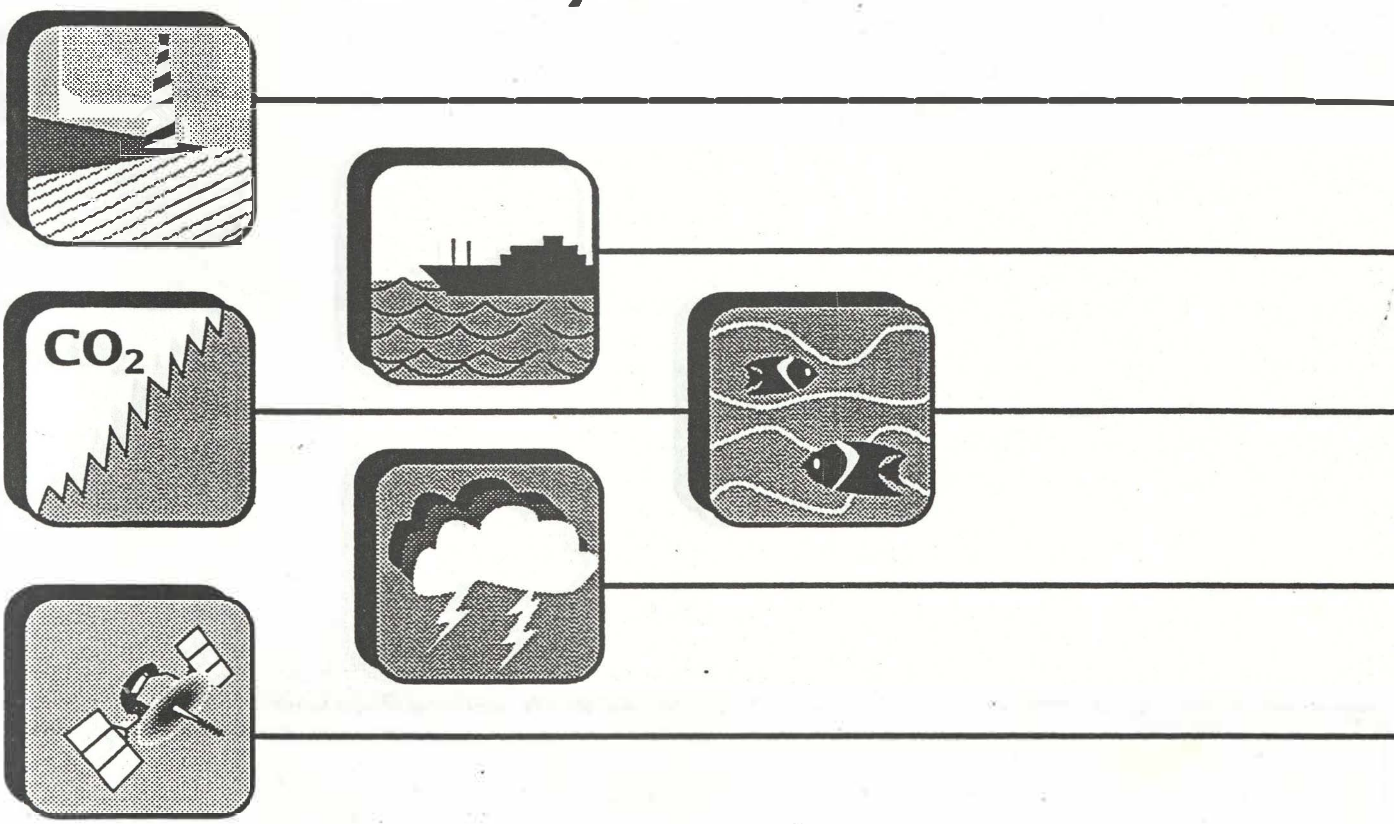
# NOAA's Data and Information Management Strategy: A Vision for the 1990s and Beyond

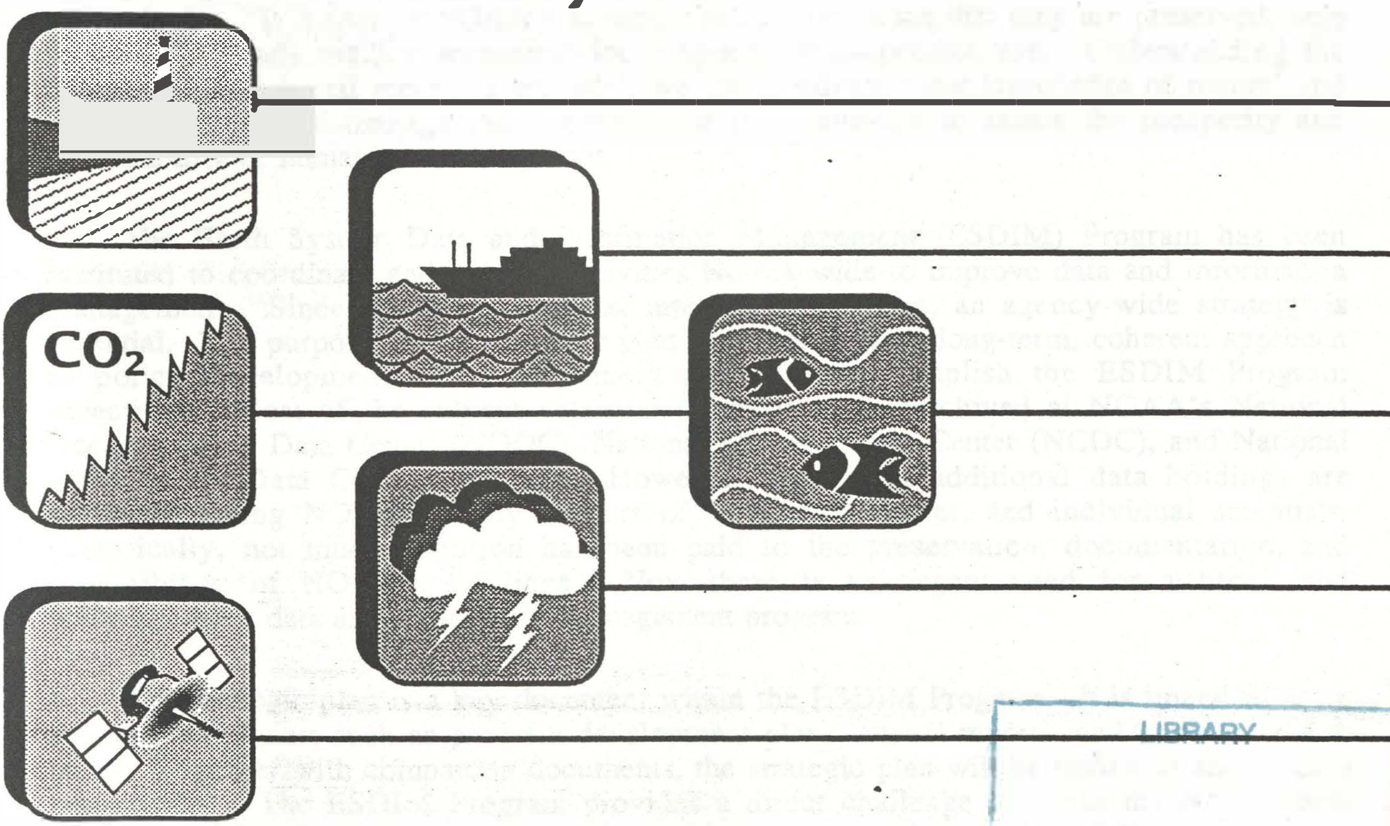



# Earth System Data and Information Management Program

**September 1992**

U.S. DEPARTMENT OF COMMERCE
National Oceanic and Atmospheric Administration

# NOAA's Data and Information Management Strategy: A Vision for the 1990s and Beyond




# Earth System Data and Information Management Program



**September 1992**

U.S. DEPARTMENT OF COMMERCE
National Oceanic and Atmospheric Administration

# PREFACE

The National Oceanic and Atmospheric Administration (NOAA) is committed to providing the science, products, and services necessary to respond to urgent national and global problems related to natural hazards, the environment, and natural resources. To carry out its mission for the 1990s, and beyond, NOAA is responsible for the development, management, and stewardship of vast amounts of high-quality Earth system data and information. These data and information must be accurate, complete, timely, and fully integrated across the spectrum of NOAA's Line Office and programmatic functions. After any operational needs and responsibilities are met, such data and information collected are still valuable. It is part of NOAA's scientific obligation to see that they are preserved, kept secure, and made readily accessible for long-term retrospective use. Understanding the Earth's environmental record is crucial if we are to advance our knowledge of natural and human systems and manage the integration of these systems to assure the prosperity and sustainability of humanity on this planet.

The Earth System Data and Information Management (ESDIM) Program has been instituted to coordinate and oversee activities NOAA-wide to improve data and information management. Since this is a complex, integrating program, an agency-wide strategy is essential. The purpose of this strategic plan is to document a long-term, coherent approach to policy development and management actions to accomplish the ESDIM Program objectives. Most of the subject data and information are archived at NOAA's National Oceanographic Data Center (NODC), National Climatic Data Center (NCDC), and National Geophysical Data Center (NGDC). However, significant additional data holdings are scattered among NOAA's many Centers of Data, laboratories, and individual scientists. Historically, not much attention has been paid to the preservation, documentation, and accessibility of NOAA's holdings. Now there is an urgent need for a broad and comprehensive data and information management program.

The strategic plan is a key document within the ESDIM Program. It is linked to more detailed documents such as program development plans, annual reports, and implementation plans. Together with companion documents, the strategic plan will be reviewed and revised periodically. The ESDIM Program provides a direct challenge to all in the NOAA data management family to contribute to a better understanding of the world we live in.

John A Knauss

John Knauss
Under Secretary for Oceans and
Atmosphere and Administrator

## TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

## LIST OF FIGURES

# EXECUTIVE SUMMARY

## PURPOSE

By order of the U.S. Department of Commerce, the National Oceanic and Atmospheric Administration (NOAA) is given a mandate to: determine requirements and develop long-range plans for data management; manage receipt, processing, archiving, dissemination, publication, and application of climatological, geophysical, oceanographic, and related marine fisheries data; and encourage international cooperation and participation in data archiving programs. Thus, NOAA is the designated steward of our national treasure of Earth system data. To be able to conduct multidisciplinary studies on decadal time-scale data sets, proactive, long-term data management is essential. With constant change in data collection programs—instruments, platforms, observing sites, practices, and personnel—attention to data continuity is essential. Stewardship requires that NOAA's data be preserved, documented, and accessible to the extent that the data will be equally useful today, 20 years hence, and beyond.

The purpose of this document is to present the strategy by which NOAA proposes to carry out its data and information management responsibilities under this mandate. It presents a vision for a vastly improved NOAA-wide Earth system data and information management program for acquiring, quality-controlling, storing, preserving, and providing easy user access to all of its environmental data bases over the full range of operational through retrospective scientific use. It recognizes that a comprehensive, far-reaching data system modernization program is needed if NOAA is to carry out its critical responsibilities into the 21st century.

## NOAA'S DATA MANAGEMENT CRISIS

The present state of management of environmental data and information in NOAA can be greatly improved. Large amounts of data are inaccessible, incompatible, or have deteriorated to an unusable state. Many of the problems stem from the lack of coordinated data collection and processing procedures or from an aging technological infrastructure that does not have the capacity for adequate data communication, processing, and storage. This crisis is dramatically summarized in figure E-1. Clearly, NOAA's priceless treasure of historical Earth system data is at risk of being lost forever. And, as if this were not serious enough, the future looks worse. NOAA now acquires about 10 terabytes* of environmental

* 1 terabyte = 1,000,000,000,000 bytes

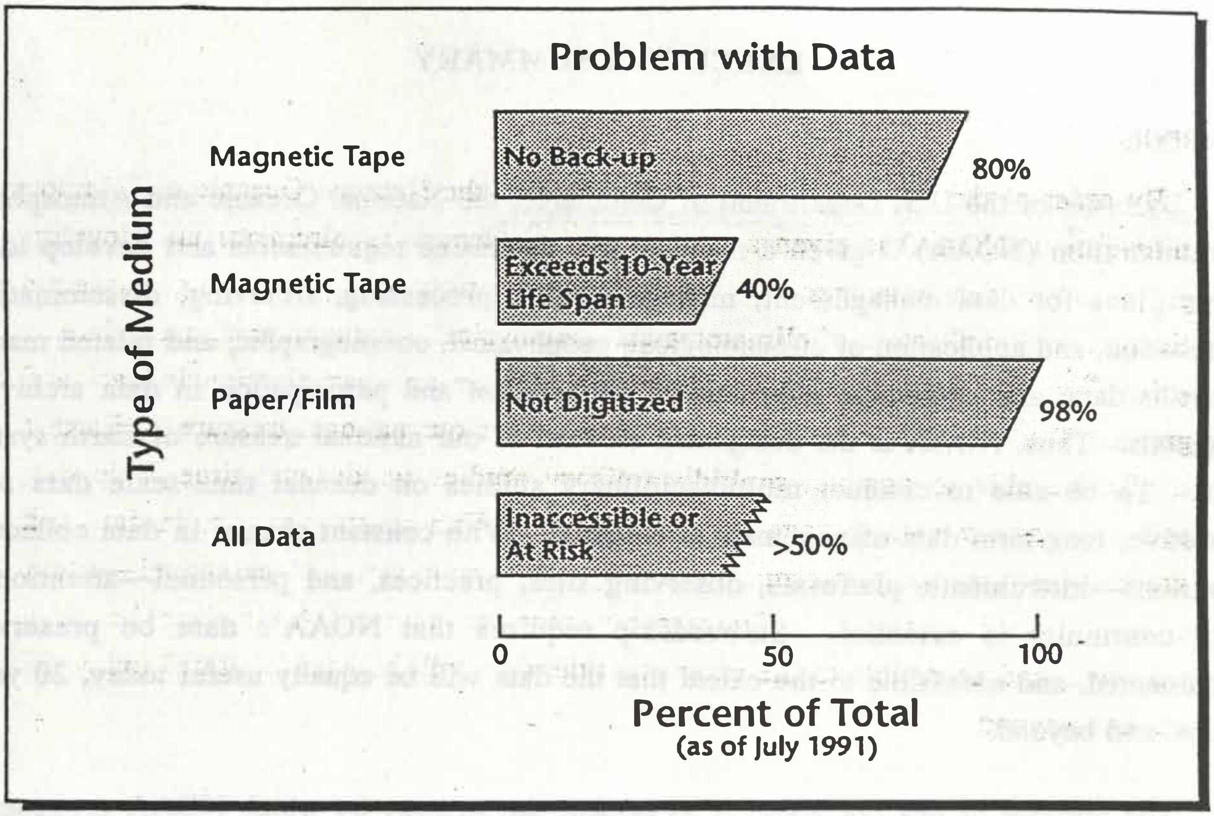



**Figure E-1. NOAA's Data Management Crisis**

data every year, and its data management systems are operating at full capacity. In fact, NOAA has not been able to keep pace with the 10 percent per year digital data volume growth of the last few years. It is estimated that NOAA will have to handle an increase of approximately 100 terabytes per year by the year 2000. Therefore, the current problems must be remedied and the trends reversed. To field the vast improvements needed by the end of this decade, the time for action is now.

## Earth System Data and Information Management (ESDIM) Program Objectives

To resolve this crisis, NOAA has initiated the ESDIM Program. The purpose of the ESDIM Program is to integrate common elements of data and information management through the modernization of NOAA's technological infrastructure, with careful consideration of the research and operational needs of all NOAA programs. The ESDIM Program recognizes this must be a NOAA-wide effort and involve the contributions and cooperation of all NOAA Line and Program Offices. The ESDIM Program will make maximum use of NOAA's vast resources and experience and build on existing systems and near-term, agency-wide data management capabilities and structure, by molding them into a broad, coordinated system, with the "look-and-feel" of a centralized system.

The ESDIM Program objectives can be summarized as follows:

- Build a top-level consensus within NOAA on data and information issues and formulate a vision of the agency's data and information management strategy for the 1990s and beyond
- Rescue critical NOAA environmental data currently at risk of being lost
- Improve access to NOAA environmental data and information for scientists and decisionmakers
- Modernize and interconnect environmental data systems throughout NOAA to increase their capability and responsiveness
- Assist in developing standards for data documentation, data quality, and network connectivity
- Provide agency-wide guidance on developing policies related to environmental data management.

## STRATEGIC APPROACH

NOAA recognizes that a studied, visionary approach is required to meet the far-reaching objectives of the ESDIM Program, and that this approach must be coordinated NOAA-wide to carry out the agency's mission. Building upon existing resources where possible, the ESDIM Program will identify where operational efficiency can be enhanced through new technology or innovative approaches to the task of environmental data and information management. The ESDIM Program will promote a modern approach to systems, operations, and services that will meet the needs of NOAA's broad spectrum of users in the 1990s and beyond.

The strategic approach is the philosophy and doctrine of the ESDIM Program strategic plan. The principal elements of this strategic approach are:

- **NOAA-Wide Participation**—NOAA consists of numerous offices and laboratories conducting many programs, where diverse environmental phenomena are being observed, recorded, researched, and archived. These activities result in numerous products and services, and contribute to a vast environmental data base. To make these results useful and accessible, the ESDIM Program will promote the agency-wide coordination and cooperation which is essential.
- **End-to-End View**—The ESDIM Program has adopted an end-to-end view of data management*, as illustrated in figure E-2. This permits the optimization of the overall data management process to: achieve the desired level of quality assurance

---

* Patterned after approach in referenced NOAA Federal Plan for Meteorological Information Management.

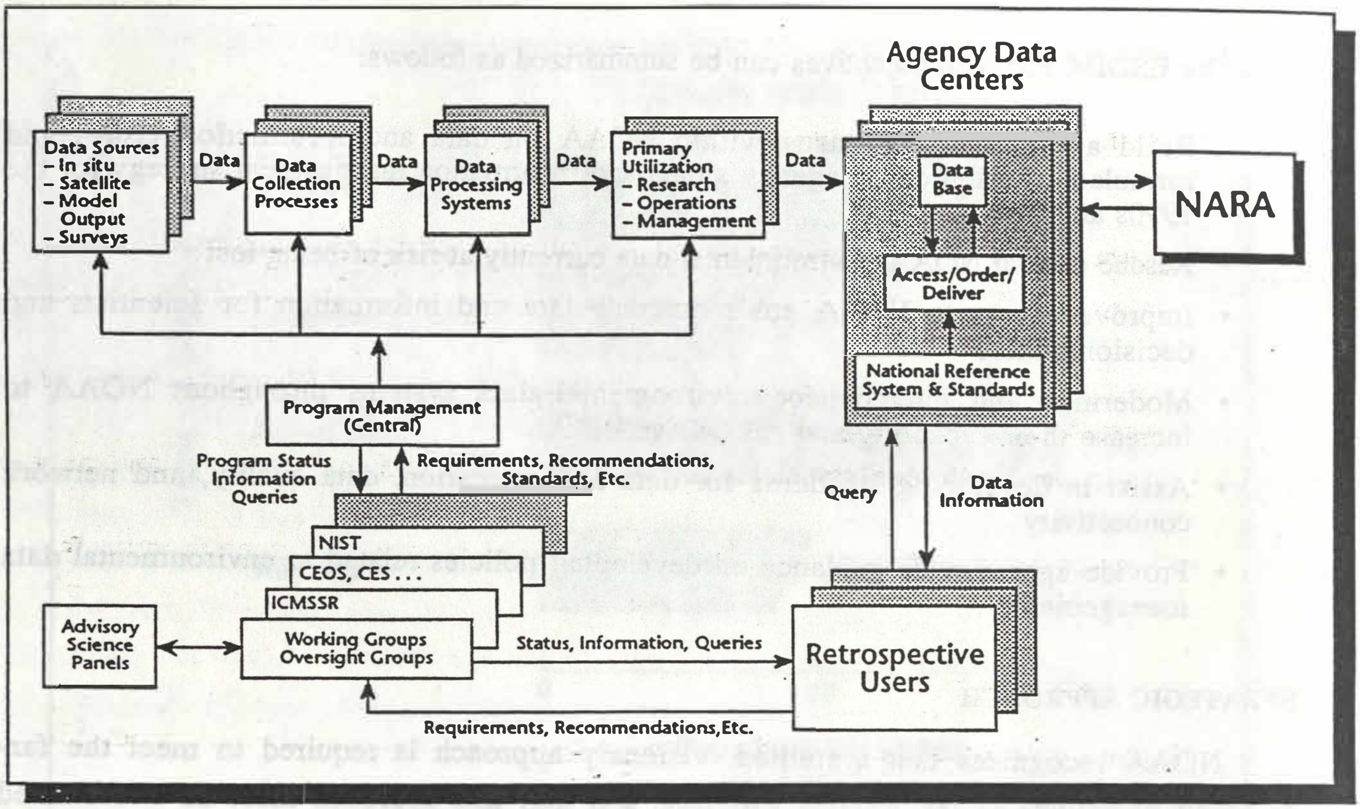



**Figure E-2. End-to-End View of Data and Information Management**

and documentation of data; archive and preserve data and derived data products for selective access and use; and, promote standards, systems, operations, expert services, and scientific oversight necessary to facilitate use of data and information by researchers and policymakers.

- **Build on Existing Capabilities**—Qualified NOAA personnel are already in place to perform the data management tasks, from planning, data collection, quality control, processing, and archiving to data dissemination. Integrating existing NOAA resources in a new data management approach takes advantage of the expertise and broad capabilities which have evolved from more than a century of making environmental observations. Particular attention will be paid to ensuring greater accessibility to the holdings of the Centers of Data, laboratories, and individual collections.
- **Establish Data Policy**—A comprehensive and coherent NOAA data policy is essential to implement the ESDIM Program. A draft policy has been prepared which covers issues such as the cost, availability, and location of NOAA's data holdings. The draft policy will be reviewed by the NOAA Program Development Board and the ESDIM Program team to formulate it to best meet the needs of the entire user community, as well as NOAA organizations.
- **Promote Education**—The end-to-end view of data and information management will require somewhat of a cultural change in NOAA. This will necessitate interprogram and interoffice education seminars and publications, which emphasize the NOAA-wide goals and purposes of the program. External programs for general science

users, policymakers, decisionmakers, and the general public will be needed to provide information on the ESDIM Program objectives and accomplishments and to identify the beneficial use of environmental data and products.

- **Develop Functional and Technical Requirements**—Identifying a comprehensive set of requirements is the first and most important step of any systems development process. The ESDIM Program will develop the requirements for data and information system modernization. Initially, functional requirements should be identified to provide top-level guidance. Then, detailed system technical requirements will be defined to permit system specifications and implementations. The ESDIM Program will coordinate overall system integration activities.

## IMPLEMENTATION STRATEGY

The proposed modernization of NOAA's data and information system constitutes a long-term program with many interrelated elements. The implementation of the ESDIM Program will require the following:

- **Rescue Data and Information at Risk**—Immediately migrate high-demand NOAA ***in situ*** data and satellite data from deteriorating tapes to more stable media, and begin digitizing endangered analog and tabular data presently on deteriorating media. Develop a long-term archive maintenance program.

- **Increase User Access to Data**—Provide more powerful and robust tools for user access to data and information at the National Data Centers and other sources of NOAA data. Provide online data set browse and visualization capabilities. Permit online data ordering and provide streamlined distribution capabilities.

- **Data Exchange and Networking**—Install and operate a wide-area electronic network for handling high-volume environmental data and information. Develop all elements of the NOAA data system using national and international standards to support interoperability; use modular and open systems architectures to permit system growth and technology insertion.

- **Information Systems**—Develop and maintain a comprehensive master directory of environmental data and information holdings, technical publications, and "gray literature" across NOAA and other organizations; coordinate with other national and international catalog interoperability programs.

- **New Data Acquisition**—Develop policies and procedures for assimilation and processing of new environmental data and information emerging from major national and international programs.

- **Data Assembly and Quality Assurance**—Assemble, validate, and document reference-quality data and information products describing Earth system behavior over significant spatial and temporal extents. Ensure accurate and complete metadata.

- **Science User Participation**—Promote mechanisms by which feedback from the science data user community can improve NOAA's services to increase productivity and research quality.

## NATIONAL AND INTERNATIONAL PROGRAM COORDINATION

The primary function of the ESDIM Program is to support the implementation of the long-range strategy that has been described. The view of the ESDIM Program is much broader than individual NOAA programs or projects that have more narrowly defined data management requirements. In response to this challenge, an integrated NOAA-wide approach to Earth system data and information management will be actively pursued. It will require active involvement of all NOAA Line and Program Offices.

Nationally, there are several mechanisms for inter-agency coordination that address data formats, catalog interoperability, and communications. The Committee on Physical, Mathematical, and Engineering Sciences (PMES) of the Federal Coordinating Council for Science, Engineering and Technology (FCCSET), in the President's Office of Science and Technology Policy (OSTP), addresses all aspects of information science and technology. The Interagency Working Group on Data Management for Global Change (IWGDMGC) promotes interagency coordination on data management topics, such as data policy, standards compliance, and interoperability for a Global Change Data and Information System (GCDIS). This working group reports to the Subcommittee on Global Change, which is a part of the Committee on Earth and Environmental Sciences (CEES) of FCCSET.

Internationally, the ESDIM Program must coordinate its activities with counterpart environmental science agencies to develop uniform data standards and formats. This cooperation can be led through NOAA's activities and responsibilities for operating World Data Centers and should include participation in the preparation of an international master directory.

## PROGRAM MANAGEMENT

The NOAA-wide data and information management activities will be managed by the ESDIM Program in the Office of the Deputy Assistant Administrator for Environmental Information Services. To ensure that the NOAA-wide strategy goals and objectives are met, the ESDIM Program will establish a number of support teams, drawing upon specialists in major aspects of the program. These teams will provide technical and operational advice, assist in the selection of program element priorities, and support overall system configuration management and quality assurance.

To realize this ambitious vision, the ESDIM Program will solicit support and cooperation from all NOAA Line and Program Offices, the user community, and counterpart national and international organizations, and will promote education in the broadest sense, to promulgate the merits of this undertaking. With more accessible and reliable information, this program will increase scientific productivity and improve scientific judgment, and will support intelligent political and economic decisions for global environmental management. This is a vision that can not be allowed to fail, for it offers a critically needed basis for positive action to sustain the Earth's environment and the quality of human life.

# SECTION 1
# OVERVIEW AND STATEMENT OF THE PROBLEM

Data and information provide the foundation of environmental science studies. The central role that Earth system data and information assume in environmental studies relates to the ever-changing nature of our environment. With the increasing data base volumes, and the complexity and multidisciplinary nature of environmental studies, effective data and information management is of paramount importance. The Earth System Data and Information Management (ESDIM) Program, begun in FY 1991, is a NOAA-wide effort to address critical data management concerns so that NOAA can be responsive to present and future demands for its environmental data. This section sets the context in which NOAA must address existing and anticipated problems through the ESDIM Program.

## 1.1 NEED FOR STEWARDSHIP IN DATA AND INFORMATION MANAGEMENT

Large numbers of environmental measurements are being made in support of NOAA operational programs, such as weather monitoring and forecasting. It is of the utmost importance and priority that these data be made available to a broader spectrum of users in a timely and organized manner. We live in an "information age," an age in which data accessibility, quality control, visualization, and format commonality are essential for scientific studies of retrospective data.

To gain an understanding of the Earth system processes to the point where the dominant elements of such processes and their extent of influence are revealed, a variety of data sets must be studied through multidisciplinary science programs. Proactive, long-term data management is essential to conduct multidisciplinary studies on decadal and longer time-scale data sets. A progression of seemingly unrelated, sometimes episodic events can be studied properly only if a nearly continuous series of observations are available. A good illustration of this is shown in figure 1-1. Changes of global average temperature compiled from COADS raw observations are shown in the top curve and indicate almost random fluctuations over the period. However, the lower curve, produced by the Intergovernmental Panel on Climate Change (IPCC, 1990) makes use of retrospective data management capabilities and shows unambiguously a gradual warming trend. Therefore, valuable information about global warming can be extracted from this progression of data when proper data management techniques are applied. To ensure the "complete" progression of data, from the earliest point in time possible until the present time, NOAA must care equally for all data, from the early data sets, to those that are being observed currently. This is stewardship.

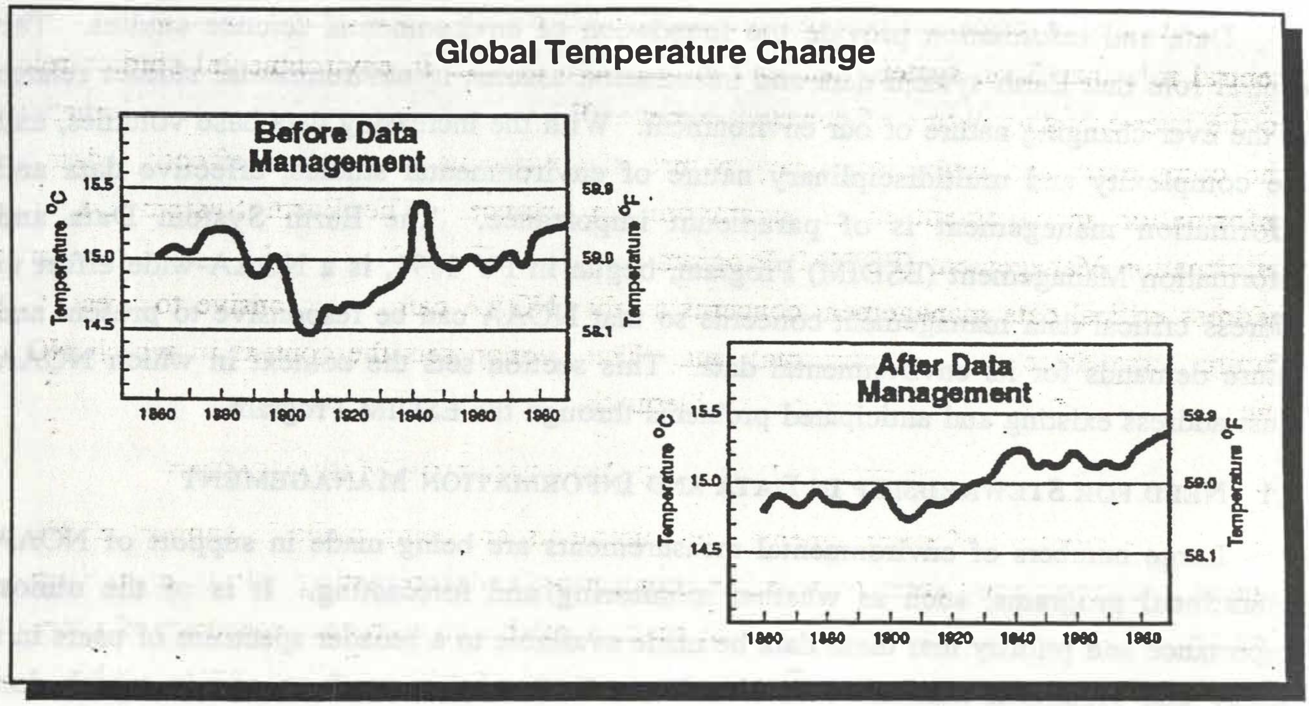



**Figure 1-1. Importance of Data Management**

Change is inherent to long-term data collection programs. Instruments, platforms, observing sites, practices, and personnel all change. Earth system problems such as marine pollution, ozone depletion, fisheries stock reduction, sea level rise, and global warming require long-term observational data to develop an understanding of the problem and to formulate mitigation strategies. When that long-term record is broken, the lack of continuity causes analytical paralysis. It is essential for NOAA to provide data continuity in the face of change. This also is stewardship.

As steward for the Nation's Earth system data and information, NOAA must respond to critical data needs addressing the scientific, economic, and environmental issues that may well dominate global interests in the 21st century. This role requires that a long-range, NOAA-wide program be developed to provide data guaranteed to meet the "20-year test." This test, coined in the U.S. Global Change Research Program, implies that data are preserved, documented, and accessible to the extent that the data will be as useful 20 years hence as they are today.

## 1.2 Environmental Science Programs and the Importance of Effective Data and Information Management

The global physical environment, that is, the atmosphere around and above us, the oceans that make up 70 percent of the planet's surface, and complicated land masses, is critical to human well-being, yet is largely taken for granted. The impacts of man's activities on this global environment are not well understood. Environmental science programs are the means to better understand important but complex Earth system phenomena and processes. Understanding achieved through measurements and research directly affects human welfare today, as well as secures the welfare of future generations.

Natural phenomena, including hurricanes, blizzards, droughts, tornadoes, storm surges, tidal waves (tsunamis), earthquakes, volcanic eruptions, and man-made disasters, such as oil spills, oil well fires, and nuclear reactor meltdowns, can have substantial adverse effects on human welfare and quality of life. Understanding the impact of these events, and taking corrective or preventive measures, can minimize potential human and economic losses. In 1991 alone, two major environmental events made this point vividly clear: the eruption of the volcano Pinatubo in the Philippines and the battering of the U.S. coast by Hurricane Bob. Most of the results of Hurricane Bob are readily observable, with property damage estimated at $1.5 billion in Massachusetts, Rhode Island, and Connecticut. The outcome of the Mt. Pinatubo eruption, however, is more subtle and it may take years of careful monitoring and study to isolate possible effects, such as global cooling due to masking by airborne volcanic ash and sulfur dioxide. Global cooling of 1°F in one year is expected. On the other hand, global warming due to "greenhouse gases" is estimated to be a few tenths of a degree each decade. Both of these events will require extensive scientific effort to determine the complex, conflicting environmental effects. Numerous multidisciplinary data sets are needed to carry out such extensive studies; simply put, well-documented, high-quality data must be readily accessible to retrospective science users. These data must also be of global extent to achieve the spatial and temporal correlations required, and must cover decades to identify and quantify trends accurately.

A natural disaster is, in effect, the outcome of a "dramatization" of ceaseless environmental change. Merely understanding the catastrophic behavior of the environment is not enough. Changes such as global warming and cooling, depletion of the ozone layer, and rise of sea level and resultant flooding, affect the entire globe, and have a direct relationship to our climate, agriculture, productivity, physical well-being, quality of life, and national security. They must be studied thoroughly to determine the precautions and corrective actions that we must take to minimize their impact. These studies can not be undertaken without access to archives of vast quantities of science quality environmental data and information that span decades of multidisciplinary global measurements. Effective data and information management is pivotal to data quality, archiving, and access.

Man-made disasters can be equally devastating. The Prince William Sound oil spill by the vessel Exxon Valdez, the Chernobyl nuclear reactor meltdown, and the environmental sabotage during the recent war with Iraq are vivid examples of disasters of human origin. They, too, need to be studied in terms of their environmental impact: plume trajectories, particle dispersion, and effects on flora and fauna. Here, effective data management will result in the much-needed, quick response to effect rapid evaluation and correction.

Disasters drive home the need to continuously monitor and to better understand our Earth's environment. Science programs and operational services use data symbiotically. Through a feedback mechanism, environmental data used primarily to satisfy operational needs are subsequently sent to archives for retrospective research that will lead to increased understanding of these phenomena and the ability to predict more precisely their future occurrence or impact. Thus, operational quality is enhanced through research aided by effective data management.

In addition to science applications, the same data and information are important to decisionmakers in private and government sectors. The legal, commercial, construction, industrial, and insurance sectors use them to ascertain past environmental conditions for damage litigation, to obtain criteria for event scheduling and engineering design, and to establish insurance thresholds and standards. Environmental data and analyses will play an increasing role in regulations, and national and international policy development. Therefore, relevant, accurate, timely environmental data and information must be made routinely available for all potential users.

## 1.3 NOAA'S CHARGE

Through its mission to secure our national interests in environmental sciences, NOAA has acquired, and continues to acquire, comprehensive Earth system data and information holdings. Major NOAA facilities that hold data and information are listed in Appendix A. These holdings are truly a treasure that provides a description of the planet Earth that is unsurpassed by holdings of other national agencies or foreign countries. The importance that the Department of Commerce (DOC) places on data and information management is indicated in its mandate to NOAA:

**The National Oceanic and Atmospheric Administration shall...**

- Determine requirements and develop long-range plans for...data management
- Manage receipt, processing, archiving, dissemination, publication and application of climatological, geophysical, oceanographic, and environmental data
- Encourage international cooperation and participation in...data archiving programs.

Department of Commerce
Organization Order #25-5 (1989)

To carry out this mandate, the ESDIM Program must promote a NOAA-wide, end-to-end view of data and information management.

## 1.4 THE PROBLEM AND ITS COMPLEXITY

Updating and strengthening NOAA's data management system will prevent loss of critical data. If current management practices persist, the situation will get dramatically worse due to the significant increase anticipated in environmental data flow, as pointed out by the General Accounting Office (GAO) in a report to the Senate entitled *Environmental Data: Major Effort Is Needed to Improve NOAA's Data Management and Archiving*. In that report, the following projection is made:

- **The volume of digital data archives is expected to increase from 96 terabytes* in 1990 to 410 terabytes** in 2000.**

The increasing scale of environmental science projects will require considerable management attention and resources, as well as the use of advanced data storage technologies. Figure 1-2 shows NOAA's annual expenditures for data management as a percent of its total budget, in relation to its growing cumulative volume of archived digital data. Clearly, as the magnitude of NOAA's data handling requirements has increased over the decade, the resources available to respond have not kept pace.

In fact, the system is so strained that some of NOAA's valuable environmental data may be lost. One potentially serious source of loss is due to the deterioration of storage

* 1 terabyte = 1,000,000,000,000 bytes

** This figure has been revised by recent estimates to 750 terabytes.

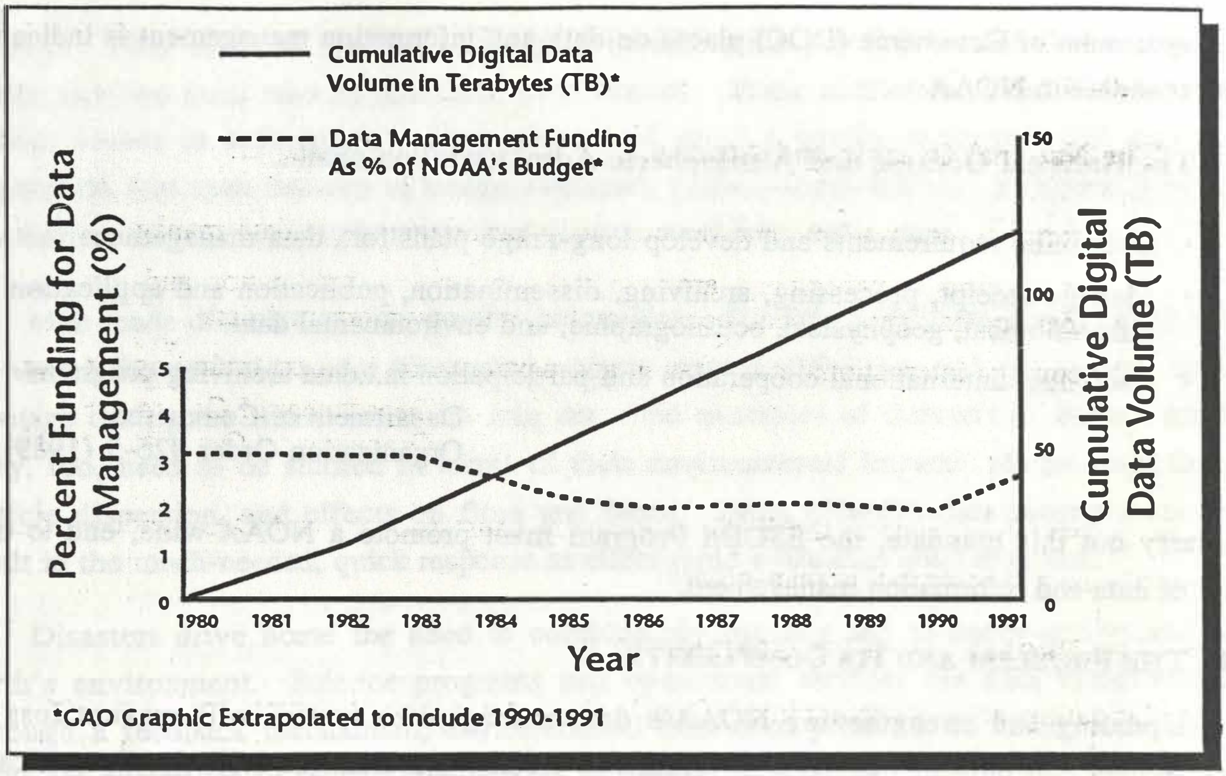



**Figure 1-2. GAO Graphic of NOAA's Data Crisis**

media. For example, many of NOAA's archived magnetic tapes have exceeded their expected ten-year lifetime, and most are not backed up, should they be accidentally destroyed (Reference GAO report). Similar deterioration is also occurring with film and paper data records. There is also a potential loss of data that have never been placed in an archive. Their existence is not widely known and they are not subject to requisite data maintenance procedures. Often an expert in a specialized field has compiled vast amounts of episodic data that are kept in personal files and are not transferred to a NOAA archive. These data are at risk of being lost due to project or career termination.

Another critical problem is that most of NOAA's data are now inaccessible to a certain degree. To extract information, direct interaction with the repository personnel is required, which is time-consuming and costly. A customer normally does not know what data and information are available; and directories, catalogs, and browsing capabilities are not generally available. Descriptions of data sets, in terms of formats, parameters, quality, discontinuities, and gaps, are not uniform, and at times not generally available. NOAA's National Data Centers and other repositories of data are not interlinked. Procedures for receiving data from these repositories are not standardized. These deficiencies and others make access to the data sets slow and inefficient, and at times, may completely bar access.

(For example, a block of Geostationary Operational Environmental Satellite (GOES) data which has been stored on magnetic media requires approximately the same length of time to access and transfer, as the original observation time for that block.)

Environmental research involves complex and unusual requirements for access to data and information. Earth system studies have multidisciplinary, multidecadal, and global data requirements. To evaluate comprehensively the dominant factors (often interactive and nonlinear) that influence a particular environmental phenomenon, many relevant data sets must be consulted. The recent eruption of Mount Pinatubo is a case in point. It is far too dominant an event to neglect in global climatic change studies. In addition, the structure of Earth system data and information is inherently complex. The data sets are heterogeneous, some contain few parameters, others a great many; temporal and spatial coverage vary substantially; and, different storage media, operating systems, and computers are employed at numerous, geographically dispersed archives.

The information referenced by the GAO report is now almost 2 years old. More recent analyses of the data handling requirements from the Next-Generation Doppler Radar (NEXRAD), the GOES, and the Polar Orbiting Environmental Satellite (POES) project the annual archive growth from these systems to approach 100 terabytes per year, circa 1996. This is illustrated in figure 1-3. Considering this level of ingest, the total archive requirement for the Year 2000 has been revised, and is now projected at about 750 terabytes, as illustrated in figure 1-4. This represents a five-fold increase over the current archive.

In 1991, the non-digitized analog records were estimated to be the equivalent of 50 terabytes of digital data (Rotar, 1991).* The estimated growth in non-digitized records is illustrated in figure 1-5. While all types of data show steady increases in volume, the digital data volume will experience the steepest ascent, placing the greatest burden of data handling on the various centers that process and archive data. The increased volume of incoming data brings requirements that compete with other Data Center tasks for limited resources. Figure 1-6 illustrates the anticipated decline in data that NOAA can archive, in percent, compared with the increasing volume of data held by NOAA Data Centers. The volume of digital

* NOAA Data Center Requirements Study, July 12, 1991

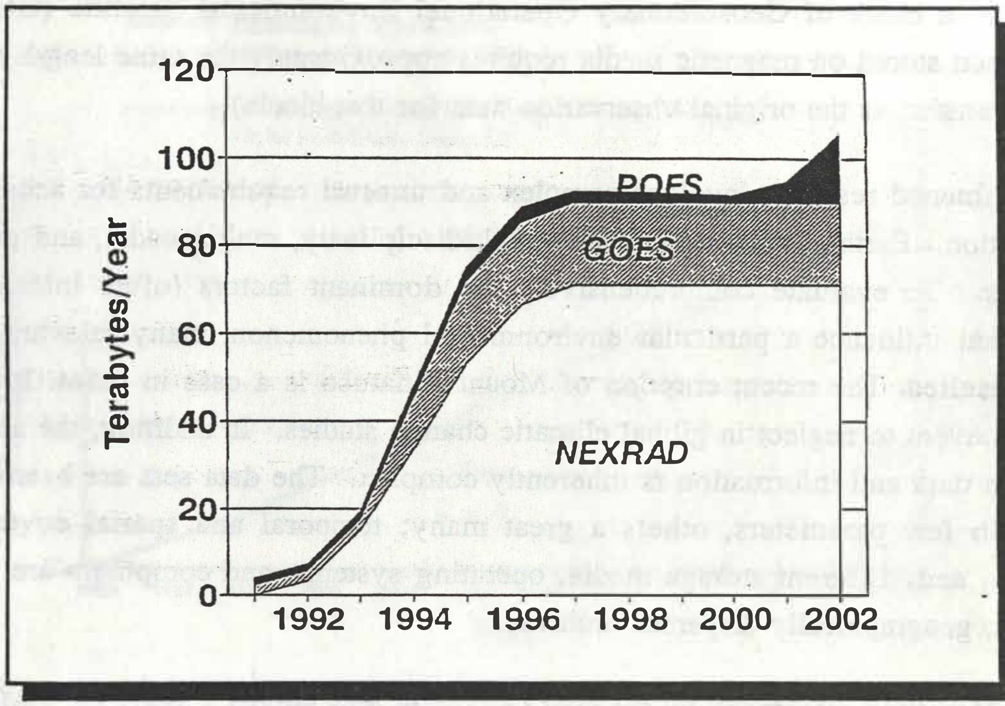



**Figure 1-3. Major Contributors to Annual Archive Growth**

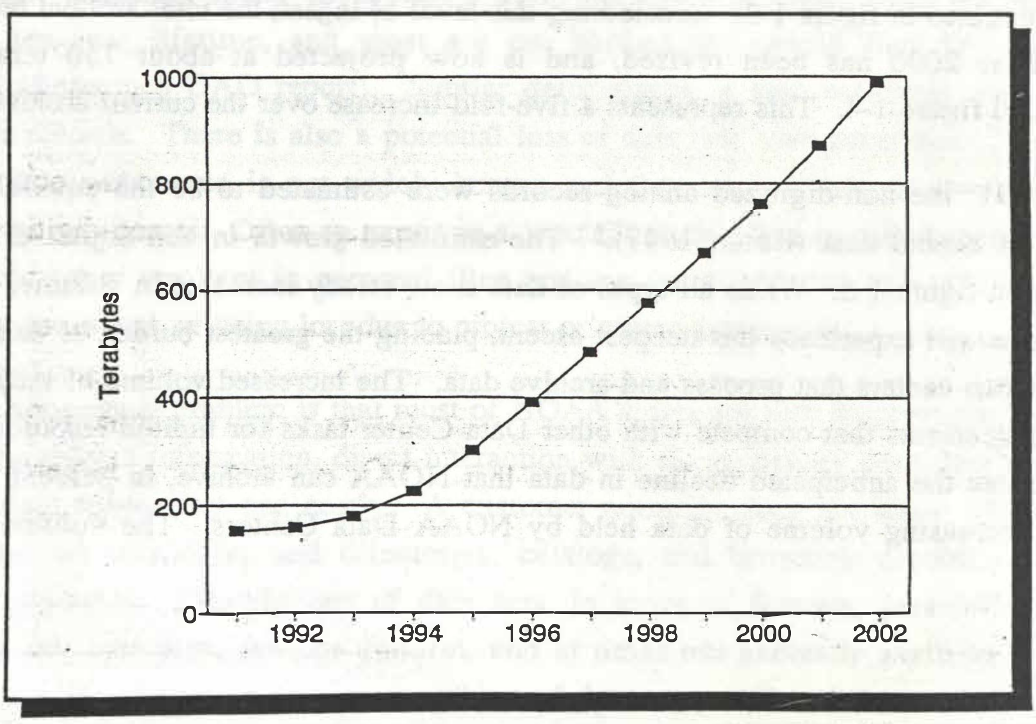



**Figure 1-4. Projected Archival Digital Data Volume**

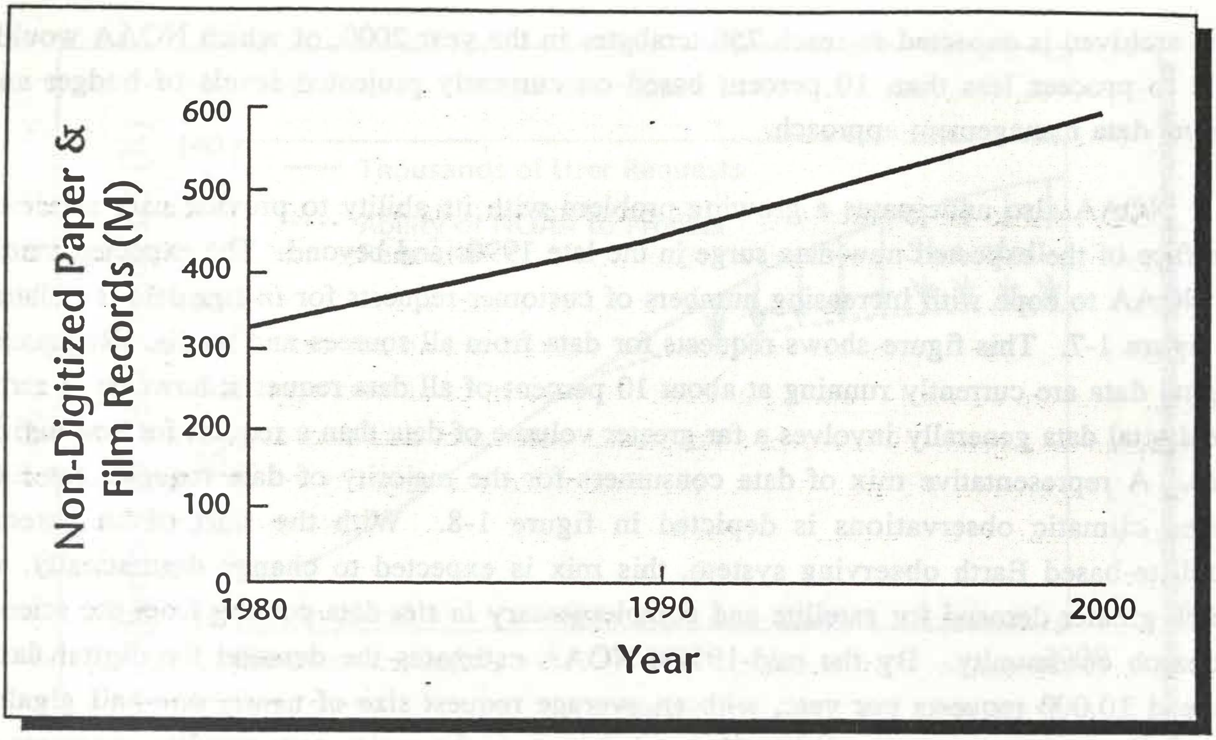



**Figure 1-5. Actual and Projected Archival Non-Digitized Data Volume, 1980-2000**

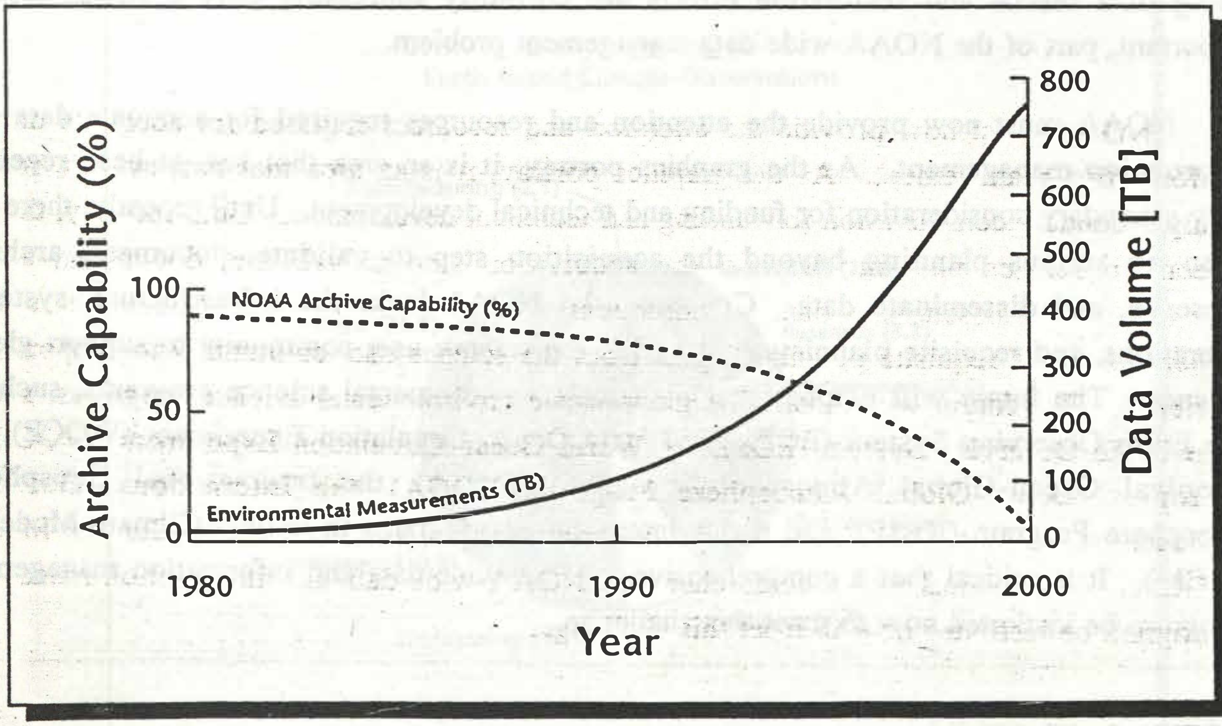



**Figure 1-6. Trends in NOAA's Ability to Handle Its Holdings**

data archived is expected to reach 750 terabytes in the year 2000, of which NOAA would be able to process less than 10 percent based on currently projected levels of budget and a failing data management approach.

NOAA also anticipates a growing problem with its ability to provide user services in the face of the expected new data surge in the late 1990s and beyond. The expected inability of NOAA to cope with increasing numbers of customer requests for *in situ* data is illustrated in figure 1-7. This figure shows requests for data from all sources and media. Requests for digital data are currently running at about 10 percent of all data requests; however, a request for digital data generally involves a far greater volume of data than a request for non-digitized data. A representative mix of data consumers for the majority of data requests for Earth-based climatic observations is depicted in figure 1-8. With the start of an extensive satellite-based Earth observing system, this mix is expected to change dramatically, with much greater demand for satellite and complementary *in situ* data coming from the scientific research community. By the mid-1990s, NOAA estimates the demand for digital data to exceed 10,000 requests per year, with an average request size of nearly one-half gigabyte* (equivalent to one compact disk). Beyond that time, the major new satellite programs will begin to come online, the number of requests will increase, and the average request size is expected to be at least 10 times greater. Figure 1-9 illustrates some of the major shortfalls in the present state of NOAA's data and information management. This indicates that the on-going data rescue and restoration efforts are currently addressing only a small, though important, part of the NOAA-wide data management problem.

NOAA must now provide the attention and resources required for adequate data and information management. As the graphics portray, it is an area that has, at best, received only secondary consideration for funding and technical development. Until recently, there has been no serious planning beyond the acquisition step to validate, document, archive, preserve, and disseminate data. Consequently, NOAA lacks the infrastructure, systems, operations, and requisite planning inputs from the science user community to support global science. The future will emphasize global-scale environmental science programs, such as, the Earth Observing System (EOS), the World Ocean Circulation Experiment (WOCE), the Tropical Ocean-Global Atmosphere Program (TOGA), the International Geosphere-Biosphere Program (IGBP), and major international programs in Global Climate Modeling (GCM). It is critical that a comprehensive, NOAA-wide data and information management program be instituted now to meet this challenge.

---

* 1 gigabyte = 1,000,000,000 bytes

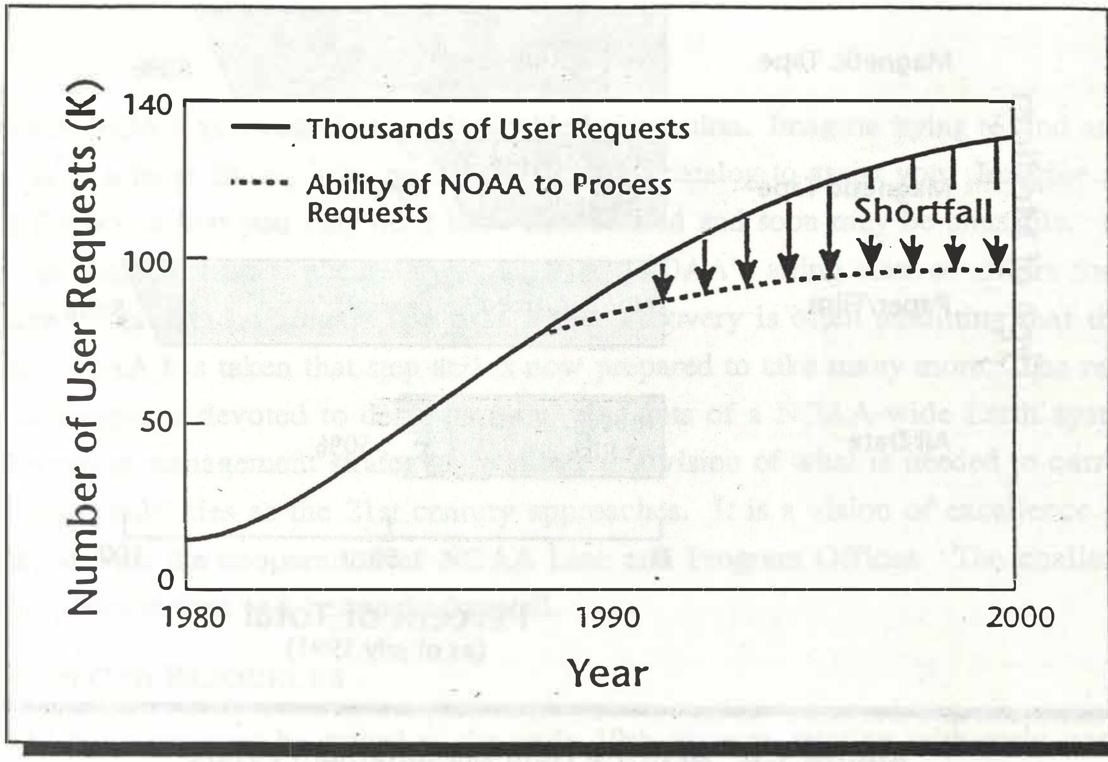



Figure 1-7. Projected Shortfall in Filling User Requests

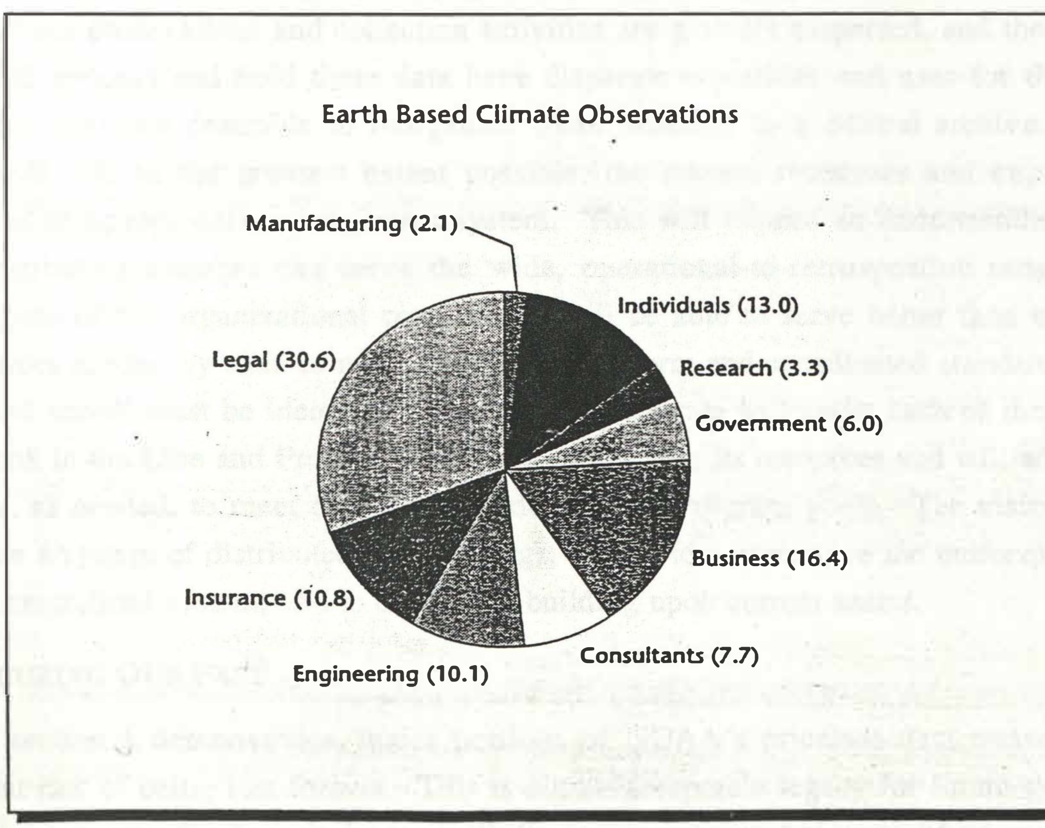



Figure 1-8. 1991 Customer Request Profile (%)

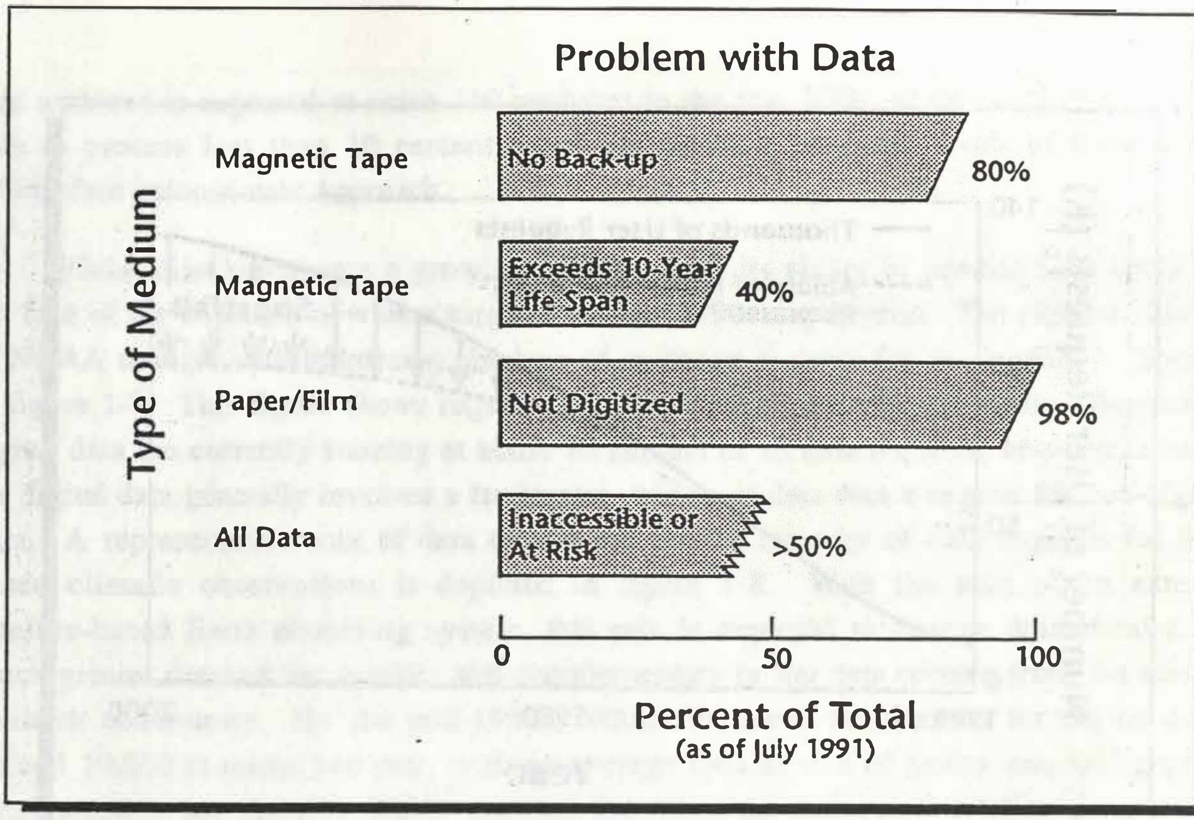



**Figure 1-9. NOAA's Data Management Crisis**

# SECTION 2
# THE VISION

Picture NOAA as a vast library of valuable information. Imagine trying to find and check out a book in a large library with no librarian and no catalog to assist you. Imagine also that some of the books that you may need have deteriorated and soon may be unusable. Unfortunately, this characterization accurately summarizes NOAA's ailing state of affairs for a large percentage of its data holdings. The first step to recovery is often admitting that there is a problem. NOAA has taken that step and is now prepared to take many more. The remainder of this document is devoted to describing the elements of a NOAA-wide Earth system data and information management strategy. NOAA has a vision of what is needed to carry out its ESDIM responsibilities as the 21st century approaches. It is a vision of excellence that will be achieved with the cooperation of NOAA Line and Program Offices. The challenges are many and contributions will be sought from all.

## 2.1 Using Our Resources

NOAA's roots may be traced to the early 19th century, starting with early coastal and marine life surveys. Today, NOAA has a vast array of personnel, instruments, facilities, and platforms involved in handling oceanographic, atmospheric, geophysical, and fisheries data. The data from observations and collection activities are globally dispersed, and the organizations which process and hold these data have disparate objectives and uses for them. It is neither practical nor desirable to reorganize these holdings to a central archive. Instead, NOAA will use, to the greatest extent possible, the current resources and experience to develop an integrated data management system. This will require an understanding of how each contributing element can serve the wide, operational-to-retrospective range of data users. Some of the organizational components will be able to serve better than others, and the resources needed by each to meet some set of uniform and coordinated standards for data quality and access must be identified. The ESDIM Program will assist each of these components found in the Line and Program Offices in identifying its resources and will add to these resources, as needed, to meet the NOAA-wide ESDIM Program goals. The vision, then, is to produce a system of distributed data holdings, which, to a user, have the uniform look-and-feel of a centralized system, and to do this by building upon current assets.

## 2.2 Securing Our Past

As section 1 demonstrates, major portions of NOAA's priceless data treasures are at significant risk of being lost forever. This is not an acceptable legacy for future generations. NOAA accepts the fundamental responsibility to secure its holdings of historical Earth system data. The ESDIM Program vision is that all data that are critical to a better understanding of our global environment will be preserved and made readily accessible. Some strides are currently being made in this direction through projects to migrate high-demand NOAA satellite and *in situ* data from deteriorating tapes to new, more stable media, and to begin digitizing endangered analog and tabular data presently on deteriorating paper and film media. This rescue and restoration operation is, however, a mammoth task that has just begun and will continue into the foreseeable future. The paper pages and film reels each numbers in the tens of millions; add to these tens of thousands of microfiche, negatives, photoprints, and charts, and about 100,000 magnetic tapes that have already exceeded the life expectancy for this medium. The salvage process requires considerable devotion of time, labor, equipment, and facilities. There is a valuable lesson learned here, however: data preservation must be an ongoing operation in a successful, long-range data management plan. The ESDIM Program will not repeat past errors, and will establish proper archive maintenance procedures NOAA-wide. All Federal guidelines and standards for data preservation will be fully implemented in the National Data Centers. The ESDIM Program will assist in the development and prioritization of near-term projects for data rescue and restoration NOAA-wide, and will support, as needed, the long-term planning and conduct of data preservation activities for the specialized needs of the Centers of Data.

NOAA envisions a future in which the condition of all of its data is well known and is included in a comprehensive preservation plan. Multidisciplinary peer groups will be assembled to carefully evaluate the importance of selected data and assign priorities for rescue, restoration, and preservation. Some data may be deemed surplus, but no data will be lost due to simple neglect.

## 2.3 Engaging Our Future

NOAA's planned programs for the 1990s will have significant data management implications. Handling the expected volume of data will require careful planning. The opportunity exists now to build into these programs the proper data quality assurance so that the operational and retrospective data user needs are met and the continuity of critical data sets is ensured. This introduces an end-to-end approach to data management, as described in section 3.2. NOAA must assure that its data will be useful today and at any time in the future. The ESDIM Program will promote this approach NOAA-wide. Summaries of some of the major programs to be considered follow.

NOAA has recognized that its weather forecast and warning operations need major upgrades to meet national weather information requirements for the 1990s and beyond. In response, it prepared a comprehensive plan for the modernization of the National Weather Service (NWS). Scheduled to be completed in the late 1990s, this series of programs will infuse new technologies to upgrade and streamline services and greatly increase the effectiveness of NWS hydrometeorological prediction capabilities. The NWS observational programs that will require end-to-end data management attention include:

- **Next-Generation Weather Radar (NEXRAD).** A network of advanced Doppler radars to measure the motion of the atmosphere responsible for severe weather such as tornadoes, to detect heavy rainfall, and to increase lead times for predictions. With a full deployment of 175 sites by 1997, the retrospective data handling requirements have been estimated to be over 70 terabytes per year.
- **Next-Generation Environmental Satellites.** A new series of geostationary meteorological satellites to provide higher spatial and temporal resolution imagery and data to aid shorter-range warnings and forecasts; a new series of polar-orbiting meteorological satellites to provide improved, all-weather atmospheric guidance for medium- and long-range forecasts; and access to environmental data from a wide variety of non-NOAA satellites. Each satellite may potentially generate 15 terabytes of data per year, beginning in the mid-1990s.
- **Automated Surface Observing System (ASOS).** An automated electronic sensor instrument system to replace manual weather observations now taken at 250 NWS sites. About 1600 stations will generate an estimated 2 gigabytes of data per year by 1994.
- **Wind Profiler Demonstration Program.** A state-of-the-art, vertically pointing Doppler radar system, designed to be unstaffed, and providing output system diagnostic information every 6 minutes. This program is conducted jointly with NOAA's Environmental Research Laboratories. The demonstration network consists of 30 stations. By 1996, this program will produce an estimated 70 gigabytes of data per year.

NOAA's Climate and Global Change (CGC) program is a critical component of the U.S. Global Change Research Program (GCRP). The GCRP is a high-priority Presidential initiative designed to establish the scientific basis for national and international policymaking related to natural and human-induced global environmental changes. NOAA's CGC program will provide the principal operational monitoring, research, prediction, data archiving, and information dissemination services in support of the national effort. From this program, NOAA will establish a new national information service, providing high-quality predictions and assessments of the changing global climate.

NOAA's Coastal Ocean Program (COP) is a focused effort to integrate scientific and technological capabilities into well-framed strategies to improve marine products and services of all NOAA Line Offices. The COP will improve NOAA's research, modeling, and prediction capabilities, applications, and information synthesis, with particular emphasis on environmental quality, fishery productivity, and the physical impact of natural coastal hazards. This program will produce data from observation and monitoring, and generate information products from data and multidisciplinary research programs. The ESDIM Program must support the NOAA-wide development of data standards and formats to guarantee the usefulness of these data.

By the end of this century, NOAA will become a net importer of remote environmental data from both domestic and foreign satellite programs. NASA's Mission to Planet Earth program will have the most dramatic impact. Included in this program is EOS. This and other polar-orbiting satellite programs will take continuous data to image and measure the Earth system characteristics below. Cumulatively, these programs will produce several hundred terabytes of data each year going into the 21st century. The implications for the ESDIM Program are staggering and require considerable early planning.

NOAA participates in and anticipates further data exchange with numerous multiagency, national, and international programs. This will continue and will obviously have consequences for the ESDIM Program, especially for standards of data exchange. Typically, programs such as the U.S. Weather Research Program (STORM), will entail a few gigabytes of data a year. The ESDIM Program will work with these programs to ensure that even relatively small data sets are readily accessible.

NOAA is committed to providing the science, products, and services to support its expressed mission "...to describe and predict the environment." In engaging the future, this mission takes on even greater importance as NOAA must provide science quality data to support global programs that will increasingly rely upon its treasury of Earth system data and information. NOAA's products and science support will help develop information crucial to decisionmakers concerned with sustaining our global environment. New and improved products and services must be contemplated to meet future requirements. A new Earth system information center dedicated to supporting NOAA's mission may be considered for synthesizing environmental data and making information readily available to decisionmakers and other users.

## 2.4 NOAA-Wide Data Management Objectives for the 21st Century

Valuable data are effectively worthless if they can not be located or obtained by a potential user. To supply the missing librarian and catalog in the prior NOAA library analogy, the ESDIM Program envisions an ambitious, NOAA-wide, end-to-end data system modernization to provide the long-term, systematic growth of its scientific data and information management resources and improve user access. The specific objectives of the ESDIM Program are:

- Build a top-level consensus within NOAA on data and information issues and formulate a vision of the agency's data and information management strategy for the 1990s and beyond
- Rescue critical NOAA environmental data currently at risk of being lost
- Improve access to NOAA environmental data and information for scientists and decisionmakers
- Modernize and interconnect environmental data systems throughout NOAA to increase their capability and responsiveness
- Assist in developing standards for data documentation, data quality, and network connectivity
- Provide agency-wide guidance on developing policies related to environmental data management.

If the proposed modernization program is carried out in the remainder of this decade, then NOAA envisions a data and information environment for the 21st century that exhibits the following key attributes:

- **Distributed Scientific Workstations** that permit users to identify and access data and information from their desktops. Included will be online capabilities to browse, analyze, and visualize data sets. New data base architectures will be in place to support these capabilities for the massive data volumes anticipated.
- **Automatic Quality Control and Validation Systems** capable of processing the huge volumes of data collected with a minimum of manual intervention. By the end of the decade, most observing systems will be automatic and will perform the first level of quality control on-site. The next level of quality control, where data from many sources are combined and compared, will be performed at recognized data centers.
- **National Standards for Exchange of Retrospective Earth System Data** that address data format and requirements for metadata and describe the data sufficiently to allow automatic processing without any additional information. Given the multidisciplinary requirements of global change research, the standards would apply to all environmental data.
- **All Data Digitized and Cataloged.**
- **An Integrated Distributed National Earth System Data Base** that functions as a single entity. To users, all data will appear to be stored in a single location even though the data are actually stored at different network nodes. The data base will include a comprehensive directory of data sets, discipline-specific data catalogs, detailed inventories of all data sets, physical directories of all data, and station history information for point data. High-use data sets or data that are required in real time will be available online. All other data will be available from near-online

systems that automatically retrieve data from off-line media without human intervention.

- **Automatic Data Ingest Systems** that load all digital data into the archives with a minimum of human interaction. Inventory and physical directory systems will be updated automatically.

- **Online Order-Entry Systems** that permit users to interact with online systems to determine the volume of data wanted, number of input and output media units required, and the relevant charges for ordering data. The systems will provide the capability to select data, specify the media the user would like it delivered on, and the desired standard format.

To realize this vision, the ESDIM Program will solicit support and cooperation from all NOAA Line and Program Offices, the user community, and counterpart national and international organizations, and will promote education in the broadest sense, to promulgate the merits of this undertaking. With more accessible and reliable information, this program will increase scientific productivity and improve scientific judgment, and will support intelligent political and economic decisions for global environmental management. This is a vision that cannot be allowed to fail, for it offers a critically needed basis for positive action to sustain our environment and our quality of life.

# SECTION 3
# STRATEGIC APPROACH

The strategic approach is the philosophy and doctrine of the strategic plan. It provides the framework and foundation upon which the parts of the plan are built. NOAA recognizes that meeting its goals will require a studied, visionary approach. End-to-end data and information management must be thoroughly understood and its implementation carefully planned. This is not simply a project of automation. The ESDIM Program must identify how and why the Line and Program Offices currently conduct their data handling functions and what needs to be done to meet the requirements of the future. The ESDIM Program must coordinate a NOAA-wide approach to meeting the agency's mission and determine what existing environmental services should be retained and what new ones will be needed to carry out that mission. Building upon existing resources where possible, the ESDIM Program will identify where operational efficiency can be enhanced through new technology or through innovative approaches to the task of environmental data and information management. The ESDIM Program will promote a modern approach to systems, operations, and services that will meet the needs of NOAA's broad spectrum of users in the 1990s and beyond.

This section first establishes the context of the ESDIM Program, its scope, and its ground rules, then identifies the key elements of the plan and its objectives. The importance of NOAA-wide participation in Earth system data and information management, along with the overall integration and end-to-end view, is highlighted. Use of existing NOAA capabilities is emphasized. The important role of continuing education and a draft NOAA data policy is included. Finally, the ESDIM Program's functional and technical requirements are summarized.

## 3.1 NOAA-Wide Participation

The NOAA-wide approach is mandated because NOAA consists of numerous offices and laboratories each carrying out many programs and projects, where diverse phenomena are being observed, recorded, researched, and archived. These activities result in numerous products and services, and contribute to a vast environmental data base. Well-managed and reliable data are essential to develop and maintain useful environmental data bases and disseminate environmental information products such as:

- Severe storm and flood warnings and weather forecasts
- Charts of U.S. waters and airspace

- River flow and water resource forecasts
- Solar and space environment forecasts
- Climate change prediction
- Ocean and coastal analyses and assessments
- Status of living marine resource stock reports
- Health of ecosystems
- Quality of the marine environment assessments
- Living and non-living marine resources conservation plans
- Habitat and endangered species protection plans
- Tide and current predictions.

While the ESDIM Program is not specifically tasked with the operational data management concerns of these activities, it will address the need for near-real time and retrospective access to these data and products. A summary of NOAA's many products and services is presented in Appendix B. A sample of the key environmental variables may be seen in Appendix C, illustrating the broad spectrum of NOAA's data base. Figure 3-1 summarizes the multidisciplinary data base categories in NOAA's major programs and illustrates the broad scope of coordination and interoperability requirements.

A primary objective of the ESDIM Program is to provide coordination and build consensus for data and information management throughout NOAA. Many benefits are gained by coordinating the tasks. First, there would be minimal duplication of effort; most tasks can be shared throughout the various components of NOAA, drawing upon existing capabilities and talent. Once a task is completed, its products can be used by all. Second, coordination promotes increased interoperability. By sharing standards, formats, and procedures, the system as a whole will become more efficient and data exchange will become easier. Procedures and system components that conform to agreed-upon, NOAA-wide standards are beneficial because they eliminate all or part of the overhead associated with unnecessary conversions of the data or loss of time due to difficulties in adapting to ever-changing formats and protocols. Finally, by adopting a NOAA-wide approach, a universal data management system may emerge allowing NOAA to become the national leader of interagency environmental data management efforts.

| Major Program Users \ Data Base Category | Atmospheric | Oceanographic | Marine Ecosystems | Land Surface | Snow & Ice |
|---|---|---|---|---|---|
| Marine Resources | | X | X | | X |
| Mesoscale Studies | X | X | | X | X |
| Atmospheric Research | X | X | | X | X |
| Oceanographic Research | X | X | X | | X |
| Atmospheric Services | X | X | | X | X |
| Ocean Services | X | X | X | | X |
| Coastal Ocean | X | X | X | | X |
| Global Change Research | X | X | X | X | X |

**Figure 3-1. Major Users of NOAA Data Bases**

## 3.2 END-TO-END VIEW

The ESDIM Program has adopted an end-to-end view of data and information management. This view includes the following elements:

- Quality assurance of data at each step of the data handling process, from the point of collection to the point of archive
- Documentation of data to the extent that the data are credible and useful for retrospective scientific research over the long term
- Archiving and preserving critical environmental data, derived data products, and associated documentation for selective access and use by the research community at large
- Promoting standards, systems, operations, expert services, and scientific oversight that facilitate the use of data and information by researchers and policymakers.

Note that there is a distinction between the data management roles of the Line and Program Offices and the ESDIM Program. A Line Office acquires and manages data to accomplish the mission of that office, whether operational or research-oriented. Line Offices generally are not responsible for providing mission-oriented data for retrospective use by other groups.

Scientific Program Offices have specific research responsibilities and acquire and handle selected data accordingly. The ESDIM Program activities overlap those of the other NOAA elements only to the extent that is required to generate a NOAA-wide data management focus and ensure the long-term usefulness of the data and information generated. This requires a high level of monitoring and coordination which ensures that the end-to-end data management objectives are met. The ESDIM Program will provide guidelines and resources, as necessary, to make the end-to-end practice of data management an integral part of every observation, collection, analysis, and research activity in NOAA.

The end-to-end view is an important concept in data management: it reduces overlaps and duplication of effort, and it ensures that data management is coordinated and optimized over the entire process to achieve the desired quality, processing efficiency, and user accessibility. This end-to-end view is illustrated* in figure 3-2.

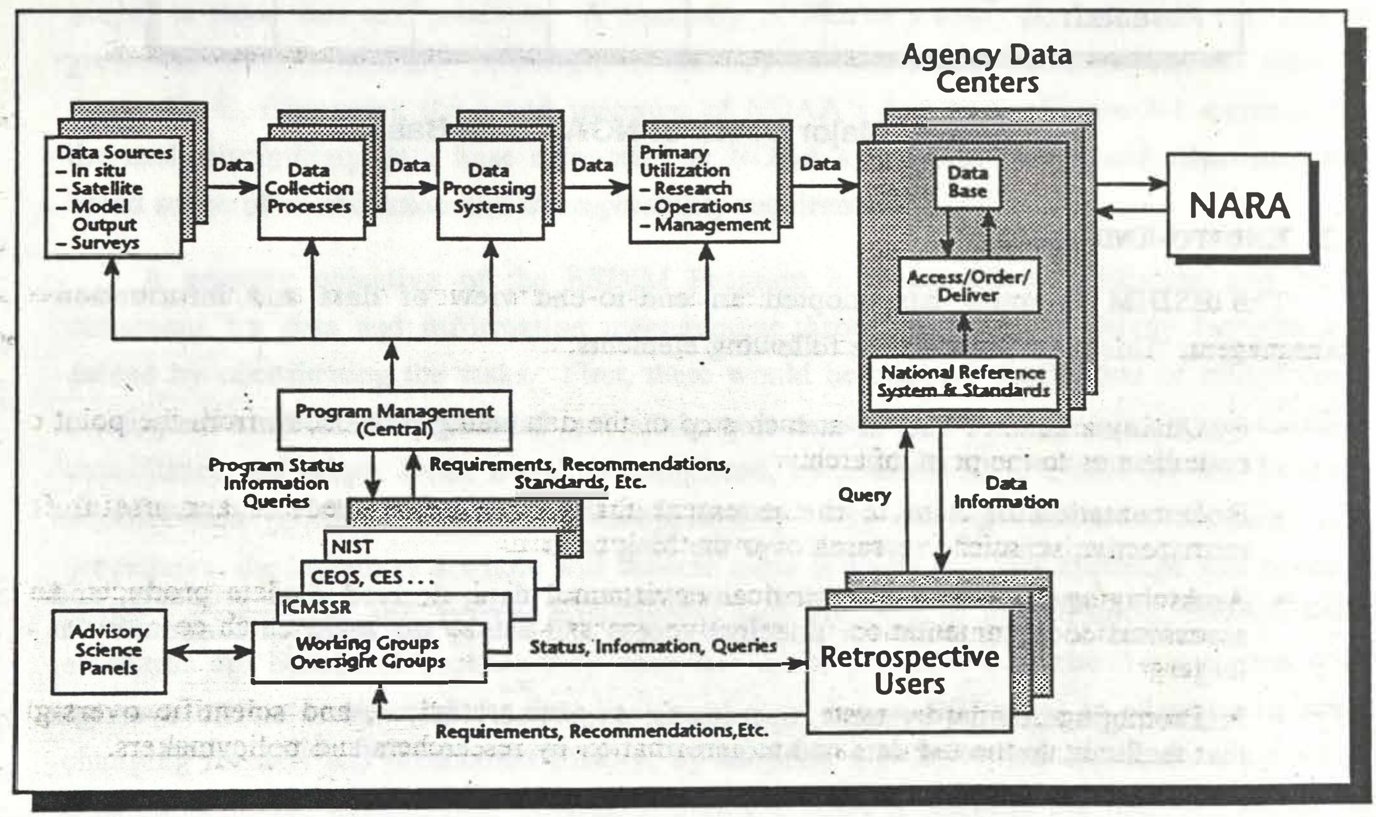



**Figure 3-2. End-to-End View of Data and Information Management**

---

* Patterned after an approach described in the referenced NOAA Federal Plan for Meteorological Information Management.

The activities of both old and new processes and procedures need to be assessed and optimized using an end-to-end view. For example, data collection activities should be planned with all steps in mind, including the final step of data dissemination. How will data quality control be conducted? Who will be the user of those data? How much data is optimum from the user and the data center perspectives? What sensor is optimal? What communications capabilities will be used? What processing will be required? What archiving space and storage media are anticipated? Questions like these need to be answered at the planning stage. Committees, some temporary and some permanent, will be assembled to provide guidance on issues pertaining to the end-to-end view. The ESDIM Program is already bringing together data managers and users from both inside and outside of NOAA to identify the most pressing needs. Science working groups are evaluating the current condition of and recommending improvements for selected large data sets. National Data Center review panels are supporting the planning for Center upgrades and recommending priorities based on scientific needs. NOAA-wide studies are underway to enhance data documentation, to develop prototype data systems, and to identify the requirements for a modern end-to-end data system. Many of these committees will provide long-term support to evolutionary program needs, and new ones will be convened as necessary.

### 3.3 BUILD ON EXISTING CAPABILITIES

The approach to effective management of Earth system data and information should rely heavily on interfacing with existing NOAA capabilities. Qualified NOAA personnel are already in place to perform all tasks from planning, data collection, processing, and archiving to data dissemination. There is also a wealth of existing facilities, data collection instrumentation, platforms, buoys, ships, airplanes, satellites, observation stations, data processing facilities, and archives. Integrating existing NOAA resources in a new data management approach may seem a logical thing to do, but it is more than sheer pragmatism; it is the involvement of the NOAA team sharing expertise and broad capabilities which have evolved from more than a century of making environmental observations.

In carrying out its mission, many NOAA elements acquire and hold environmental data that are not archived at recognized Data Centers. NOAA refers to these holdings as Centers of Data (see Appendix A). A Center of Data may become a recognized data center by implementing the archival procedures specified by NOAA and the National Archives and Records Administration (NARA). Some centers may adopt active data management programs to meet NOAA-wide quality and format standards, and pass selected holdings elsewhere for archiving.

The ESDIM Program shall try to retain as much as possible of the existing and near-term, agency-wide data management capabilities and structure by molding them into a broad, coordinated system. Re-engineering and automation will be considered where possible to render a smoother functioning and more efficient organization. Particular attention will be paid to repackaging the holdings at the Centers of Data, laboratories, and individual collections for inclusion at more accessible central archives.

### 3.4 ESTABLISH DATA POLICY

A comprehensive and coherent NOAA data policy is imperative to implement the ESDIM Program. There are existing data policies to consider; for example, the national policy issued July 2, 1991, by the Office of Science and Technology Policy (OSTP), "Data Management for Global Change Research Policy Statement." Typical of most data management policies, this document is tailored to a specific program area, in this case, the U.S. GCRP. It is much too narrow in scope to encompass the varied NOAA-wide programs. The needs of the full spectrum of users, from the general public to research investigators, are detailed in the draft NOAA policy* that follows:

> Unclassified and/or unrestricted environmental data and information produced, sponsored, collected, or obtained (by domestic or foreign exchange, purchase, or gift) by NOAA or other Federal or Federally supported activities are public property. It is, therefore, the policy of NOAA to make available those worldwide environmental data and information under NOAA's stewardship on the basis of exchange, loan, cost of reproduction, or free depending on the specific limitations described below. NOAA will continue to endorse a policy of free (full) and open exchange of environmental data with the international community.
>
> 1) All data and information products in NOAA not subject to confidentiality regulations shall readily be provided to [internal] NOAA users at the cost of reproduction when costs are extensive, otherwise free whenever possible or when special arrangements are made.
> 2) NOAA will provide data to researchers from other Federal and non-Federal groups for the cost of reproduction.
> 3) Research or experimental data needed to support time-critical NOAA operations shall be made available for those operations in real-time or non-real-time

---

* Interim NOAA policy on management of environmental data and information, effective from February 1, 1992, for one year.

depending on the requirement. Time critical operations include warning and forecasting operations, hazardous materials response activities, and others as defined by the Under Secretary.

4) New NOAA programs in their initial planning stage which focus on needs for either operational or episodic measurements shall ensure that the requirements of all NOAA Program and Line Offices are considered. NOAA Program Managers of replacement systems shall give strong consideration of the impact of the new system on data continuity and requirements for calibration or comparison with the old system. NOAA shall institute a review process through the NOAA Program Development Board (PDB) in the event of unresolved issues or requirements. For major systems acquisitions, any exception to including validated data requirements will be outlined in a decision memorandum to the Deputy Under Secretary before full-scale implementation begins.

5) Data sets may be held in temporary or specialized data centers (for indefinite periods of time) prior to submitting them to the NOAA Data Centers. However, some of the same data management requirements as imposed on the National Data Centers will apply to these specialized data centers or Centers of Data; i.e., all NOAA Program Managers shall ensure that the following conditions are met: data integrity and appropriate metadata are maintained; all other users are provided access in a timely manner; the data's existence is documented in the NOAA Earth System Data Directory and the proper transfer of the data to a designated national data archive is completed prior to the demise of the temporary or specialized data center.

6) NOAA components may produce and use supporting data sets which need not be placed in data management centers; e.g., flight recordings, photos, voice recordings, etc. Such data sets are to be documented in the NOAA Earth System Data Directory.

This draft policy will be reviewed by the NOAA Program Development Board and the ESDIM Program team to formulate it to the needs of the entire user community, as well as NOAA organizations.

### 3.5 PROMOTE EDUCATION

In a changing operating environment and organizational structure, continuing education is necessary to keep current with the various organizational and material changes being imposed on the system. Education programs are needed within and outside NOAA to promulgate the merits of this undertaking. Internally, an overview of the changes in the organization in which a particular unit is a part should be given, emphasizing NOAA-wide goals and purposes. The end-to-end view of data and information management will require somewhat of a cultural change in NOAA. This will necessitate interprogram and interoffice education seminars and publications.

Externally, there are three targets for education, First, there are those organizations that contribute to environmental data and information collection, processing, communications, or archiving. Second, general science users, policymakers, and decisionmakers must be advised as to what is available and how to access it. Third, the general public should be informed as to what is being accomplished in NOAA, the value of NOAA's environmental data as a national resource, and the direction that NOAA will take in the future. The importance of Earth system data and information should be reiterated programmatically to educational institutions and policymakers on all levels, and guidance should be provided on how to beneficially use environmental data and products.

### 3.6 Develop Functional and Technical Requirements

Identifying a comprehensive set of requirements is the first and most important step of any systems development process. The ESDIM Program will develop the functional requirements for data and information system modernization. Initially, functional requirements should be identified to provide top-level guidance. The principal system functional requirements include:

- Storage facilities sufficient to archive data to NARA standards well into the next century
- Processing capabilities sufficient to efficiently handle the increasing flood of incoming data
- A distributed workstation environment
- Priority data available online
- Online capabilities to facilitate data access, browse, and visualization
- Comprehensive data directories, catalogs and data set inventories to improve access
- Systematic data base preservation, maintenance and continuity
- Interoperability and interconnectivity of data centers
- Consistent and systematic quality control and data documentation procedures
- Mechanisms for user feedback and involvement
- Improved user services: integration, analysis, ordering, accounting, and distribution.

While wholly applicable to the recognized Data Centers, most of these functional requirements apply to the Centers of Data as well, where data management is complicated since the primary focus of these organizations is to conduct research programs or carry out certain operational functions. Processing and storage of retrospective data is a collateral duty that is important, but of secondary priority. The ESDIM Program will hold workshops to explore the many ways Centers of Data can be fully integrated into the NOAA-wide data management efforts.

A preliminary requirements study has been completed for the NOAA National Data Centers (Rotar, 1991). While these centers are only part of the future data complex, they do represent the vast majority of holdings and a principal focus for system modernization. They, therefore, are very representative of the long-term technical requirements that NOAA must target. The archive growth projections have been discussed previously (see figure 1-4); as the requirements study points out, however, depending upon the timing and output of major programs like NEXRAD, EOS, and other planned environmental satellites, the total archive could approach 1000 terabytes by the year 2000. Figure 3-3 indicates the projection of required computing capacity, including the millions of instructions per second, input/output speeds, and hard-disk storage. Finally, figure 3-4 indicates the telecommunications network data rates expected for various operations. These are estimates, based upon some assumptions that can not be fully detailed here, that indicate trends.

The ESDIM Program will provide the overall system integration activities and incremental improvements in system connectivity, data set production facilities, metadata assimilation, and online inventories. The resulting system will increase the capacity for the assimilation of new data streams. The system will provide for consistent data quality assurance and for the capture of metadata describing the data assets. It will provide the primary user interface permitting intelligent data searches, data and information browse, and online retrieval. Communications paths will be provided to link the full variety of multi-disciplinary scientific data bases such as atmospheric, geophysical, oceanographic, biological, and geodetic. Network hardware and local area network (LAN) software will be closely coordinated with DOC's Information Resources Management (IRM) staff to ensure compatibility and adherence to any established NOAA and DOC network and data base design standards.

There are many other NOAA-wide requirements within the ESDIM Program domain, such as the establishment of mechanisms and procedures for interagency and international cooperation, coordination, and data exchange. Numerous organizations, national and international, have made varying degrees of progress in efforts to improve their data and information management. Interaction and exchange with these organizations will seek to develop universal products and to ease the exchange of data and information between organizations through increased compatibility.

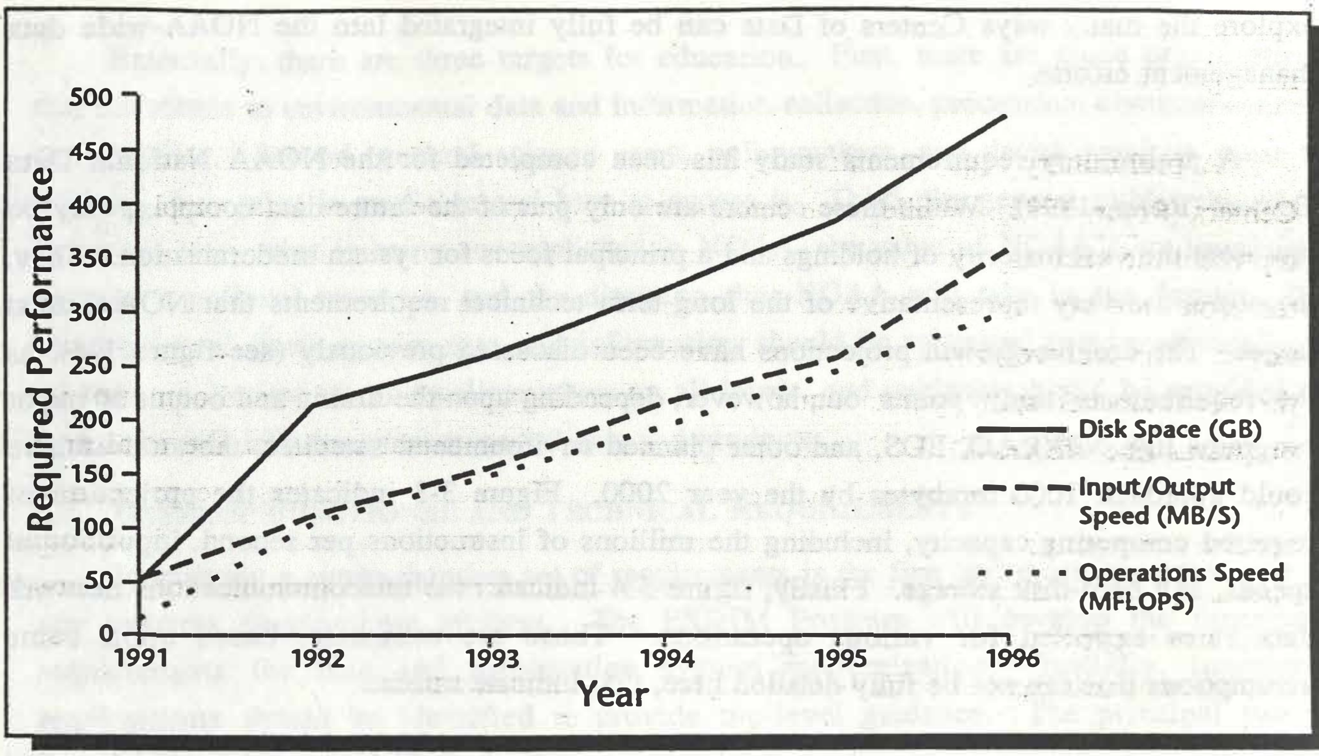



**Figure 3-3. Computing Capacity Projection (Ref P. Rotar)**

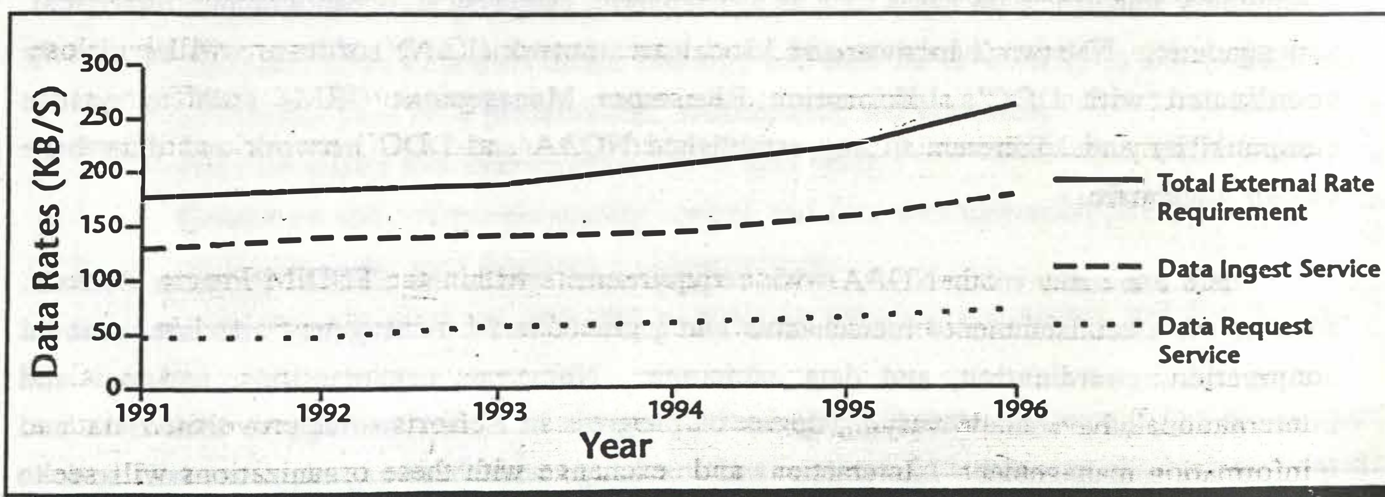



**Figure 3-4. Effective External Network Data Rates (Ref P. Rotar)**

## SECTION 4
## IMPLEMENTATION STRATEGY

An implementation strategy identifies the means by which a program is carried out. While the previous sections describe what needs to be done, this section discusses how to do it. First, the activities associated with a continuing program for data rescue and preservation are identified. Next, the program for a NOAA-wide data system modernization is outlined. Then, the principal methods to improve science quality are discussed. These include science user participation in the development of user access activities and the methods applicable to data quality control. Finally, technical approaches to achieve system interoperability, use technology insertion, and promote data exchange are identified. Implementations will be carried out mainly by the Line and Program Offices. The ESDIM Program will monitor and coordinate implementations to ensure that agency-wide needs are met.

### 4.1 DATA RESCUE AND RESTORATION

As an initial step to a long-term program for data preservation, certain data sets that are at risk must be rescued immediately. Rescue and restoration activities may be conducted in a number of phases:

- Define the "at-risk" criteria for data sets or portions of them
- Identify all at-risk data sets
- Determine condition of each data set
- Prioritize restoration work
- Begin rescue/restoration and archive in proper retention level.

For the definition of at-risk criteria, the following have to be considered: the location of the data; the storage conditions; the type, age, and condition of the storage medium; and the existence and availability of knowledge and expertise about the data set contents. Identifying the at-risk data sets will require a substantial effort, primarily because so many data sets are kept at Line Offices in the hands of experts who collected them for a specific study or purpose. The demand for retrospective use of such data sets may be low, and often the data sets are kept on paper medium in file cabinets or drawers. Here, the risk of such data sets being lost is not only due to deterioration of media but also to career termination or a personal decision to discard the data. Appendix D gives a partial list of data sets that have been identified as at risk of being lost.

The condition of each data set has to be determined. Ideally, some form of scoring system can be devised that will enable the labeling of a data set with a risk grade. This grading will give the at-risk definition a quantitative value that will cover a spectrum of conditions. Using the condition of the data set (its risk grade), together with information about demand for the data set, will enable the prioritization of the restoration work. It is clear that grading and prioritizing data sets for rescue is very subjective. Therefore, this must be done with the involvement of broadly representative user committees. Not all data will survive because of limitations in funding and time, so very conscientious decisions must be made.

The technical approach to rescue and restoration also has to be determined. There are many types of storage media on which data sets reside. Some data may need only simple backup from one magnetic tape to another; others may be transferred from magnetic reels to cassettes. Data on paper as tables or graphs, film, and microfiche require that a digitization step precedes any others. For some digital media such as punched cards or punched strips of paper, the proper reading and transfer apparatus must be in place. There will be many manual steps required to move and mount materials, so labor cost and time must be included in estimating the work flow for each rescue project.

Restoration work may then begin according to priority lists using NARA-approved storage media for the newly rescued data. Schedules and milestones must be set for the restoration work to monitor the progress. After the rescue effort is complete for a given data base, the data will be archived in a selected center according to their predetermined retention level. It is anticipated that rescue and restoration will be a continuing activity. The strategy is to conduct a concerted effort early in the program and maintain some level of effort every year thereafter.

### 4.2 MODERNIZATION

The NOAA-wide data system modernization will provide NOAA with long-term, systematic growth of its scientific data and information management resources to meet the strategic objectives. The fundamental objective is to develop and implement an end-to-end data system that effectively integrates NOAA's Line and Program Office data management activities. This system will build upon and complement existing and near-term, agency-wide data management capabilities. The ESDIM Program will supply resources as needed to develop those system aspects that are NOAA-wide. It will support the overall system engineering and integration activities, will implement incremental improvements in system connectivity, data set production facilities, metadata assimilation, online inventories and data requests, improve data protection, and increase capacity for the assimilation of new data streams. This development will lay the foundation for NOAA to meet its long-term responsibilities, which include permanently archiving the observations for the EOS program. A systematic approach to archive maintenance will be established.

The NOAA data and information system modernization initiative is envisioned to be a multiyear, phased effort. The critical planning, integration, and design phase will commence in FY 1993. The principal objectives of the initiative are to:

- Upgrade the data management system to deal with impending large incoming volumes of data from NOAA observing platforms
- Modernize user-oriented data and information system services
- Provide data gateways between NOAA and other agencies
- Ensure environmental data continuity in NOAA's observational programs.

To the greatest extent possible, competitive procurements are anticipated to provide the best available hardware, software, and integrated systems. The program will employ a building block approach in which the development of key elements will begin with proof-of-principle prototypes, concept demonstrations, and tests.

Various activities are already underway which lay the groundwork for this initiative. These include:

- Preparing a NOAA Earth System Data Directory that identifies, locates, and describes environmental data sets available from all NOAA components
- Developing Pathfinder data sets from four extensive satellite data sets, and preliminary studies of future large data volume producing programs, to gain experience in handling, archiving, and distributing very large data sets
- NOAA-wide data system prototypes such as those to be developed under the Marine Ecological Data System Project, and the Prototype Distributed Environmental Data Management System Project, to gain experience in integrating data and developing interoperable data directories, catalogs, and inventory systems
- Preliminary studies of requirements and alternative approaches to NOAA systems modernization.

The data system modernization requires that the staff involved throughout NOAA in the end-to-end data and information management process have the training and possess the skill mix necessary to conduct the operations and provide the services expected. Training and education programs are anticipated to ensure that NOAA, and ultimately the system users, will gain the maximum benefit from new technologies and procedures. Along with the anticipated increased efficiencies due to automation and the application of advanced technologies, skilled human resources will be added, as necessary, to meet increased user demands.

Some of the most important issues which must be considered in conducting the modernization initiative are discussed in the remainder of this section.

### 4.3 SCIENCE USER PARTICIPATION

Inclusion of science users in planning committees is a necessary part of the strategy. Knowing the demands and desires of the customer *a priori* will, in many cases, streamline the process and focus the efforts and outcomes. Such streamlining will eliminate redundant processing and parameters, and unnecessary resolution, precision, and sensitivity, thereby making the system more responsive to the needs of the user, more efficient, and more economical. The ESDIM Program will bear in mind that user needs and demands are continuously changing in response to contemporary subjects of study and needs for quality data. These needs simply cannot be projected decades in advance. Thus, fostering user participation will be an ongoing quest.

There are over 600 data bases incorporating over 2000 environmental variables that NOAA collects, observes, records, processes, documents, quality-controls, archives, and distributes to the user. Each parameter in NOAA's archives is, to a degree, an outcome of this end-to-end data handling process. Each of these processes—at times, few parameters share a single process—must be reviewed and optimized. Such process reviews must be conducted by committees with the proper mix of people, knowledgeable and experienced in the steps of the process. Ultimately, the strategy is to standardize the steps NOAA-wide to achieve common levels of formatting, quality, and user access.

Seeking user feedback may be the most effective tool available to management to ensure the usefulness of program results. Feedback mechanisms that can be used to strengthen NOAA systems, operations, and services must be built into the end-to-end data and information management process. Procedures to evaluate and act upon such feedback must be included to ensure that the systems and services will evolve and are continually improved. Therefore, the strategy calls for a continuing outreach program to include users on planning committees, monitor user satisfaction with the evolving systems and services, and encourage users to contribute improvements to data and information from their work.

## 4.4 Quality Control

Quality control is the procedure by which an organization ensures the integrity of a product or process. This must be a high priority for an organization whose ultimate mission is to disseminate credible and usable information. It ensures that products are uniform or conform to certain standards. This, in turn, builds user confidence and familiarity with the product. Quality control will require several things. First, random noise or extraordinary spikes, systematic errors, and biases must be identified in incoming data. Second, a data set must have extensive documentation fully describing the various attributes of the set, in the form of metadata. In the data and information domain, the attributes of a given data set include: sensor precision, sensitivity and response range, sampling rate, spatial coverage, station history, mean, variance, range, noise, errors, gaps, discontinuities, and other data documentation (metadata). To use these data correctly, a user must have these attributes available. To the extent possible, the quality assurance process should be automated using expert systems. Scientists should be involved for specialized interpretation and fine-tuning the knowledge-based controls.

As a NOAA-wide strategy, archived data sets must ultimately possess a uniformly high quality. The data users are the main judges of what level of quality is required for given applications. Therefore, the ESDIM Program must sponsor a committee consisting of representative users and NOAA data managers to develop formats, standards, and procedures for establishing and implementing quality levels for data. NOAA must produce selected "science quality" data sets for broad, multidisciplinary research efforts to fulfill its mission. The insight from this NOAA-wide program will be used to develop a long-term strategy for both end-to-end quality control for new data and to address the issues of enhancing the quality of currently archived data sets. Implementing these controls will produce science quality data that satisfy the aforementioned 20-year test.

## 4.5 User Access

User access to data and information is one of the strategic objectives of the ESDIM Program. In a system where data are to be distributed to retrospective users at minimum cost of reproduction, one might say that a given data set is inaccessible only if it is hopelessly lost. If it exists on some known medium, then it is theoretically accessible, but may be accessible only to a chosen few who know of its existence and worth, and have the means to retrieve it. However, this is not acceptable access for most users. The degree of accessibility can be measured by the efficiency with which access is gained to data and information on different storage media, using different hardware and procedures. Thus, increased access efficiency becomes a measure of the effectiveness of this program.

To achieve an increase in access efficiency, new, powerful data base management tools must be developed. Standards and guidelines for metadata, directories, and catalogs need to be developed to ensure uniformity. The introduction of these products, along with the data system modernization, will substantially improve the access to NOAA's data holdings. The long-term strategy is to produce the equivalent of an "electronic librarian." By employing a fully distributed architecture, data identification, evaluation, and ordering can all be accomplished electronically from the user's desktop terminal. The user sees a virtually uniform system of indistinguishable data sources. This is illustrated in the architecture shown in figure 4-1.

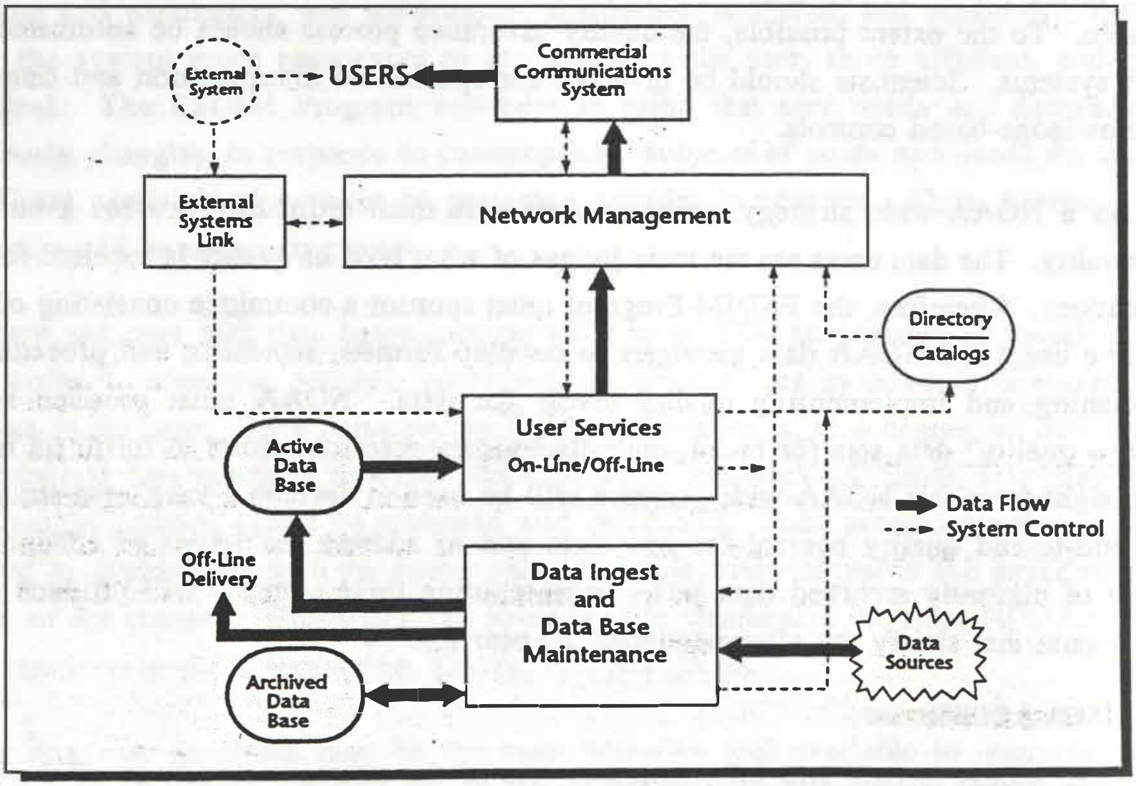



**Figure 4-1. Top-Level System Architecture**

The User Services element shown in this architecture includes a powerful and robust data base management system to support user queries. The data base architecture should permit extensive multilevel browsing. This is particularly important with large volumes of data anticipated in future measurement programs. The browsing capability makes it possible to make full use of the metadata files, directories, and catalogs. Data visualization will also be an important tool to assist users in the selection of data. Accounting, ordering, and distribution for product requests should be fast and simple.

Rapid, online, interactive access to information is especially relevant to policymakers and the general public. Data to support these activities are discipline and sub-discipline specific. These data may be large-scale synoptic descriptions of a single physical, chemical, or biological parameter, or temporal variations of the parameter. The transformation of such data into information is a process whereby these apparently disparate and often seemingly conflicting data are synthesized into a coherent description of the state of the environment. The modernization of decision support through enhanced environmental information synthesis and visualization is key to realizing the value of the Nation's data assets. The modernization initiative will address options and plans for a data system for policymakers and decision-makers through concept demonstration plans. Guidance will be sought from government decisionmakers in the design, implementation, and evaluation of this activity.

### 4.6 OPEN SYSTEM APPROACH AND TECHNOLOGY INSERTION

An "open system" possesses the quality of standardization and is, therefore, open to changes, extensions, and expansions. This quality is a thread throughout the various system levels promoting interoperability among hardware and software interfaces. This is important because of rapidly changing technology and the anticipated increasing number of users to be connected. In the design of any data handling system today, anticipation of new technology developments is necessary. The exact details of future developments are difficult to foresee, but the general trends should be considered. Increased data ingest rate and volume, and increase in demand for quality data are some examples of readily discernible trends. A system architecture ought to be sufficiently flexible to be able to take advantage of such growth trends.

Open systems have come to mean specific kinds of computer system architectures such as LAN-based distributed processing, or layered architectures conforming to the OSI 7-layer model. As valuable as these architectures are, a broader definition of an open system is needed for systems with expected lifetimes of 30 years or more, as will be the case for some NOAA data systems. This kind of system may have to survive several revolutions in technology, new generations of standards, and changing fashions in system design. One broad definition is that an open system is one in which the function or application is independent of the implementation; that is, the functionality can be "transported," such as via computer software source code, between different sets of hardware. This is important for making the system independent of particular manufacturers, but also independent of the technology at any moment in time. Hardware obsolescence is one key driver forcing the procurement of new systems, a problem that open architectures can alleviate through incremental upgrades and periodic insertion of modern technology. This mitigates service disruptions to users. Another important benefit of transportable functionality is that it preserves investment in software development, which is often the single most expensive item in the development of a new system. Accordingly, software transportability is probably the most basic aspect of an open system design, permitting performance enhancements at relatively small incremental costs.

To deal properly with the expected increases in data volume, user demand for environmental data, and the complexity and variety of scientific applications, new technology must be liberally included in the NOAA-wide data system modernization. The strategy is that all new data systems will be designed based on modular and open systems architectures to permit cost-effective system evolution using periodic technology insertion.

## 4.7 National and International Data Exchange

Data exchange with external organizations is necessary to support a comprehensive ESDIM Program. Such exchange is primarily between counterpart data centers. Nationally, there are a number of environmental data collection efforts in various agencies, e.g., NASA, NSF, DOE, EPA, USGS, USDA, and in the military and academic institutions. Figure 4-2 illustrates an interagency data and information model served by interoperable catalogs and user access capabilities. Internationally, there are similar organizations whose data must be considered for data exchange programs. The ESDIM Program will participate in the development of data standards and international directories. It will coordinate its activities through existing responsibilities for World Data Centers, international environmental data collection programs, and existing protocols for cooperation. Facilitating data exchange is essential to supporting global environmental research.

To form a viable data exchange program, standards must be set for the various phases of exchange. Important issues such as data formats, documentation (metadata files), quality control, fees, communication, and data transfer channels must be addressed. The ESDIM Program will institute systems engineering activities for standards development that conform to the Federal Information Processing Standards. These standards, developed under the

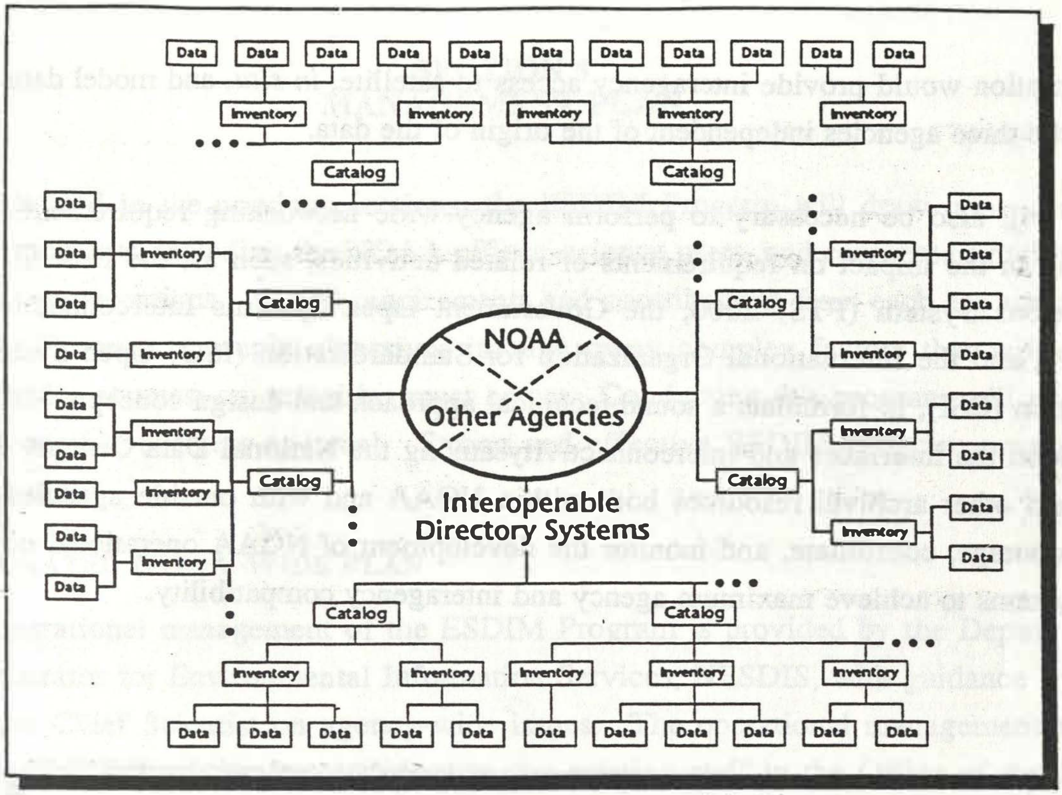



**Figure 4-2. Interagency Data and Information Exchange**

direction of the National Institute of Standards and Technology (NIST), provide a structure for improving the use and management of computers and automatic data processing systems. Similar standards will be investigated to guide the development planning for the interconnectivity interface requirements. The NOAA data system modernization will provide integrated, efficient environmental data and information services through NOAA-wide distributed capabilities. Through National Data Centers, the National Environmental Satellite, Data, and Information Service (NESDIS) is already actively involved with NIST in the development and promulgation of standards. As a strategy, ESDIM Program-developed standards will comply with national and international standards to the fullest extent possible.

The ESDIM modernization initiative envisions a 3-step program. Step 1 will provide the internal gateways necessary for its own disparate elements to exchange data. Step 2 will develop interoperability so that this access is easy. Step 3 will develop the building blocks to provide NOAA with access to the holdings of the other agencies and to provide other agencies access to NOAA's treasure of data. Planning has begun on a prototype development of a tri-agency data and information system for global change data. Participating agencies are NOAA, NASA, and USGS. The system concept is illustrated in figure 4-3. The successful demonstration would provide interagency access to satellite, *in situ*, and model data from any one of the three agencies independent of the origin of the data.

It will also be necessary to perform agency-wide networking requirements analyses, factoring in the impact on requirements of related activities, such as, the Federal Telecommunications System (FTS) 2000, the Government Open Systems Interconnection Profile (GOSIP), and the International Organization for Standardization (ISO) Open Systems Interconnection (OSI), to formulate a sound technical approach and design concepts. ESDIM will focus upon the interfaces and interconnectivity among the National Data Centers, Centers of Data, and other archival resources both within NOAA and with outside agencies. ESDIM will encourage, coordinate, and monitor the development of NOAA operational communications systems to achieve maximum agency and interagency compatibility.

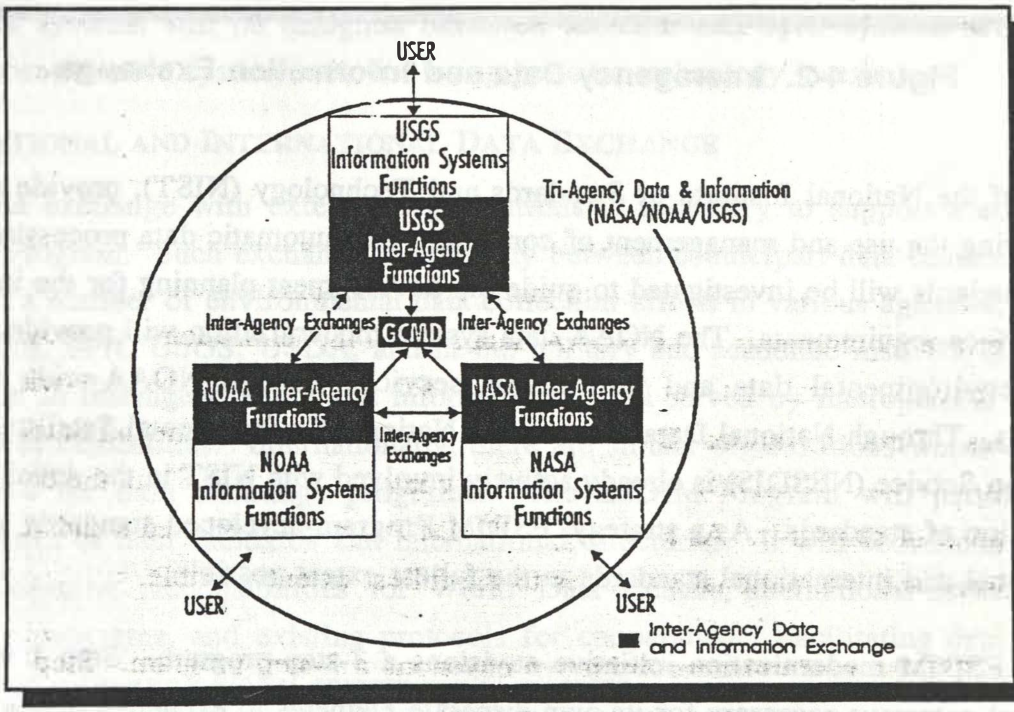



**Figure 4-3. The Tri-Agency Interoperable Data and Information Model**

## SECTION 5
## MANAGEMENT PLAN

As indicated in the previous sections, the ESDIM Program will draw upon a diverse group of participants, including the NOAA offices, science users, and counterpart national and international organizations. Varied requirements and contributions from each are anticipated. The ESDIM Program strategic objectives involve many complex factors that will vie for always limited resources; so, priorities must be set. Conducting this program will not, even remotely, resemble business-as-usual. Strong and effective ESDIM Program management will be essential.

### 5.1 INTEGRATED NOAA-WIDE PLAN

The operational management of the ESDIM Program is provided by the Deputy Assistant Administrator for Environmental Information Services, NESDIS, with guidance from the Office of the Chief Scientist on agency-wide issues. The operational management support will be provided through modest additions to the existing staff in the Office of the Deputy Assistant Administrator. As appropriate, senior technicians and other staff from each of the existing Data Centers may also be assigned to work with the ESDIM Program. Other NOAA Line Offices may also provide assistance on a temporary basis to deal with specific issues. Certain parts of the program dealing with communications and data systems acquisition will be closely coordinated with the DOC's IRM staff and Telecommunications and ADP Security Branch to ensure compatibility and compliance with any NOAA and DOC network and data base design and security standards.

The primary function of the ESDIM Program is to support the implementation of the long-range strategy that has been described. The view of the ESDIM Program is much broader than individual NOAA programs or projects that have more narrowly defined data management requirements. In response to this challenge, an integrated NOAA-wide approach to Earth system data and information management will be actively pursued. It will require active involvement of all NOAA Line and Program Offices.

Outside NOAA, the ESDIM Program will coordinate its computation and networking activities through the Committee on Physical, Mathematical, and Engineering Sciences (PMES) of the Federal Coordinating Council for Science, Engineering and Technology (FCCSET), in the President's Office of Science and Technology Policy (OSTP). A NOAA representative has been appointed to the PMES Committee's Working Group on High-Performance Computing, other members of which include representatives from DOE, DOI, EPA, NASA, NSF, and USDA. The charter of FCCSET/PMES includes all aspects of information science and technology, including supercomputing, networking, and visualization, all of which are important to successful management of the very large data bases with which NOAA is involved.

Another important external committee is the Interagency Working Group on Data Management for Global Change (IWGDMGC), which has already established important mechanisms for interagency coordination on data management topics such as data policy, standards compliance and interoperability for a Global Change Data and Information System (GCDIS). This working group reports to the Subcommittee on Global Change, a part of the Committee on Earth and Environmental Sciences (CEES); CEES is also a part of FCCSET. The ESDIM Program interacts with this working group through the NOAA representative to IWGDMGC to ensure that ESDIM Program planning is coordinated with the working group's recommendations and activities relative to the U.S. GCRP.

The ESDIM Program also plays a major role on the Office of the Federal Coordinator for Meteorological Services and Supporting Research (OFCM) Working Group on Meteorological Information Management. Many important interface and integration issues involving common data formats, catalog interoperability, and telecommunications networks are addressed through these coordinating activities.

Internationally, the ESDIM Program must coordinate its activities with counterpart environmental science agencies to develop uniform data standards and formats. This cooperation can be led through NOAA's activities and responsibilities for operating World Data Centers and should include participation in the preparation of the International Master Directory. Other protocols must be established, as necessary. To effectively support global environmental research programs, the activities developed under the ESDIM Program should account for and seek to influence data and information management, not only NOAA-wide, but worldwide.

## 5.2 Management Structure

The ESDIM Program will be managed by the Deputy Assistant Administrator for Environmental Information Services, as shown in figure 5-1.

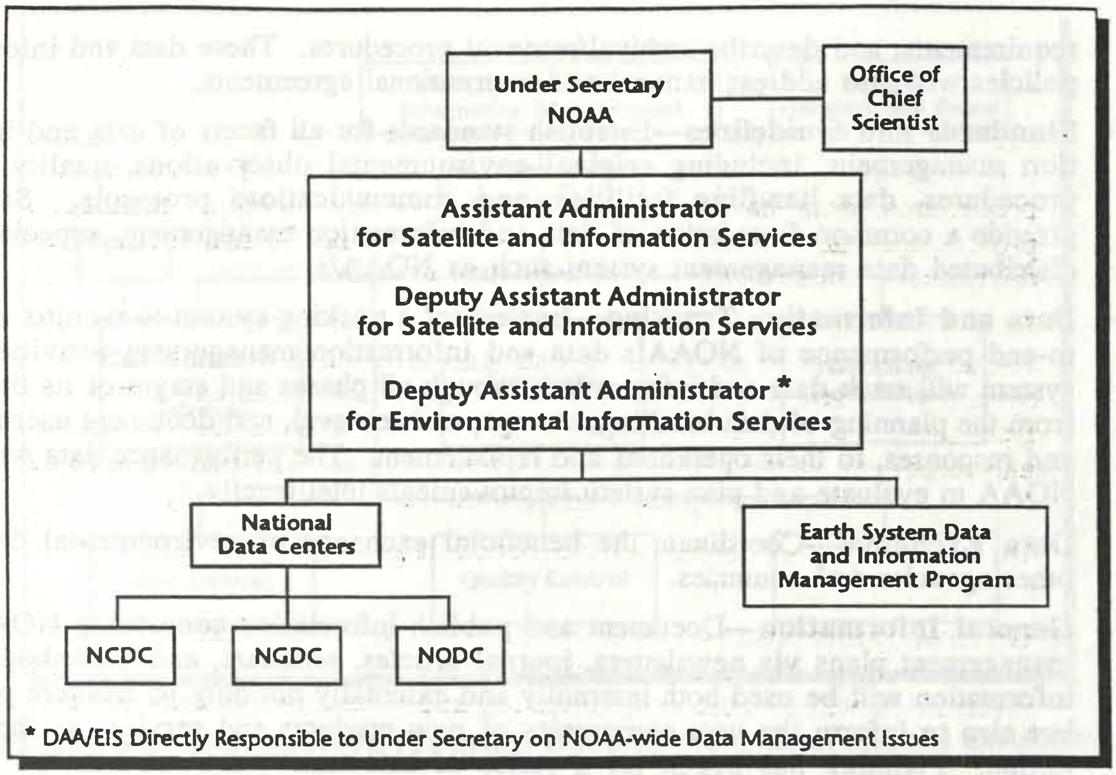



**Figure 5-1. NOAA-Wide Strategy Management Structure**

The primary responsibilities of the ESDIM Program are:

- **NOAA-Level Management**—Coordinate and integrate NOAA's data and information management activities. This includes the management and technical guidance to ensure the quality, continuity, and accessibility of NOAA data holdings.
- **Planning and Evaluation**—Involve representatives of the Line Offices and cross-cutting Program Offices in the preparation of management strategic plans, program development plans, implementation plans, and supporting documentation describing the NOAA-wide data and information management activities.
- **Oversight**—Monitor data and information management throughout NOAA for adherence to environmental data policy guidelines. Identify operations deficiencies and recommend corrective measures.
- **Data System Modernization**—Support the development and implementation of NOAA's data system modernization through the preparation of requirements analyses, concept and acquisition planning documentation, and configuration management activities. Establish the necessary infrastructure to support user access interconnectivity with the diverse NOAA environmental data and gateways to other agency and international archives.
- **Policy Development, National and International**—Develop and promulgate NOAA-wide policies needed to clearly establish data and information management responsibilities and procedures. These policies will establish data submission

requirements, and describe archival/retrieval procedures. These data and informatione policies will also address national and international agreements.e

- •e **Standards and Guidelines**—Establish standards for all facets of data and information management, including original environmental observations, quality controle procedures, data handling facilities, and communications protocols. Standardse provide a common foundation of data and information management, especially in ae distributed data management system such as NOAA's.e
- •e **Data and Information Tracking**—Implement a tracking system to monitor the end-to-end performance of NOAA's data and information management activities. Thee system will track data and information through all phases and stages of its life cycle, from the planning of data handling, storage, and retrieval, and document user requestse and responses, to their operations and replacement. The performance data will allowe NOAA to evaluate and plan system improvements intelligently.e
- •e **Data Exchange**—Coordinate the beneficial exchange of environmental data withe other agencies and countries.e
- •e **General Information**—Document and publish information concerning NOAA data management plans via newsletters, journal articles, seminars, and workshops. Thise information will be used both internally and externally not only to measure progress, but also to inform the user community of new products and services as they comee online. Planning has begun for a series of informative articles entitled "NOAAe Environmental Situation Reports," which will be provided to the general public on ae quarterly bases.e

## 5.3 Operational Structure

To ensure that the NOAA-wide strategy goals and objectives are met, the ESDIMe Program will establish a number of support teams, drawing upon specialists in major aspects of the program. These teams will provide technical and operational advice, assist in thee selection of program element priorities, and support overall system configuration managemente and quality assurance. A possible support team structure is illustrated in figure 5-2. Ase indicated by the figure, the ESDIM Program is assisted in its deliberations on the definition,e selection, and review of the program elements by the NOAA Program Development Boarde (PDB) comprised of Assistant Administrators and Program Directors. The PDB guidance ise implemented through the ESDIM Program team, which consists of representatives from eache Line Office and cross-cutting program, and through the supporting teams.e

## 5.4 Initial Milestones

The ESDIM Program strategy spans this decade and beyond. Activities ande implementations will be subject to annual budget funding approvals, and many midcoursee corrections are anticipated. e

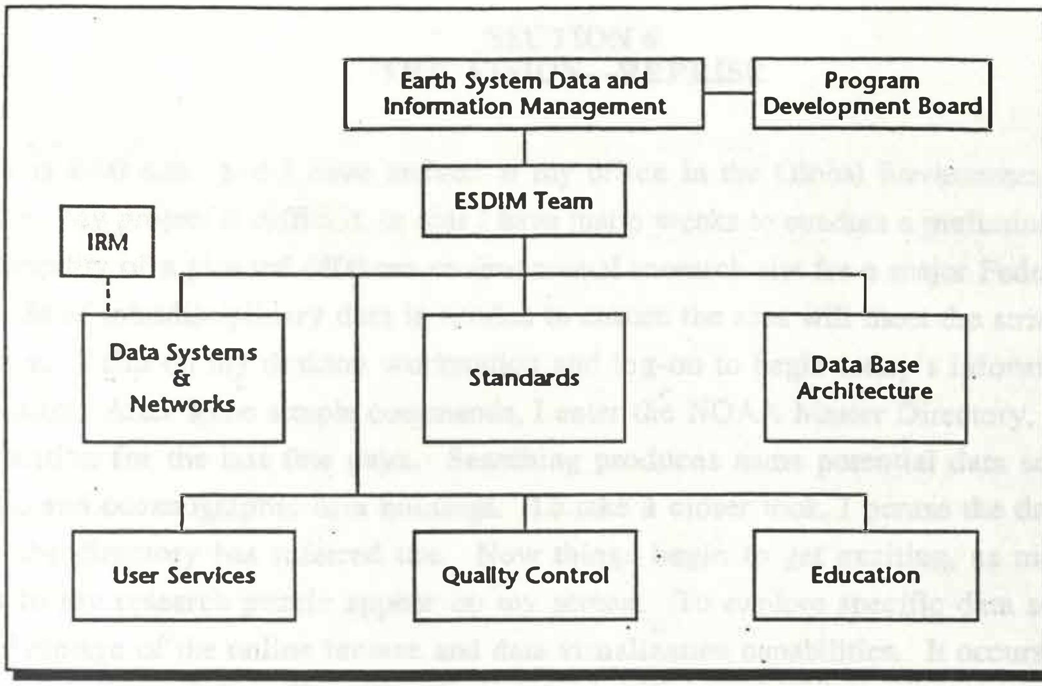



**Figure 5-2. NOAA-Wide Program Development Operational Structure**

Intense activity is expected in data rescue early in the program. NOAA-wide rescue and restoration projects were initiated in FY 1991 NOAA-wide, and will continue at a priority level for the foreseeable future. A summary of activities funded by the ESDIM Program is presented in Appendix E. Planning for the NOAA-wide data system modernization began in FY 1992 with the identification of requirements. Based on the huge volume of data anticipated by the late 1990s (particularly from EOS), preliminary studies of candidate technologies, data handling architectures, and prototype systems should begin immediately. Planning is underway on a proof-of-concept for intra- and inter-agency data gateways. Evaluation and planning has begun for demonstration and prototype, user-based, integrated data and information systems and data continuity projects.

Since each of the strategic objectives competes for funds with the others, and all have roughly equivalent urgency, priorities must be considered carefully. This requires that a funding strategy be developed that optimizes the return on NOAA's data and information management investments.

## SECTION 6
## THE VISION—REPRISE

It is 8:00 a.m., and I have arrived at my office in the Global Environmental Research Institute. My project is difficult, in that I have just 6 weeks to conduct a preliminary review of the suitability of a planned offshore environmental research site for a major Federal study. A multitude of interdisciplinary data is needed to ensure the area will meet the stringent criteria set down. I flip on my desktop workstation and log-on to begin today's information-gathering session. After some simple commands, I enter the NOAA Master Directory, the object of my attention for the last few days. Searching produces some potential data sources in the climatic and oceanographic data holdings. To take a closer look, I peruse the data catalog to which the directory has referred me. Now things begin to get exciting, as more potential pieces to my research puzzle appear on my screen. To explore specific data sets further, I take advantage of the online browse and data visualization capabilities. It occurs to me that I am not even sure of the storage location of the data I am examining, but they look extremely relevant to my project and are well documented, so I can see immediately how to use them. Therefore, I order the data, knowing that the reproduction and distribution fees, if any, will be automatically applied to my credit account. I continue the search for other data sets for my collection.

It is now 11:30 a.m., and I have had an extremely fruitful morning, having ordered four more data sets. Now, I have the rest of the day to study the CD-ROMs full of data that have already been delivered to me. Not bad for a project that is only 3 weeks old; and I thought being based in Provincetown, Massachusetts, would present problems to carrying out my research.

[This fictional scenario illustrates the ultimate achievement of the broad and ambitious data and information management program that has been described in this plan. It is not only possible, but probable, because it is driven by a solid need. NOAA and the ESDIM Program welcome this challenge.]

## REFERENCES


Federal Coordinating Council on Science, Engineering and Technology, Office of Science and Technology Policy, *Our Changing Planet: The FY 1990 Research Plan*, Committee on Earth Sciences, The Executive Office of the President, Washington, D.C., July 1989.

Federal Coordinating Council on Science, Engineering and Technology, Office of Science and Technology Policy, *Data Management for Global Change Research Policy Statements*, Executive Office of the President, Washington, D.C., July 1991.

Intergovernmental Panel on Climate Change, *Climate Change, The IPCC Scientific Assessment*, ed. J.T. Houghton, G.J. Jenkins, and J.J. Ephraums, Cambridge University Press, Cambridge, 1990.

Rotar, Paul, *NOAA Data Center Requirements Study*, National Center for Atmospheric Research, Boulder, Colorado, September, 1991.

U.S. Department of Commerce, National Oceanic and Atmospheric Administration, Earth System Data and Information Management Program, *FY 1993 NOAA-Wide Program Development Plan (Draft)*, Washington, D.C., September 1991.

U.S. Department of Commerce, National Oceanic and Atmospheric Administration, Earth System Data and Information Management Program, *NOAA-Wide Program Development Plan (Draft)*, Washington, D.C., February 1991.

U.S. Department of Commerce, National Oceanic and Atmospheric Administration, Earth System Data and Information Management Program, *Science Quality Data Sets*, Washington, D.C., 24 September 1991.

U.S. Department of Commerce, National Oceanic and Atmospheric Administration, Federal Coordinator for Meteorological Services and Supporting Research. *Federal Plan for Meteorological Information Management, Final Draft*, Washington, D.C., May 16, 1990.

U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Environmental Satellite Data and Information Service, Data Access Project Office, *NOAANET Functional Requirements*, Washington, D.C., August 1988.

U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Environmental Satellite Data and Information Service, *NESDIS Strategic Plan 1991-2000*, Draft, Washington, D.C., April 1991.

U.S. Department of Commerce, National Oceanic and Atmospheric Administration, *NOAA's Strategic Plan 1991-2000 An Agenda for Agency Advancement*, Clearance Draft, April 15, 1991, Washington, D.C., June 1991.

U.S. Senate, Report to the Chairman, Committee on Commerce, Science and Transportation, General Accounting Office, *Environmental Data; Major Effort is Needed to Improve NOAA's Data Management and Archiving*, GAO/IMTEC-91-11, Washington, D.C., November 1990.

U.S. Weather Research Program, *Strategic Plan for Environmental Information Management: A Vision for the Year 2000*, Unpublished Working Paper (D. McGuirk), 1991.

# APPENDIX A
# EXISTING FACILITIES

In a broad sense, NOAA is officially charged with maintaining environmental records for the Nation. In response to that charge, NOAA's National Environmental Satellite, Data, and Information Service (NESDIS) operates three National Data Centers, in which the bulk of the agency's data and information holdings are archived—the National Climatic Data Center, in Asheville, NC; the National Geophysical Data Center, in Boulder, CO; and the National Oceanographic Data Center, in Washington, DC. NOAA also maintains many Centers of Data that manage data for individual scientific specialties. These Centers of Data (examples are listed below) are managed by the individual NOAA Line Offices. Each Center of Data, such as in National Marine Fisheries Service (NMFS), may have several satellite laboratories that also manage significant amounts of distributed data. Examples of NMFS's 20 satellite laboratories include: Auke Bay, Alaska; Beaufort, North Carolina; Bay St. Louis, Mississippi; Naragansett, Rhode Island; and Honolulu, Hawaii.

| Title | Location |
|---|---|
| Bathymetric/Geodetic Data Base | Rockville, MD |
| National Meteorological Center | Camp Springs, MD |
| National Marine Fisheries Service Headquarters | Silver Spring, MD |
| NMFS Northeast Fisheries Science Center | Woods Hole, MA |
| NMFS Southeast Fisheries Science Center | Miami, FL |
| NMFS Southwest Fisheries Science Center | La Jolla, CA |
| NMFS Northwest Fisheries Science Center | Montlake, WA |
| NMFS Alaskan Fisheries Science Center | Seattle, WA |
| Joint Ice Center | Suitland, MD |
| Great Lakes Environmental Research Laboratory | Ann Arbor, MI |
| Ocean Applications Branch | Monterey, CA |
| Ocean Products Center | Camp Springs, MD |
| Equatorial Pacific Information Collection | Seattle, WA |
| Particle Deposition Air Resources Laboratory | Silver Spring, MD |
| University of Hawaii | Honolulu, HI |
| National Snow and Ice Data Center | Boulder, CO |
| National Tide and Water Level Data Base | Rockville, MD |
| Global Monitoring for Climate Change | Boulder, CO |
| Satellite Data Processing and Distribution/NESDIS | Suitland, MD |
| Pacific Marine Environmental Laboratory | Seattle, WA |
| Environmental Research Laboratory | Boulder, CO |
| Atlantic Oceanographic and Meteorological Laboratory | Miami, FL |
| NOAA Central Library | Rockville, MD |

# APPENDIX B
# NOAA PRODUCTS AND SERVICES
## (In a typical year)

## *National Ocean Service*

- produces 10,000 aeronautical charts
- issues 1,500 nautical charts and bathymetric maps
- provides 40,000 ocean observations, 500 sea temperature analyses and 300 ice analyses
- maintains coastal zone management partnerships with 29 states and territories
- manages 8 marine sanctuaries and 265,000 acres in estuarine reserves
- monitors 250 sites in coastal areas to assess environmental quality change
- responds to 100 hazardous materials spills
- enhances its national inventories of human activities and natural resources in estuarine, coastal, and oceanic areas, such as national inventories of 7,000 coastal pollution sources
- expands its national data base on characteristics of 122 estuaries

## *National Weather Service*

- issues 12,000 severe weather and flood warnings and prepares 2.5 million routine public weather forecasts
- produces forecasts for:
  - aviation weather: 1 million
  - fire weather: 70,000
  - agricultural weather: 40,000
  - marine advisories: 40,000
  - river forecasts: 700,000
- operates the most powerful computer available to model the changing state of the atmosphere 4 times a day
- receives nearly 20 million surface weather observations, 1 million ship observations, 1 million aircraft observations, and 1 1/2 million weather balloon soundings
- maintains a sophisticated tsunami monitoring system in the Pacific Basin

## *Office of Oceanic and Atmospheric Research*

- supports 80 major oceanic and atmospheric field experiments
- supports 11 environmental research laboratories
- publishes 1,000 scientific articles
- makes 150 scientific presentations
- maintains the National Sea Grant College Program with 200 institutions and 3,000 scientists
- supports 20 postdoctoral fellowships
- forecasts changes in the space environment
- maintains 4 global baseline monitoring stations

## *National Marine Fisheries Service*

- maintains the largest and most complex marine fisheries management system in the world
- oversees 2 million square miles of ocean area where 300 species are harvested comprising nearly 20% of the world catch
- assesses 146 living marine resources stocks, including 28 marine mammal stocks
- administers 36 Fishery Management Plans
- undertakes 2,000 sampling days at sea
- monitors 400 protected species permits
- reviews 7,500 Federally licensed development projects and seeks habitat mitigation for 200,000 acres
- deploys over 500 observers for fisheries and marine mammal data collection
- undertakes 350 enforcement operations resulting in more than 2,000 convictions
- inspects 200 seafood processing establishments

## *National Environmental Satellite, Data, and Information Service*

- maintains the world's first established and largest system of civil operational environmental satellites
- collects 34,000 cloud images and 8,100 atmospheric soundings from GOES while simultaneously collecting 26.6 million observations from remote platforms
- operates polar orbiting satellites providing 16.8 million global temperature and water vapor profiles, and 400,000 global ozone measurements
- estimates over 1 million wind measurements
- maintains worldwide monitoring of tropical storms, flash floods, and volcanoes
- relays 8,200 emergency locator signals from stranded ships and aircraft
- maintains climate, oceanographic, and geophysical data bases with nearly 200,000 billion bytes serving 130,000 public requests and 35,000 researcher requests

## *NOAA Corps*

- operates 18 NOAA research and survey ships
- flies 15 NOAA aircraft for research and reconnaissance missions
- performs scientific management duties in every NOAA Line Office
- fulfills duties as the 7th U.S. uniformed service, including readiness training for national defense mobilization

## APPENDIX C
## NOAA'S ENVIRONMENTAL VARIABLES

The list of NOAA's environment variables spans the full breadth of parameters describing the Earth system. The list encompasses directly measured and derived parameters of the Earth system from the bottom of the oceans to the top of the atmosphere. All the science disciplines are represented in the partial list shown below.

**External Forcing**
Solar (surface) irradiance

**Concentrations of Radiatively & Chemically Important Trace Species**
$CO_2$
Stratospheric $O_3$
$N_2O$
$CH_4$
Chlorofluoromethanes
[dichlorodifluoromethane]
Tropospheric $O_3$
CO
Stratospheric aerosols
[carbon tetrachloride
methyl chloroform]
Atmospheric aerosols
[tropospheric aerosols]
Water vapor

**Atmospheric Response Variables**
Surface air temperature
Tropospheric temperature
Pressure
Precipitation
Stratospheric temperature
Winds
Tropospheric water vapor
Components of earth radiation budget
Cloud amount, type, height

**Earth Surface Data**
Topography (absolute height)
Bathymetry
Coastline position
Sediment characteristics

**Earth Surface Properties**
Index of vegetation cover
Snow cover
Surface albedo
Soil moisture
Land surface temperature
Surface radiation budget

**Paleoclimates**
Atmospheric composition
Ice volume & extent
Land & ocean temperature
Vegetation

**Geophysical Fields**
Gravity & geoid
Global seismic properties
Magnetic field
Mineral deposits
Thermal vents

**Ocean Variables.**
Sea surface temperature
Surface radiation budget
Incident solar radiation
Ocean wind stress
Surface winds
Surface currents
Surface wave field
Shelf/slope front position
Upwelling fronts and indices
Subsurface
- circulation
- temperature
- salinity
- currents
- density
- dissolved oxygen
- depth of pycnocline
- stratification

Ocean color
Ocean chlorophyll
Sea ice extent
Sea level
Biogeochemical fluxes
Ocean $CO_2$
Nutrients
Primary productivity
Survey species composition
- species identification
- location
- number
- length
- age at length
- sex
- maturation

Commercial fisheries landings
- species identification
- landings weights
- fishing location
- fishing effort
- species lengths
- species age at length

Recreational fisheries landings
- species identification
- number
- landings weights
- fishing location
- fishing effort

Fish pathologies
- PCBs
- heavy metals

Ecosystem surveys
- zooplankton biomass
- phytoplankton
- marine toxic substances
- pollution

# APPENDIX D
# A PARTIAL LIST OF DATA SETS AT RISK

| Data Sets at Risk | Parameters Affected | |
|---|---|---|
| **Satellite** | | |
| AVHRR | Sea surface temperature<br>Vegetation index<br>Earth radiation budget<br>Aerosols<br>Cloudiness | Surface temperature<br>Visual and infrared imagery<br>Snow cover<br>Sea ice<br>Tropical rainfall estimates |
| GOES | Cloud type classification<br>Cloud amount fields<br>Cloud height fields<br>Cloud motion fields | Temperature soundings<br>Humidity soundings<br>Visual and infrared imagery<br>Precipitable water |
| TOVS | Temperature profiles<br>Humidity profiles<br>Total ozone information | Precipitation estimates<br>Stratospheric temperature<br>Surface temperature |
| SBUV/2 | Vertical and total ozone amounts | |
| SSM/I | Microwave imagery<br>Snow cover<br>Sea ice | Precipitation estimates<br>Cloud liquid water |
| SSM/T | Temperature soundings | Humidity soundings |
| SEM | Total energy detector data | Medium energy and electron detector data |
| **In Situ** | | |
| Surface synoptic observations (00, 6, 12, 18, 24z) | Temperature<br>Humidity<br>Precipitation<br>Wind vectors<br>Cloud cover | Cloud type and ceilings<br>Weather event<br>Pressure<br>Sunshine |
| Surface daily observations (hourly) | Temperature<br>Humidity<br>Precipitation<br>Wind vectors | Cloud cover<br>Weather event<br>Pressure |
| Marine surface observations | Temperature<br>Humidity<br>Precipitation<br>Wind vectors | Cloud cover<br>Weather event<br>Pressure |
| Global gridded atmospheric/ oceanic analyses | Atmospheric temperature fields<br>Atmospheric pressure fields | Forecast fields |
| Upper air synoptic data | Atmospheric temperature<br>Atmospheric humidity | Atmospheric wind vectors |
| Solar optical observing network | Solar activity | |
| Ocean thermal profiles | Ocean heat content, ocean currents | |
| Polar snow and ice data | Snow/ice observations | |
| Bathymetry data | Ocean depths | |
| Tide gauge records | Water levels | |

## APPENDIX E
## FY 1991/1992 NOAA-WIDE
## ESDIM FUNDED ACTIVITIES

**National Environmental Satellite, Data, and Information Service**

Data Rescue and Data Access at the National Geophysical Data Center
Data Rescue and Data Access at the National Climatic Data Center
Library Information/Data Rescue and Access Program
Data Rescue and Access at the National Oceanographic Data Center
GOES Satellite Data Rescue
DMSP Satellite Data Rescue
Satellite and Surface Climatology

**National Marine Fisheries Service**

Rescuing and Accessing Marine Ecological Data
Marine Ecological Data System Prototype Development

**National Ocean Service**

Sea Ice Data Rescue
EEZ Bathymetric Data Rescue
Data Inventory/Description
Ocean Tide and Current Metadata Rescue and Access
Digital Shoreline Data Rescue and Access
Tide and Lake Level Data Base
Geodetic Data Rescue and Access

**National Weather Service**

Climate Data Continuity Project
Development of an Interim Radar Data Archived Support Center

**Oceanic and Atmospheric Research**

GMCC Observatory Data Rescue and Accessibility
Rescue and Improved Access to Great Lakes Precipitation and Air Temperature Data
Rescue and Improved Access to Energetic Particle Data
Rescue and Improved Access to Research Doppler Radar Data
Rescue and Improved Access to Pacific CTD and XBT Data
Rescue and Improved Access to Atlantic Coast Meteorological Data
Integrated Atmospheric Deposition Data Accessibility
NOAA Data System Requirements Study

**Coastal Ocean Program Office**

Rescue and Improved Access to Coastal AVHRR SST Products and Ocean Buoy Data
Prototype Network Integration to Support CoastWatch Improved Data Distribution

**STORM Program Office**

Prototype Distributed Environmental Data Management System

**Office of Administration**

Develop NOAA-wide Telecommunications Network Strategy
Support Activities for the Development of a NOAA Interoperability Profile

**NOAA-wide**

The NOAA Earth System Data Directory

# APPENDIX F
# LIST OF ACRONYMS

| | |
|---|---|
| ASOS | Automated Surface Observing System |
| AVHRR | Advanced Very High Resolution Radiometer |
| AWIPS | Advanced Weather Interactive Processing System |
| CTD | Conductivity-Temperature-Depth |
| CEES | Committee on Earth and Environmental Sciences |
| CGC | Climate and Global Change |
| COADS | Comprehensive Ocean-Atmosphere Data Set |
| COP | Coastal Ocean Program |
| DMSP | Defense Meteorological Satellite Program |
| DOC | Department of Commerce |
| DOE | Department of Energy |
| DOI | Department of Interior |
| EEZ | Exclusive Economic Zone |
| EOS | Earth Observing System |
| EPA | Environmental Protection Agency |
| ESDIM | Earth System Data and Information Management |
| FCCSET | Federal Coordinating Council for Science, Engineering, and Technology |
| FTS | Federal Telecommunications System |
| GAO | General Accounting Office |
| GCDIS | Global Change Data and Information System |
| GCM | Global Climate Modeling |
| GCMD | Global Change Master Directory |
| GCRP | Global Change Research Program |
| GMCC | Geophysical Monitoring for Climate Change |
| GOES | Geostationary Operational Environmental Satellite |
| GOSIP | Government Open Systems Interconnection Profile |
| IGBP | International Geosphere-Biosphere Program |
| IPCC | Intergovernmental Panel on Climate Change |
| IRM | Information Resources Management |
| ISO | International Organization for Standardization |
| IWGDMGC | Interagency Working Group on Data Management for Global Change |
| LAN | Local Area Network |
| NARA | National Archives and Records Administration |
| NASA | National Aeronautics and Space Administration |
| NCDC | National Climatic Data Center |
| NESDIS | National Environmental Satellite, Data, and Information Service |
| NEXRAD | Next-Generation Weather Radar |
| NGDC | National Geophysical Data Center |

| | |
|---|---|
| NIST | National Institute of Standards and Technology |
| NMFS | National Marine Fisheries Service |
| NOAA | National Oceanic and Atmospheric Administration |
| NODC | National Oceanographic Data Center |
| NSF | National Science Foundation |
| NWS | National Weather Service |
| OFCM | Office of the Federal Coordinator for Meteorological Services, and Supporting Research |
| OSI | Open Systems Interconnection |
| OSTP | Office of Science and Technology Policy |
| PDB | Program Development Board |
| PMES | Physical, Mathematical, and Engineering Sciences |
| POES | Polar Orbiting Environmental Satellite |
| SBUV | Solar Backscatter Ultraviolet Spectrometer |
| SEM | Space Environment Monitor |
| SSM/I | Special Sensor for Microwave/Imaging |
| SSM/T | Special Sensor for Microwave/Temperature Sounder |
| SST | Sea Surface Temperature |
| STORM | Stormscale Operational and Research Meteorology Program (U.S. Weather Research Program) |
| TOGA | Tropical Ocean-Global Atmosphere Program |
| TOVS | TIROS Operational Vertical Sounder |
| USDA | United States Department of Agriculture |
| USGS | United States Geological Survey |
| WOCE | World Ocean Circulation Experiment |
| XBT | Expendable Bathythermograph |

## 1877 Meteorological Record for Smithville, North Carolina, Compiled by a Sergeant of the U.S. Army Signal Service

[FORM 20.]

# METEOROLOGICAL SUMMARY.

| DATE. | BAROMETER. Mean. | Mean of— A.M. observation. | Mean of— P.M. observation. | Mean of— Night observation. | Range. Highest. | Range. Lowest. | Range. Difference. | THERMOMETER. Mean. | Mean of— A.M. observation. | Mean of— P.M. observation. | Mean of— Night observation. | Range. Maximum. | Range. Minimum. | Range. Difference. | Mean relative humidity. | WIND. Prevailing direction. | Number of miles. Noon to 6 P.M. | 6 P.M. to midnight. | Midnight to 6 A.M. | 6 A.M. to noon. | Maximum velocity during month. | Total. | Amount of rain and melted snow. Amount in inches. | Number of days on which rain or snow fell. | Number of auroras. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1877. | | | | | | | | | | | | | | | | | | | | | | | | | |
| January. | 30.201 | 30.223 | 30.176 | 30.214 | 30.490 | 29.409 | 1.081 | 46.4 | 41.3 | 49.4 | 44.1 | 64 | 17 | 47 | 80.4 | SW | 1837 | 1443 | 1088 | 1287 | 36 | 5065 | 2.98 | 15 | 0 |
| February. | 30.118 | 30.144 | 30.088 | 30.125 | 30.495 | 29.583 | 0.912 | 47.8 | 43.0 | 53.1 | 45.8 | 65 | 30 | 35 | 69.9 | NE | 1832 | 1014 | 1460 | 1527 | 27 | 5383 | 2.77 | 5 | 0 |
| March. | 30.051 | 30.074 | 30.019 | 30.054 | 30.604 | 29.427 | 1.177 | 53.0 | 49.2 | 58.2 | 51.6 | 73 | 27 | 46 | 72.4 | SW | 2163 | 1763 | 1705 | 2125 | 49 | 7746 | 3.73 | 13 | 0 |
| April. | 29.964 | 29.989 | 29.943 | 29.960 | 30.393 | 29.371 | 1.022 | 59.7 | 57.7 | 63.9 | 58.1 | 79 | 36 | 43 | 78.3 | SW | 2665 | 1804 | 1450 | 2028 | 50 | 7057 | 5.18 | 10 | 0 |
| May. | 30.053 | 30.069 | 30.021 | 30.053 | 30.418 | 29.889 | 0.529 | 65.4 | 62.7 | 72.3 | 62.3 | 87 | 42 | 45 | 72.9 | E | 2074 | 1421 | 1458 | 1605 | 27 | 6258 | 2.85 | 6 | 0 |
| June. | 30.067 | 30.088 | 30.034 | 30.067 | 30.288 | 29.703 | 0.585 | 76.4 | 75.1 | 81.1 | 74.4 | 95 | 57 | 38 | 79.2 | SW | 2237 | 1857 | 1282 | 1633 | 28 | 7109 | 5.19 | 8 | 0 |
| July. | 30.021 | 30.047 | 30.005 | 30.024 | 30.313 | 29.810 | 0.503 | 80.9 | 79.1 | 85.6 | 79.2 | 98 | 64 | 34 | 76.7 | SW | 2344 | 2066 | 1580 | 1744 | 36 | 7703 | 4.81 | 11 | 0 |
| August. | 30.014 | 30.017 | 29.985 | 30.020 | 30.271 | 29.750 | 0.521 | 79.9 | 77.1 | 84.3 | 78.7 | 93 | 67 | 26 | 75.0 | SW | 2114 | 1770 | 1396 | 1638 | 24 | 6918 | 2.78 | 10 | 0 |
| September. | 30.029 | 30.056 | 29.999 | 30.031 | 30.282 | 29.690 | 0.592 | 74.3 | 72.0 | 77.8 | 72.1 | 91 | 60 | 31 | 78.9 | N | 1927 | 1740 | 1551 | 1720 | 31 | 6938 | 11.87 | 17 | 0 |
| October. | 30.089 | 30.101 | 30.068 | 30.092 | 30.407 | 29.393 | 1.014 | 66.2 | 63.5 | 71.1 | 64.4 | 82 | 46 | 36 | 79.4 | N | 1728 | 1400 | 1491 | 1692 | 51 | 6311 | 6.84 | 12 | 0 |
| November. | 30.112 | 30.137 | 30.089 | 30.116 | 30.454 | 29.706 | 0.748 | 57.7 | 53.7 | 64.7 | 56.2 | 78 | 28 | 50 | 77.9 | N | 1879 | 1758 | 1859 | 1806 | 36 | 7202 | 6.56 | 11 | 0 |
| December. | 30.198 | 30.223 | 30.179 | 30.199 | 30.571 | 29.271 | 1.300 | 50.5 | 46.6 | 54.3 | 49.8 | 66 | 24 | 42 | 80.9 | N | 1845 | 1644 | 1635 | 1762 | 50 | 6426 | 7.05 | 11 | 0 |
| Sums. | 360.927 | 361.151 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 757.2 | 724.0 | 842.8 | 748.7 | 971 | 498 | 473 | [illegible] | – | 24095 | 19222 | [illegible] | [illegible] | 445 | 81416 | 65.01 | 129 | |
| Annual mean. | 30.077 | 30.096 | 30.049 | 30.080 | 30.445 | 29.559 | 0.887 | 64.1 | 60.1 | 67.7 | 61.7 | 80.9 | 41.5 | 39.4 | 76.6 | SW. | 2008 | 1610 | 1485 | 1742 | 37.1 | 6785 | 5.42 | 10.8 | 0 |

GENERAL REMARKS.

January. Middle of month remarkable for dense fog. Mean Bar. .070 lower than 1876; Mean Temp 6°.3 lower than 1876

February. No unusual phenomena during month, Weather generally clear.

March. Rainfall above the average. Severe storm during the month.

April. Two severe storms, Wreck of vessel and loss of life occasioned by the one of 13th inst. No results of the other known.

May. Barometer lower than corresponding month of 1876, and rainfall less; No severe storms, Weather generally fair.

June. Small monthly range of barometer, No severe storms during month; Weather pleasant.

July. Excessively hot during last week. Gales on 3d, 18th, 19th - 20th. Lunar Halo on 29th. Thunderstorm on 25th, 26th, 29th, 30th.

August. Temp. & Press. were very uniform. No Storms. Some Rain squalls from S & SE on 28th. Thunderstorms on 1st, 3d, 7, 9th. Halos on 20, 22, 25, 26

September. Remarkable for preponderance of Northerly winds & cloudiness. Much rain during latter half of month, Gales 1, 20-27; Thunderstorms 6, 7, 12, 13-20. Lunar halo 21st

October. Very fine during greater part of month. Severe gale on 4th. Lunar Halo on 14.

November. Gales on 2, 8th, 21, 22, 23d-26. Thunderstorm on 28; First Frost on 12th. Remarkably cold on 30th; Lunar Halo on 20; Meteor on 20, 4 P.M.

December. Mean Temp. 10° higher than 1876; Severe SE & SW. Storm on 5th; NE & E. Gale on 29th. Frost on 1, 2, 7, 8, 9th. Lunar Halo on 15th.

Station Smithville N.C.

Robert Seyboth
*Sergeant, Signal Service U. S. Army.*

All entries previous to July were made by my predecessor.
R. S.

Source: NCDC