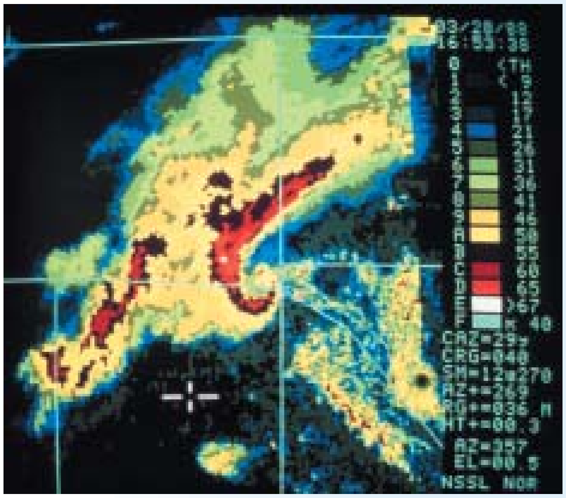
Increasing Complexity of Data Management

Data do not become archived and available for dissemination to users simply because they are delivered to NOAA. Data ingest—the process by which data are received, processed and prepared for archive—may be relatively straightforward or complex and time-consuming, depending upon the characteristics of the data and the delivery medium used. Today, the vast majority of environmental data are provided in digital form. However, this does not mean that these data are readily usable.

Agencies and researchers who develop observing systems and instruments are primarily concerned with obtaining observation data for *their* mission—they may have little interest in facilitating the archive and re-use of their data because that is not their mission. Therefore, the data formats that they design tend to be optimized for the program's use, and the archive delivery system and media chosen by the program are the most cost-effective from the standpoint of the *program*. This usually results in inflated archive and dissemination costs over the full life cycle of the data. For example:

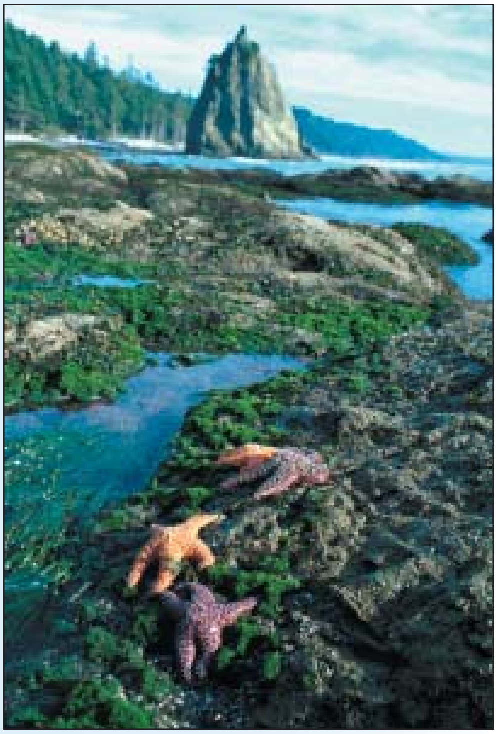
- NEXRAD Doppler radar data are received in digital form on 8mm tapes. Each tape contains approximately 2 to 3 days of data for one station, and data are reported for approximately 164 stations. Due to the extremely slow tape speed, reading or copying just one tape takes at least 4 hours. The labor-intensive nature of this process in effect limits the distribution of this data set (to other users) to small amounts of data.

- NOAA has more than 38,000 U-matic tapes containing the GOES data collected for more than 20 years. Use of this antiquated, 25-year-old tape technology for data storage means that the data are practically inaccessible for today's researchers.



The distinctive red hook echo on this NEXRAD Doppler Radar image indicates tornadic activity.

The actual data provided for archive are changing, requiring more human resources to make the data useful and accessible to users. The content of the data to be archived is becoming more complex, with new types of measurements being introduced. Increased data



Olympic Coast National Marine Sanctuary. (Photo credit: Nancy Sefton.)

complexity brings with it a requirement for more human resources—and in some cases, new or different skills—to manage the data than was previously required.

Another area of major emphasis is the Nation's marine sanctuaries. Both the volume and the complexity of the data being gathered from these areas will increase, and management of these new data sets will require additional investments that are presently unidentified. In addition, NOAA is involved in characterizing and mapping freshwater, estuarine, and marine species communities and their habitats on local, regional, and national scales to meet mandated responsibilities for Essential Fish Habitat regulations. This effort is growing rapidly and requires significant collection and management of spatial data in GIS-compatible formats.

Access to some of NOAA's data sets must be restricted or they may only be distributed in aggregate form due to confidentiality or legal requirements. For example, the locations of cultural resources such as shipwrecks within National Marine Sanctuaries, catch information voluntarily provided by commercial fishers and the fishing industry, and information related to ongoing legal cases cannot be freely disseminated. In cases such as these, special data security measures must be employed to ensure adherence to the appropriate confidentiality or proprietary restrictions for each data set.

New emphasis is being placed on the standardization of data set structure and data accessibility so that relevant data sets from diverse sources can be readily identified, accessed, and analyzed together to solve pressing environmental issues. The use of geographic information systems (GIS) will increase dramatically as NOAA strives toward a holistic approach in analyzing spatial data from an ecosystem perspective.

Because data sets are becoming more complex, there is a growing amount of information that must accompany the new data to make them useful. Both the metadata (i.e., information about the data) and the data themselves must be managed in a way that ensures not only their preservation and availability, but also ensures that they are related to each other and that the relationship is helpful to users. This activity requires ongoing human resources to maintain the metadata and to manage the data for active data sets.

Rescuing Data and Information

NOAA conducts many environmental data rescue activities to preserve historical data before they are lost or have become unrecoverable, thereby preserving these data to assist in finding solutions to today's—and tomorrow's—environmental problems.

These activities involve rescuing data from around the world, and converting data from deteriorating media to modern media.

The goals of NOAA's former Environmental Data Rescue Program were to preserve the meteorological, climatological, geophysical, and oceanographic data stored throughout NOAA, and to make this information more accessible to researchers and the general public. This successful (yet short lived) program rescued millions of environmental records from being irretrievably lost due to deteriorating media. However, once these data were converted to more stable storage media, they needed to be ingested, validated, and processed as are any new data coming into the archive. Therefore, these rescued data competed for the same resources that are required for the ingest of new data, causing even more pressure on overburdened resources.

Less than 50 percent of the hydrographic surveys are available in digital format for use in GIS applications. Paper chart and source data archives need to be scanned, geographically rectified, cataloged, supplemented by metadata, and made accessible to users to facilitate charting efforts, environmental research, and environmental resource management. Although much historical tide, water-level gauge height, and water current data have been the subject of previous data rescue efforts, much data remains in need of rescue.

Rescue efforts are underway to convert more than 10,000 historical shoreline maps into digital format. However, the rescue of the 600,000 aerial photographic images of coastal areas, airports, and some sites of natural disasters has not yet been addressed. Much historical data of importance to understanding fisheries stocks—such as commercial vessel logbooks and old data sheets from fisheries research vessels—are in danger of being lost due to deteriorating physical media.

NOAA also conducts environmental data rescue activities to preserve historical data from countries around the world. At the end of the Cold War, millions of extremely valuable, historical climate records became available to scientists around the world. NOAA utilizes international organizations, such as the United Nations, and bilateral agreements between nations to collect and preserve data from many of these countries. It is imperative that these data be rescued before they are lost forever due to physical deterioration, and to make the data available for use.

NOAA is conducting the Global Ocean Data Archaeology and Rescue (GODAR) Program under the aegis of the United Nations Intergovernmental Oceanographic Data Exchange. The purpose of GODAR is to ensure that the historical deep-ocean and coastal data

New Use for Archived Satellite Data: Exploration of the Polar Basins

The sea ice that stopped Captain James Cook from exploring further south in the 1770s continues to hamper conventional exploration of polar ocean basins. However, NOAA satellite data that were originally collected for another purpose are now being reprocessed and used to map the detailed gravity, sea depth, and structure of these polar basins. Ultimately, these same altimeter data may be used to detect possible long-term thickness changes in sea ice in response to climate change.

that are so important in developing baselines are not lost. This multi-nation activity has resulted in the largest ocean profile database in the world, with data still remaining to be rescued.

Russia has a wealth of historical environmental records, mostly in manuscript form, going back 100 years. With as many oceanographic stations as the rest of the world combined, these data can fill many of the gaps in NOAA data sets, enabling improved understanding and prediction of both oceanic and atmospheric change. NOAA has begun to digitize some of these records to preserve the data and to make the data available to researchers and industry. However, other valuable data sets have yet to be rescued from many countries—*as well as from the U.S.*

Where Are the Fish?

It costs money to fish the ocean. Large ships with good equipment are needed to be able to sail the ocean, find the fish, and haul them in. But ocean data must also be used to know where to look for fish. Commercial fishermen save millions of dollars each year by using ocean temperature data available from NOAA to locate the catch.

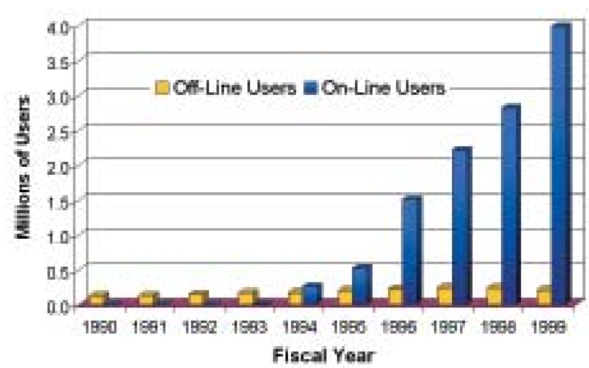
The increasing volumes of incoming data are outgrowing NOAA's storage capabilities and technologies, and they are affecting NOAA's data-rescue capabilities as well. With new media and mass storage technologies being introduced at an ever-accelerating pace, NOAA's ingest and storage capabilities—as well as its media migration capabilities—are falling farther and farther behind modern technology. There is an urgent need to migrate much of our archived data to new media before these data are unrecoverable.

Increasing Numbers of Users

Use of the Internet and the availability of on-line data and information are primarily responsible for the tremendous increase in direct users of NOAA data. However, direct user requests represent only a small fraction of the total number of individuals who rely on NOAA data each year.

Customers for NOAA data prefer to order data and information on the Web.

Requests for Data and Information



New User Groups

New user groups are evolving. One such group—weather derivatives—is potentially a $75 billion international industry. This vital new industry includes financial management companies, insurance and reinsurance companies, energy companies, and other industries whose costs are affected by weather and climate extremes in environment. This new risk management tool uses financial instruments rather than traditional insurance policies to manage the risk of losses due to extremes in weather and climate.

Tornado near Anadarko, Oklahoma, May 3, 1999.

Both the development and settlement of these financial instruments are based upon access to accurate, objective, and very timely national and international data from a source that is considered to be accurate and unbiased. NOAA operates the networks that collect these data, and NOAA archives these data; therefore, NOAA is the only source for the data that are the basis for this new industry.

Evaluation of long-range historical environmental data is the basis for the terms of these financial instruments that are traded on the financial markets. The evaluation of the actual environmental conditions during the agreement's period of performance is the basis for the settlement. This requires accurate and complete environmental data currently provided by NOAA for any specified geographic area on Earth, for the dates specified—all within days of the actual environmental observations being recorded. This is placing a very high demand on NOAA to deliver quality data, with almost immediate turnaround, and on a continual basis.

Currently, these requirements cannot be met because NOAA observations required for the resolution of the trades are collected by *volunteers* and recorded on paper forms that typically take 90 days for NOAA to receive and verify.

Weather Disasters and Economic Loss

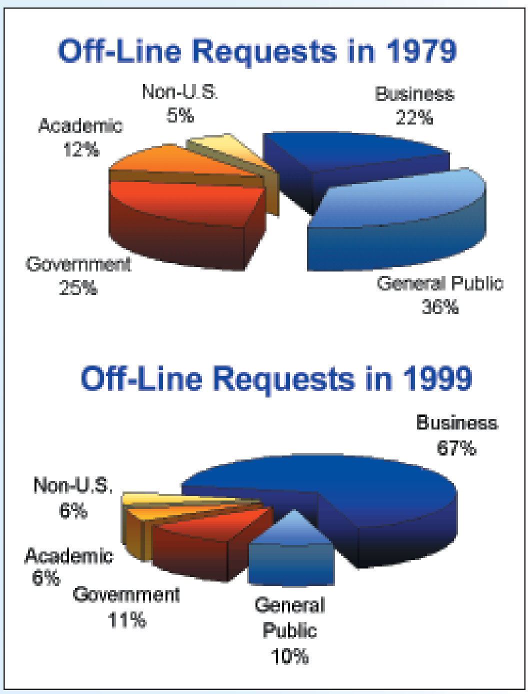
1998 was one of the costliest hazard years for our Nation. The U.S. sustained seven disasters, each costing more than $1 billion. The Southern drought and heat wave alone resulted in almost $9 billion in agriculture and ranching damage. An epidemic of drought and wildfires have persisted in recent years, resulting in economic losses on many levels.

Increasing Numbers of Requests

Over the past 20 years, there has been a tremendous increase in the number of business users who request data and information from NOAA. The initiation of an on-line store service for ordering data products and the acceptance of credit cards as a payment option have also contributed to the changing user profile by simplifying the process of obtaining data and information from NOAA. A small sample of the business uses of NOAA environmental data include the following:

- **Attorneys and Consulting Meteorologists:** Evidence and expert testimony in court cases.
- **Insurance Industry:** Rate determination and claims settlement.
- **Engineering, Marine, Architectural, and Construction Industries:** Design and construction guidelines, site selection, environmental impacts, construction deadline penalties and extensions.
- **Utilities:** Projections of demand, computation of rate adjustments, air pollution studies.
- **Agribusiness:** Determine optimal geographic locations by crop type, plan the application of herbicides and pesticides, study effects of climate variation on crop yield.
- **Navigation:** Nautical charts, the *Coast Pilot,* and other marine products are used by commercial and recreational users of the Nation's waterways.
- **Fishing Industry:** Locating prime locations for fishing (through ocean temperature data).
- **Financial Services:** Utilize temperature, precipitation, drought, and flood data to reduce risk and minimize losses.

Business use of NOAA data, as reflected in off-line data requests, has tripled in the past two decades.



In general, user requests increased throughout the 1990s. Although off-line data requests doubled, the truly exponential growth has been in the number of on-line users. However, NOAA receives some user requests that must be denied because they would be too difficult and time-consuming to service.

While on-line requests have increased, it is important to realize that of the 760-terabyte NOAA digital data archive, only 3 terabytes are available on-line. This is due to constraints in the migration of digital data to modern media and usable formats. In addition, much of NOAA's archive has not yet been converted to digital form.

On-Line Ordering and Data Delivery

Historically, NOAA received requests for data via telephone, fax, or mail. The requested data were usually supplied to the user on paper (e.g., printouts, copies, maps, publications, etc.), or—in the case of digital data—magnetic tapes were supplied. Over the past few years, compact discs have been used to provide many data sets to users.

With the growth of the Internet, users are expecting on-line ordering and on-line search and browse capabilities, with electronic file transfers for data delivery. Users are no longer content to wait days or weeks for their data or information requests to arrive at NOAA, be subsequently processed, and mailed back to the user.

In response to user demand, NOAA has made a number of its most requested data sets available to users through the Internet to improve customer service and to reduce the costs for servicing user requests. By the end of 1999, many customers were both utilizing new NOAA on-line ordering systems and receiving environmental data from on-line NOAA sources.

For example, NOAA developed an Internet-based browse tool that enables users to view a time-series graph of various daily and monthly meteorological parameters. Due to its popularity, NOAA received a large number of requests for the data from which these graphs are produced. To meet user demand and to alleviate the staff time required to fill these data requests, the tool was updated to allow the user to download the data from the Internet. On average, almost 25,000 users per month accessed this tool during 1999.

An additional challenge to data accessibility is to allow users ready access to NOAA data—regardless of its actual geographic or organizational location. For example, a user investigating the relationship among weather, oceanography, and the abundance of fish stocks should be able to locate and download the necessary data without being concerned that the various data sources may actually reside in several different locations.

On-line data delivery typically provides benefits to both the user and NOAA by providing easier and faster turnaround for the user, while requiring less NOAA staff time to service the more common, high-frequency requests. On-line data delivery also eliminates media and shipping costs. Benefits increase when large numbers of these requests can be filled without direct staff intervention, by providing the user with on-line browse and search capabilities via the Internet. Annual downloads by users have increased from 4.7 terabytes (in FY1998) to 18 terabytes (in FY2000).

Increasing Requests for Information

As on-line access to NOAA's data expands, the user's average level of technical sophistication and scientific expertise is changing. On-line users are searching for information and answers to specific questions rather than for access to data.

The needs of the business community and industry are becoming more complex, seeking interrelated data and supporting information

User Request Could Not Be Filled Due to Aging Archive Technology

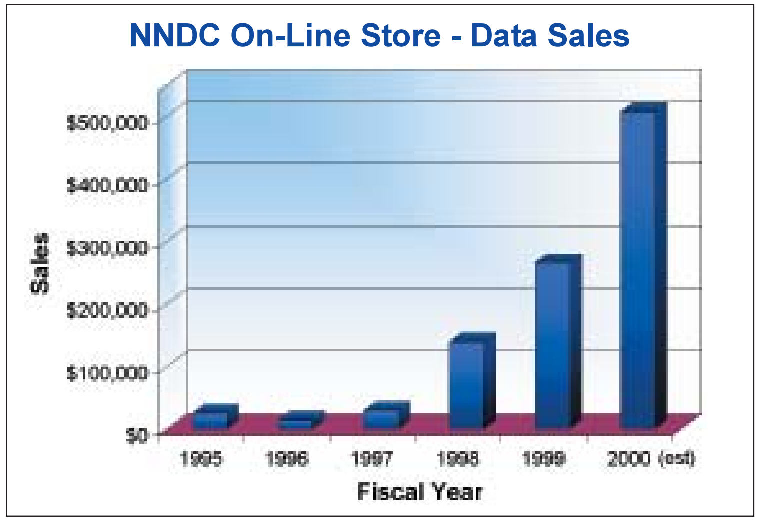
In 1998, the U.S. Weather Research Program (USWRP) and university researchers requested several years of NEXRAD Doppler radar data for multiple radar sites. These data were to be used for a study to improve the understanding of the causes and predictability of severe weather in the U.S.

NOAA could not provide the data requested because it did not have sufficient resources (both human and equipment) to produce tapes containing the requested data in the required 60-day time frame.

The NEXRAD archive was more than 200 TB in volume at the time, and the data are archived on the same 8mm tape technology on which they are received from the NWS. Since this unwieldy tape technology requires about 4 hours to read 1 tape, this request would have required more than 8,000 hours of computer time to extract the data requested—utilizing 24-hour operations of a computer and 12 tape drives for 60 days to produce an estimated 1,643 tapes.

Had NOAA attempted to fill this request, it would have been impossible to keep up with the routine operations of archiving the incoming NEXRAD data because many of the same resources required for the ingest process would have been in use to fill the data request.

In the past three years the quantity of data downloaded has doubled annually.



and documentation, rather than just seeking one particular type of data. For example, one researcher asked, "How do I get the hurricane out of the data?" Evidence of a hurricane can be found in many different data sets—in wind measurements from land and sea; in satellite data and images; and in ocean wave and water-level measurements. A researcher who wants to study hurricanes must have an understanding of the different types of data involved, in order to accurately request the different data sets that he or she will need. The task of producing integrated "event" data sets is both time-consuming and expensive, and NOAA is increasingly being asked to provide data in a form such as this—transforming data into information that is more meaningful to the user.

Mitigating the Effects of Environmental Change

There is clear evidence that Earth's climate varies, which includes severe and extreme environmental events. Changes in the frequency, geographic coverage, and intensity of environmental events—droughts, floods, hurricanes, extreme snowstorms, etc.—cause changes that stress the infrastructure and living resources of the U.S., with significant financial impacts. As world population increases, the possibility is substantially increased that a significant environmental event will affect some center of human activity.

With access to adequate data, scientists can predict the occurrence of many climatic anomalies and make skillful predictions as to their consequences. NOAA provides data to assist in the prediction of these events, and promotes environmental hazards mitigation efforts.

- The 1997–1998 El Niño event—one of the most intense of the 20th Century—was successfully predicted by NOAA's National Center for Environmental Prediction (NCEP) using

NOAA data. Data from two key data sources—ocean subsurface temperature measurements collected by ships and buoys from many nations, and satellite altimetry data—were integrated with other NOAA data in the NCEP model that predicted that anomaly.

- At the request of the reinsurance industry, NOAA developed a "natural hazards" Web page to educate the public on various meteorological and geologic natural hazards and how to mitigate their effects. NOAA also developed the on-line Natural Hazards Resource Directory. This Web page provides contact information for various resources and shows how these resources can helpful.

Societal concerns about Earth's changing environment and the desire to mitigate the effects of natural hazards have influenced the scientific community's demands on NOAA for data and information. As interest in these environmental issues increases, there is a corresponding increase in pressure on NOAA to provide data and information to support these studies.

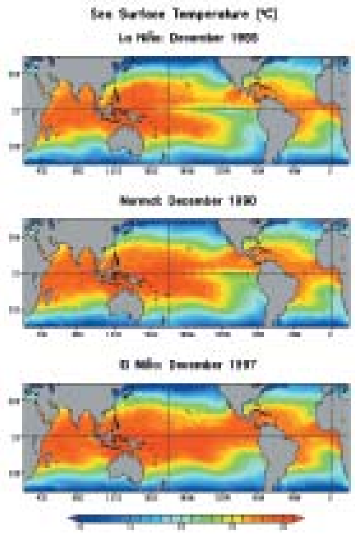
NOAA is being pushed to develop an increasing number of new, on-line data presentations to keep pace with user demand. The demand for specialized data services such as event-specific Internet Web pages for both natural (e.g., hurricanes) and man-made (e.g., oil spills) disasters will increase in the future.

The future of NOAA's environmental prediction capabilities will include an integrated suite of forecast products with lead-times ranging from minutes to decades. This concept is predicated on NOAA's ability to portray the probability that an event will take place far enough in advance so that the public and responsible decision-makers can save lives, mitigate the effects of potential losses, and promote the economic well-being of the Nation. Long-term data records provide statistical and probabilistic information.

For example, the Advanced Hydrologic Prediction Services (AHPS) is based on computer models of rainfall, water runoff, and flow in river channels—and the historical record of these parameters for given forecast points. The AHPS can extend the current 1-, 2-, and 3-day river stage forecast lead-times to weeks and months in advance. The probability information, or uncertainty, associated with these long lead-time forecasts is essential for risk-based decision-making, and will allow the preparation of maps that identify areas of potential flood inundation. Additionally, this type of advanced river forecasting will provide an estimated $400 million per year in additional economic benefits due to improved water-resource management practices. However, these forecasts are not possible

El Niño Costs Millions

El Niño-related climate variations often have widespread and devastating impacts. The frequency, severity, and paths of storms in the Pacific, the occurrence of short-term droughts, floods, and severe weather in certain regions of the U.S., and the viability of Pacific coast fisheries are all affected by El Niño and La Niña events.



The 1986–1987 El Niño event provides a well-documented example. This event was followed by a severe drought in the Northern Hemisphere that significantly reduced crop production and disrupted transportation on internal waterways. Losses for non-farm, down stream businesses amounted to $10 to $15 billion.

South Dakota farm blown out during the Dust Bowl, 1938.

Societal concerns about Earth's changing environment and the desire to mitigate the effects of natural hazards have influenced the scientific community's demands on NOAA for data and information.

Floodwaters, Saint Genevieve, Missouri, 1993. (Photo credit: FEMA.)

without the long-term weather and river records on which the probabilities are based.

NOAA National Marine Fisheries Service (NMFS) is responsible for managing the Nation's fisheries, much of which is based on developing predictions for fish populations in a process called stock assessment. Stock assessments are used for determining catch targets, to maximize economic return, and prevent over-fishing. Assessments are based on demographic models that use long-term data sets on catch, the size and age structure of the population, and mortality as inputs. However, the status with regard to over-fishing is unknown for 544 out of the 844 stocks managed by NMFS. The National Research Council has recently recommended that stock assessments be expanded from single-species to ecosystem models. Therefore, the requirements for long-term biological and oceanographic data sets for fisheries management are expanding greatly.

NOAA's environmental data and information support the development of the national economy in a global environment. Environmental data and information are increasingly important to our society—for the individual who makes day-to-day business, environmental, and social decisions; to policy makers who make global decisions that will affect all of us for years to come. NOAA's data and information will only continue to increase in importance for environmental forecasting, protection of life and property, environmental monitoring, national information policy formulation, and national security. Providing highly credible, century-scale, climatic data perspectives in a near-real-time mode has become an absolute requirement for NOAA.

Clockwise: Coral colony, sea star, sea urchin, rosy rockfish—colorful occupants of National Marine Sanctuaries administered by NOAA. (Photo credits: Cordell Bank Expeditions, and Laura Francis.)

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.