

Appendix C. Responding to the Challenge: Progress in Data Management

C.1 NOAA-Wide Activities

NOAA is responsible for the management and stewardship of a rapidly increasing amount of data and information. These resources impact all of NOAA's activities and include holdings of climatological, geophysical, oceanographic, marine fisheries, hydrographic, and cartographic records. Much of these data are held and archived in NOAA's National Data Centers: the National Climatic Data Center (NCDC), the National Geophysical Data Center (NGDC), and the National Oceanographic Data Center (NODC). Some of the data is stored by NOAA's Line Offices or by the office or scientist who originally collected it. Data are stored on a variety of paper, film, or digital media.

Throughout NOAA, there are a variety of programs that support data and information management activities.

NOAA Information Technology Architecture

The Department of Commerce Chief Information Office Council has issued a requirement that all bureaus submit for review Information Technology (IT) Architecture plans. The origin of the IT Architecture requirement lies in the Information Technology Management Reform Act of 1996, which requires that the Chief Information Officer of an executive agency be responsible for “developing, maintaining, and facilitation the implementation of a sound and integrated information technology architecture for the executive agency.” Between the Act and other Office of Management and Budget regulations, agencies are also required to institute and use a capital planning and investment analysis process to develop both a Strategic IT Plan and an Operational IT Plan.

NOAA is developing an enterprise-wide IT Architecture that defines the building blocks that make up the overall information system for the agency. The IT Architecture encompasses four views of the organization:

- business process—the actual business efforts being supported.
- data and information—the resources needed to support the business.
- applications—the programs used to process the data and information.
- technology infrastructure—the hardware and services supporting the business.

Administrative Systems				
Archiving and Access				
High Performance Computing				
IT Security				
Messaging/Directory Services				
Networking/Shared Telecommunications				
Office of Marine and Aviation Operations				
NESDIS	NMFS	NOS	NWS	OAR

NOAA IT architecture domains (horizontal bars) and segments (vertical bars).

NOAA envisions that its personnel as well as others will be able to access and use appropriate information wherever and whenever the information is needed for the benefit of NOAA, its partners, and its customers.

NOAA is such a large and diverse organization that it would be impractical to have one uniform architecture for the entire organization. The chart (above) shows the NOAA IT architecture structure. It is based upon the following approach:

- NOAA’s basic business units are its Line Offices, and results can be obtained more effectively by basing the Architecture around the Line Offices. These parts of the Architecture have been named “segments.”
- Certain areas of technology and types of systems have impact across all of NOAA and need to be dealt with at the NOAA-level. These parts have been named “domains.”
- Data and other interfaces between specific Line Offices and programs are important, but initially, they will be dealt with in the context of each segment or domain.

The horizontal rows represent the NOAA-wide domains where one architecture is needed for the entire organization. The vertical columns (with organizational acronyms) are the segments being developed by NOAA’s Line Offices.

NOAA has an enormous, around-the-clock data processing, dissemination, and storage infrastructure which is needed to support widely-distributed operations with numerous business partners and the public. NOAA is dependent upon a large IT infrastructure and recognizes that its IT must be planned in a cohesive and enterprise-wide manner. The goal is to achieve more efficiencies and higher levels of integration and interoperability internally with its Line Offices and with partners.

The strategic direction that the NOAA Archive and Access IT Architecture will take is governed by the following key elements:

- Enterprise-wide business objectives;
- A contextual framework or IT vision for transforming the Agency's IT infrastructure;
- IT objectives defining the tangible benefits of successfully fulfilling the vision;
- IT principles that assist NOAA decision-makers in formulating key IT development and investment decisions; and,
- A team of individuals representing each of the Line Offices to bring requirements and expertise to define, refine, build, and implement the crosscut architecture.

Principles support NOAA objectives and can be thought of as the rules or guidelines that will be implemented to realize our objectives. The "letters" following each item below represent the Application (A), Business (B), Data and Information (D), Metadata (M), and Technology (T) aspects of each principle:

- NOAA will develop and support functions for archive and access of its environmental data and information holdings. (A1)
- NOAA will provide intuitive state-of-the art access to environmental data and information. (B1)
- NOAA will provide technology training so that personnel remain current. (B2)
- Within NOAA, Departmental, and/or Federal policy and guidelines, NOAA's Line Offices will determine where, when, and/or how its environmental data and information are archived and how it is made available (corporate asset). (B3)
- Data and information are corporate assets. (D1)

- NOAA will maintain an environmental data and information metadata catalog. (D2)
- NOAA will support its mission for stewardship of the Nation's environmental data and information. (M1)
- NOAA will position itself to make appropriate use of emerging technology (user capability, internal, NOAA use). (M2)
- NOAA will use commercial, off-the-shelf software in preference to home grown solutions. (M3)
- NOAA will protect and preserve its environmental data and information holdings (deep archive, working archive, and disaster backup). (T1)

Environmental Services Data and Information Management (ESDIM) Program

The Environmental Services Data and Information Management (ESDIM) Program was created in response to concerns from the NOAA Under Secretary for Oceans and Atmosphere about data management in the organization. Data management in NOAA deals with acquiring, ensuring quality control, preserving, storing, and providing user access to its data holdings.

A study by NOAA's Advisory Panel on Climate and Global Change was commissioned in 1989 to review NOAA's data management and to make recommendations for more effective information management. The panel's February 1990 report, along with a November 1990 GAO report, *Environmental Data: Major Effort Needed to Improve NOAA's Data Management and Archiving*, were the catalysts that started the ESDIM Program in early 1991.

ESDIM sponsors projects within NOAA Line Offices in the areas of access, rescue, continuity, and innovation. During its early years, the ESDIM Program concentrated on data rescue. Data rescue is the saving or salvaging of data on paper, film, or digital media, that would otherwise be lost, converting it to a stable, usable media, and then archiving and/or making it available for access.

NOAA's data include paper records from the beginning of the Nation's history through the present time. Paper deteriorates over time depending on the physical properties of the paper and the conditions under which they were stored. Microfilm and microfiche, used to rescue deteriorating paper records, are themselves susceptible to deterioration and must be rescued. Digital media are

used to rescue paper and film records and to store newly acquired data. Digital media are also susceptible to deterioration. Data stored on older, less dense magnetic tape are being rescued to more dense, smaller sized tape cartridges or to optical media. In time, data stored on today's best media will have to be rescued.

The amount of new data to be archived is rapidly increasing as new data collection systems are placed into service. These new systems include satellites and NEXRAD weather radars. To archive these data, they must be converted from the original media, particular to the collection system, to the data center's archival media. This conversion is also data rescue.

Access to NOAA's data and information is a significant area of concentration for the ESDIM Program. Access is the process of making available data and information held by an individual, office, or organization to a much larger audience.

High Performance Computing and Communications (HPCC) Program



Placing conduit for the Boulder (Colorado) Research and Administrative Network, a fiber optic network that supports networking research and technology. The project was supported by NOAA's HPCC program.

The mission of the NOAA High Performance Computing and Communications (HPCC) Program is to accelerate the adoption of advanced computing, communications, and information technology throughout NOAA so that the agency can accomplish its mission better and in a more cost-effective way. The Program supports acquisition of advanced platforms for high-end research computing, research on advanced programming techniques for some of the world's fastest computers, development and deployment of advanced networking technologies, and application of advanced information technologies to NOAA's missions.

The NOAA HPCC Program objectives are:

- To enable more accurate representation of the atmosphere-ocean system through the use of increased computing power, resulting in improved weather forecasts, climate predictions, and predictions related to living marine resources, and making possible better decisions by Government and industry on issues that affect both the environment and the economy.
- To enable NOAA to disseminate its vast holdings of real-time and historical information to users more completely, in a more usable form, and in a much more timely manner through increased use of the Internet and follow-on networks of the National Information Infrastructure.

- To enhance NOAA's scientific productivity through the use of advanced collaboration and model analysis tools, enabling faster, more effective communications among researchers, and improved analysis, diagnosis and visualization of model output, leading to greater understanding of modeled phenomena and more rapid improvements in NOAA prediction and information services.

Each year, the HPCC program reviews the opportunities for advanced technology exploitation and infusion. Since the beginning of the Program, the areas of emphasis have encompassed a number of projects, technologies and investigators throughout NOAA. Some of these projects explored technologies that were relatively mature and ready for transfer to operational elements of NOAA; some were at the leading edge of technology and entailed high risk. In many ways the riskiest projects are the most important for they show us what is possible.

Data Rescue Activities

The focus of the Environmental Data Rescue Program, conducted from 1996–1999, was to preserve the meteorological, climatological, geophysical, and oceanographic data stored by the NOAA National Data Centers and throughout NOAA, and to make these data more accessible to researchers and the general public.

Millions of global environmental records—including handwritten ships' logs from the 19th Century and weather observations from America's early years—are in danger of being lost. The records, many of which are located in NOAA's three national data centers, include millions of paper, film and tape records. The paper is deteriorating, the acetate-based film is degrading, and the tape is becoming obsolete. As the media deteriorate with age, records are in danger of being lost. To remedy the situation, NOAA implemented the data rescue program, which rescues data sets by converting them to more stable storage media.

During the course of this rescue effort, the following data sets have been rescued: 732,000 oceanographic surface and profile observations recorded on paper, 13.5 million paper meteorological forms, 440,000 microfiche cooperative weather observation forms, 131,000 pages of coastal data recorded on paper, 250 bathymetric maps on paper, 4,000 geophysical observations on paper, and 88,000 reels of microfilm containing meteorological data.

The Climate Database Modernization Program (CDMP), initiated in FY2000, will make major climate databases accessible via the 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 manner; and third, provide the infrastructure required so that access to these data is quick and convenient.

In the first year of CDMP, several major projects have been undertaken. A digital database was developed which contains daily climate observations for 1892–1947 for 20,000 cooperative stations. Records containing hourly observations from 1949–1999 have been imaged, thus building a digital database for records previously only available on paper or microfiche.

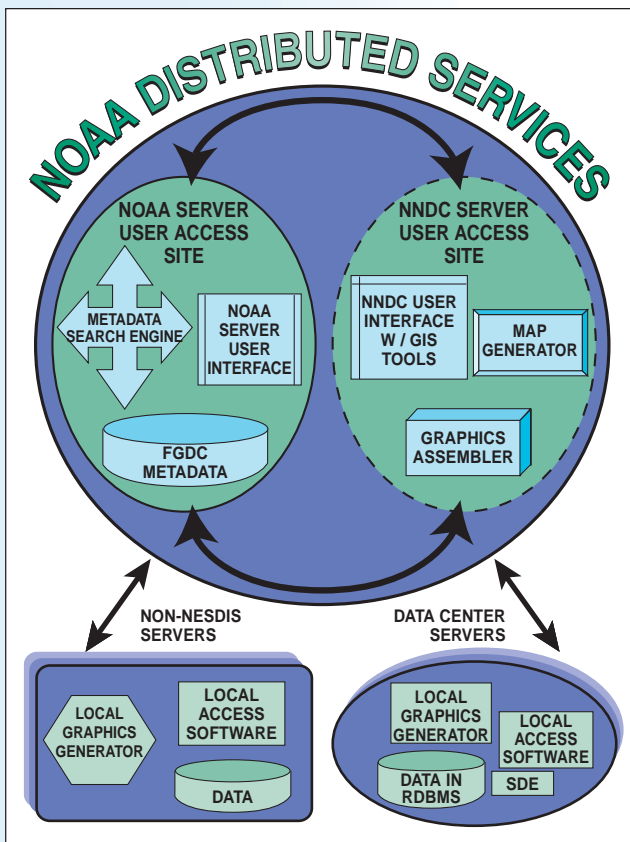
Further work includes imaging and digitizing climate record books for 1861–1990 and the imaging of all NCDC serial climate publications, 1890–1997. Station history information has been imaged and placed on-line. Incoming paper records will be imaged, digitized, and placed on-line within 48 hours of receipt. To make images Web accessible, new software and hardware have been installed. Images are being made available in ever increasing numbers as the loading process continues.

NOAA Data Directory

The NOAA Data Directory contains descriptions of more than 14,000 NOAA data sets with metadata that conforms to the Federal Geographic Data Committee (FGDC) Metadata Standard. The NOAA Directory allows a search of the descriptions using full-text indexing applications accessed via Web pages. The NOAA Directory is the keystone of the NOAA Server system. The data descriptions in 10 nodes of the NOAA Directory are available for searching purposes from the FGDC Clearinghouse, a service that has more than 100 nodes from Federal government, State, and local organizations.

NOAA Server

The NOAA Server system is a set of nationally distributed computers interconnected via the Internet that use Web technologies to provide public access to environmental data and information products. The NOAA Server uses the NOAA Data Directory to search for data sets, and connects users to data systems for browsing and ordering services.



Office of Global Programs

The Climate Change Data and Detection (CCDD) Program element of the NOAA Climate and Global Change Program focuses on the development, analysis, and stewardship of high-quality climatic reference data sets. These efforts will further NOAA's understanding of climate variability and climatic change over time-scales from days to centuries. These CCDD scientific goals are partitioned into four subprogram elements: Program-Specific Data Management, Data Archeology, Data Set Enrichment, and Climate Change Detection and Attribution.

Currently, the scientific projects supported by the CCDD element are centered on efforts to:

- Provide data and information management support to ensure the availability of critical data sets for a variety of national and international programs and assessments of primary interest to NOAA's Climate and Global Change Program. These international programs include the World Climate Research Programme (WCRP), the International Geosphere-Biosphere Programme (IGBP), the Global Climate Observing System (GCOS), and the Intergovernmental Panel on Climate Change (IPCC). National programs and assessments include activities such as the new Water Cycle initiative, the Global Water and Energy Cycle (GEWEEX) Continental-Scale International Project (GCIP) and its new initiative, GEWEX Americas Prediction Project (GAPP), the U.S. Climate Variability Program (CLIVAR), and the U.S. National Assessment activities, among others.
- Develop, quality control, and quantify time-dependent biases (homogeneity), as well as evaluate data sets for crosscutting science needs necessary to improve our ability to describe, understand, and predict seasonal, interannual, decadal, and longer-term climate variations and changes. This includes:
 - Data set enrichment: Calibrate, validate, and blend existing data sets from a variety of observing systems including space-based, *in situ*, and model data.
 - Climate change detection: Document the quantitative character of observed climate variations and changes.
 - Climate change attribution: Attribute changes in the observed climate record to specific climate forcings.

C.2 NESDIS Activities

NNDC Virtual Data System (NVDS)

The NNDC Virtual Data System (NVDS) began with NOAA's vision of a single unified system to more effectively manage the valuable data at the NOAA National Data Centers, and to provide faster and easier customer access.

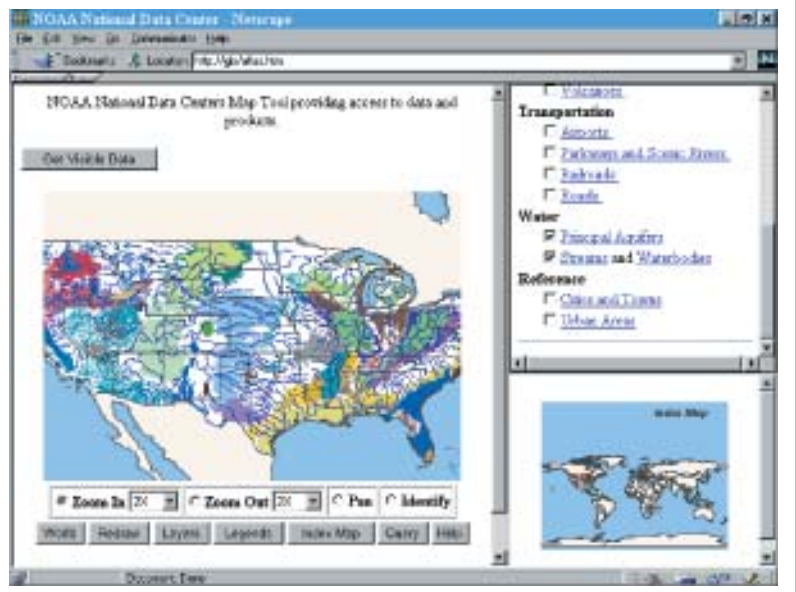
NVDS allows users to locate, access, browse, and order data and information products in an intuitive manner without regard to the data's physical location, organizational structure, underlying discipline, or storage format. NVDS provides on-line customer ordering of both on-line and off-line data.

The On-Line Data Store was developed to provide direct access to the NOAA National Data Centers. Major features include charging for on-line data access and an annual on-line subscription service that utilizes credit card billing for customers.

Data Archive Activities

The National Climatic Data Center developed, in collaboration with NASA, a NASA/NOAA Prototype Long-Term Archive system. The system archives NASA satellite data at a NASA facility, and the data are accessible through the NOAA on-line ordering system. The experience gained with this prototype will be used to help estimate costs associated with the long-term archive of NASA's Earth Observing System's (EOS) data.

The Atlas feature within NVDS empowers users with a GIS Web server that can access spatial-enabled relational databases, and display multiple discipline layers.



The Satellite Active Archive (SAA) is an operational component of the Office of Satellite Data Processing and Distribution (OSDPD). SAA is a digital library of real-time and historical data that provides electronic distribution of data and derived data products from U.S. polar-orbiting environmental satellites. The system allows users to search satellite inventories, preview sub-sampled Earth images and download data using a World Wide Web browser.

In 1997, NCDC and OSDPD officially assumed the responsibility for the distribution of Defense Meteorological Satellite Program (DMSP) data from the NASA Marshall Space Flight Center Distributed Active Archive Center. NCDC processes requests for both subscription and ad-hoc orders for DMSP data, and provides help desk support for users of these data.

Comprehensive Large Array Data Stewardship System (CLASS)

The CLASS Project is designed to enhance NESDIS capability to provide virtual environmental data and information archive and access services to the Nation through the effective application of modern, proven techniques and technology. The project places special emphasis on the ability to efficiently archive the vast quantities of satellite and *in situ* observational data currently being collected and to be collected; to provide rapid access by the public to those data in a cost-effective manner; and to safely and permanently preserve those valuable data for future generations to use.

The heart of the CLASS Project is upgraded communications capabilities, increased computer storage and power, implementation of high capacity acquisition and ingest capabilities, use of commercially available, modular hardware and software, and to expand Web access to the data and information.

This NESDIS-managed system is planned to support NOAA's requirements for environmental data and information archiving and access. CLASS will interface directly with data acquisition systems and with information systems operated by other Federal and State government agencies for the purpose of providing rapid archiving capability for the vast volumes of observational data and information being and to be collected.

C.3 National and International Coordination

National Spatial Data Infrastructure (NSDI) and Federal Geographic Data Committee (FGDC)

The development of the National Spatial Data Infrastructure (NSDI) is coordinated by the Federal Geographic Data Committee (FGDC). NSDI encompasses policies, standards, and procedures for organizations to cooperatively produce and share geographic data. The 16 Federal government agencies that make up the FGDC are developing the NSDI in cooperation with organizations from State, local, and tribal governments, the academic community, and the private sector.

NOAA staff participate in various standards committees of the FGDC. The NOAA Assistant Administrator for Satellite and Information Services serves as the Department of Commerce representative to the FGDC Steering Committee. The Department of Commerce and NOAA received the Vice President's Hammer Award in 1997 for contributions to the National Information Infrastructure.

National Environmental Data Index (NEDI)

The National Performance Review and *Access America* report called for the creation of a National Environmental Data Index (NEDI) to identify environmental data and information created, stored, and managed by Federal government agencies. NEDI is a Federal government effort managed by NOAA to provide environmental information users with access to the totality of the government's environmental data and information descriptions.

NEDI provides an Internet gateway to distributed environmental data and information descriptions to improve awareness of—and facilitate access to—data and information holdings. NEDI accesses databases held by the Department of Agriculture, Department of Commerce, Department of Defense, Department of Energy, Department of Interior, Department of Transportation, the Environmental Protection Agency, the National Aeronautics and Space Administration, the Library of Congress, and the National Science Foundation.

Most Federal government agencies are developing indexes to their data holdings under Office of Management and Budget Circular A-130, *Management of Federal Information Resources*. NEDI provides users access to these separate indexes of environmental data and information via a single query. In 1998, the NOAA NEDI Program was the recipient of the Vice President's Hammer Award as part of the National Partnership for Reinventing Government.

*Working Group on Information Systems and Services
(WGISS) of the Committee on Earth Observation
Satellites (CEOS)*

NESDIS represents NOAA interests at the Committee on Earth Observation Satellites (CEOS) and its Working Group on Information Systems and Services (WGISS). CEOS addresses the coordination of the civil Earth observation satellite programs of the world's government agencies, in addition to agencies that receive and process data acquired remotely from space. WGISS addresses the information systems and services that help CEOS agencies achieve this coordination, and that enable user ease-of-access to the Earth observation data holdings of members worldwide.

WGISS works with the CEOS participating agencies to improve satellite operations and data management by increasing the use and benefit of the Earth observation data. Also, WGISS addresses the needs of data and information users by promoting simpler and wider access to the resources they require.



*CEOS Synthetic Aperture
Radar calibration
transponder.*