

QE
47
.A1
E37
1990

Earth System Data and
Information Management

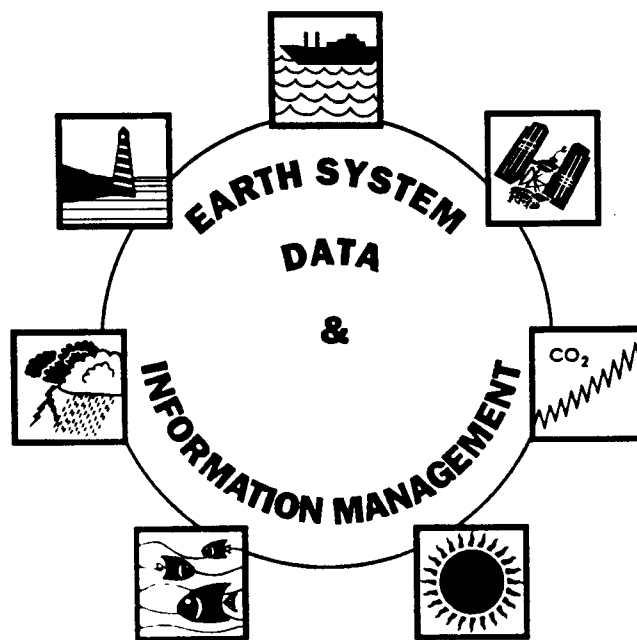


DRAFT



Earth System Data and Information Management Program

NOAA-Wide Program Development Plan



LIBRARY
 SEP 13 1991
 N.O.A.A.
 U.S. Dept. of Commerce

LIBRARY
 SEP 13 1991
 N.O.A.A.
 U.S. Dept. of Commerce

September 1990

FOREWORD

Increasingly, national and international priorities are entwined with environmental matters. Problems of acid rain, coastal erosion, ozone depletion and rainforest burning are daily topics of both individual and government attention. Concepts such as environmental risk and sustainable economic development are also expressions of the new imperatives, rivaling even military matters in importance. Responding to these challenges will require new levels of effort in measuring, understanding, and predicting Earth system processes. The key to success in all these areas is improved management of data and information about the Earth itself.

The National Oceanic and Atmospheric Administration (NOAA)—the "Earth Systems agency"—is uniquely positioned among U.S. Government entities to take the lead in environmental data and information management. First and foremost, NOAA is a science agency, with a broad spectrum of real-world operations. Its duties are precisely to measure, understand, and predict the behavior of environmental systems, and to inform policy-makers and the general public about the environment. At the same time, NOAA's position in the Department of Commerce gives it a broad interest in the economic implications of environmental matters. NOAA's activities are thus decidedly interdisciplinary, and often interagency and international in scope.

Essential to NOAA's activities in the 1990s, and beyond, will be access to huge amounts of Earth systems data and information, the grist of all environmental work. These data and information must be accurate, complete, timely, and fully integrated across the spectrum of NOAA's line office and programmatic functions. Although generally collected for topical purposes, such data and information are really long-term assets; it is part of NOAA's scientific obligation to see that they are kept secure and made readily accessible over time frames of decades to centuries. Only by understanding the environmental record in its historical context, can the Nation proceed safely and sustainably into the future.

Critical data and information management problems in NOAA must be corrected now in order to meet the challenges of the 1990s and beyond. Much of NOAA's environmental data and information holdings are presently inaccessible. In the next 2 years, NOAA stands to lose 40% of its data through deterioration of records and lack of modern computer processing and storage facilities. This problem will only worsen unless corrective action is taken immediately.

NOAA has developed a plan to correct these deficiencies, drawing upon 20 years of environmental data and information management experience. This document sets forth an Earth System Data and Information Management (ESDIM) Program Plan for FY 1991 through FY 1996 to address critical technical deficiencies and to help meet the environmental challenges facing the Nation.

Vernon E. Derr, Director
NOAA Data and Information Management Program Office

Table of Contents

FOREWORD	i
EXECUTIVE SUMMARY	iii
I. PROGRAM OVERVIEW	1
II. BACKGROUND/RATIONALE	5
III. PROGRAM DESCRIPTION	11
IV. PROGRAM IMPLEMENTATION	15
V. RELATIONSHIP TO OTHER PROGRAMS	25
APPENDIX A	
A1 Data Rescue	A1-1
A2 Data Access	A2-1
A3 The NOAA Earth System Data Directory	A3-1
A4 Standards	A4-1
APPENDIX B NOAA Data Center Characteristics	B-1
APPENDIX C NOAA Data Center Upgrade	
C1 Requirements Analysis	C1-1
C2 Requirements Initiative	C2-1
APPENDIX D Environmental Data Management Program Strategy (from NOAA Strategic Plan for 1990s)	D-1
APPENDIX E Acronym List	E-1

EXECUTIVE SUMMARY

Management of environmental data and information resources is becoming an increasingly visible and important issue for the scientific community. This is of particular importance to the National Oceanic and Atmospheric Administration (NOAA), which routinely measures and collects large amounts of environmental data and information in its own work, and is also officially charged with maintaining environmental records for the Nation. This responsibility is also a major opportunity for the agency over the next decade--to lead the scientific community and industry in tandem, developing a national data and information management infrastructure for the Earth system sciences. Through its activities over time, NOAA has become the steward of a treasury of Earth systems data and information--the most comprehensive, long-term, and up-to-date environmental description of the Earth that exists today. This treasury contains answers to urgent environmental questions facing the Nation.

Unfortunately, while the importance of environmental data and information has increased rapidly, NOAA's data and information management capability actually has decreased. Eighty percent of NOAA's 200,000 magnetic tapes have never been backed up, and some 40% exceed the typical life expectancy of such tapes (7 to 10 years). Actual tape damage has already occurred. Over 90,000,000 original hardcopy records are not digitized at all, but instead exist as virtually unusable pieces of paper, film, and glass slides. More than half of NOAA's data is now inaccessible, damaged, deteriorating, or otherwise at risk. In addition, the quantity of new data and information to be acquired is growing rapidly even as resources have been decreasing. This state of affairs cannot continue. NOAA considers its data and information management mission to be of utmost importance and is proposing to reprogram \$7 million in FY 1991 to address critical problems.

In response to these challenges, and based on 20 years of environmental data and information management experience, NOAA has developed the Earth System Data and Information Management (ESDIM) Program to be accomplished over a period of 6 years, beginning with reprogrammed funds in FY 1991. This program is summarized below, and presented fully in the following document.

Two major themes of work will be undertaken in the ESDIM Program:

A. Rescue of NOAA Data and Information

Immediately, rescue high-demand NOAA satellite data including AVHRR, GOES, and TOVS data sets; prioritize requirements for digitization and/or correction of other endangered data; digitize fragile hardcopy data; and provide means to assimilate complementary data and information streams from new sources, both within and outside NOAA.

B. Access to NOAA Data and Information

Plan and integrate data access improvements throughout NOAA, including access to data from over 1,000,000 microfiche containing atmospheric data; install data on random access media; develop on-line inventories and data browse capabilities; implement an on-line NOAA directory and usage tracking system for all NOAA's environmental data and information; make selected data readily available on CD-ROMs; and provide low cost and efficient data exchange throughout NOAA.

During the early years of ESDIM Program implementation, the requested resources will be used to correct and/or mitigate the most immediate data and information management problems--media decay and facility obsolescence. The media decay problems are addressed by the Data Rescue theme, and the facilities obsolescence is addressed by the Access to NOAA Data and Information theme.

The 6-year budget for the ESDIM Program is:

<u>Themes</u>	Millions of Dollars					
	FY 1991	FY 1992	FY 1993	FY 1994	FY 1995	FY 1996
Data Rescue	4.9	9.4	10.0	10.6	10.6	10.6
Data Access	2.1	4.6	4.0	3.4	3.4	3.4
Total	7.0	14.0	14.0	14.0	14.0	14.0

For Fiscal Year 1991, it is requested that this budget be funded by an internal reprogramming of \$7 Million, allowing NOAA to begin work immediately on its most critical data and information management concerns. In Fiscal Year 1992 and out-years, the base support established by the reprogramming will be continued, and an additional \$7 Million per year in new appropriations is requested, bringing the total ESDIM Program funding to \$14 Million per year.

Importance of Program

The two primary benefits of the ESDIM Program will be:

- 1) increased scientific productivity and improved scientific judgment, based on more accessible and more reliable data and information, and
- 2) increased operating efficiencies in data centers, utilizing up-to-date equipment and standardized technologies.

The success of all NOAA's scientific work, and the national priorities which it supports, depends on the accountability and accessibility of environmental data and information. These data and information must be accurate, complete, stable and fully-integrated across the spectrum of NOAA's line office and programmatic activities; in addition, they must be made easily accessible, in a timely and cost-effective way.

However, NOAA cannot continue to fulfill its data and information management responsibilities with the resources presently available. In default of the ESDIM Program, it is estimated that 40 per cent of NOAA's data and information holdings will be lost. As example, the environmental parameters that suffer as a result of lost high-priority satellite data are listed in the following table. In addition, NOAA will be unprepared to handle the three-fold increase in data acquisition rates planned and programmed for the mid 1990s.

High-Priority Satellite Data Sets at Risk	Parameters That May Be Lost
AVHRR	Sea surface temperature Vegetation index Earth radiation budget Aerosols Cloudiness Surface temperature Visual and infra-red imagery Snow cover Sea ice Tropical rainfall estimates
GOES	Cloud type classification Cloud amount fields Cloud height fields Cloud motion fields Temperature soundings Humidity soundings Visual and infra-red imagery Precipital water
TOVS	Temperature profiles Humidity profiles Total ozone information Precipitation estimates Stratospheric temperature Surface temperature

Table i Implications If Satellite Data Is Lost

To meet these needs and obligations, NOAA must obtain additional resources. NOAA must reassert its leadership role in archiving, protecting, refining, and utilizing the treasury of environmental data and information which it has collected over the years. If it does not, both NOAA's own stature as a science agency and the progress of environmental science will be tragically affected. Most critically, policy-makers will be forced to take actions of potentially enormous scope without the full benefit of the environmental knowledge in which they have invested.

The ESDIM Program is proposed as a comprehensive and responsible approach to providing much needed improvements in NOAA's environmental data and information management at reasonable cost. The time is right for action. After many years of being "everyone's second priority," data and information management is finally being recognized as a first-order problem for NOAA and its sister environmental agencies. Seizing the moment to undertake the ESDIM Program in NOAA will establish crucial infrastructure for the agency and the Nation in the 1990s and beyond. It is unconscionable to continue under the present circumstances, failing to develop and utilize these resources for the benefit of NOAA, the Nation, and the world.

I. PROGRAM OVERVIEW

Program Objectives

The objectives of the Earth System Data and Information (ESDIM) Program are: 1) to rescue critical NOAA data and information currently at risk; and 2) to improve access to NOAA data and information for research scientists and other users.

The ESDIM program will specifically focus on the Earth system sciences, as practiced at NOAA, encompassing meteorology, limnology, oceanography, biology/biogeochemistry, geophysics, and solar terrestrial physics.

Specifically, the ESDIM program will rely on high levels of automation and decentralization. Accordingly, the program will formulate and promote adoption of standards for the handling and documentation of environmental data and information; and, based on these standards, implement technologies to meet the burgeoning user demands for environmental data transmission, storage/retrieval, and display subsystems.

Planned activities of the ESDIM program include:

- Immediately, to rescue the high-demand NOAA satellite data sets currently at risk of being lost (specifically AVHRR, GOES, and TOVS); and within 6 years, to make 90% of all high-priority NOAA data and information readily accessible;
- To modernize processing and storage facilities at the NOAA Data Centers, such that they can accommodate large quantities of on-line (or nearly on-line) data and information; and to improve access to the Data Centers;
- To develop and implement a comprehensive directory identifying all major data and information holdings within the agency, and specifying electronic access paths to them (see Appendix A);
- To make preparations (in policy, procedures, technology) for the major new influxes of environmental data and information which NOAA must handle in the 1990s, including interagency and international streams;

- To establish quality assurance measures for NOAA-maintained data and information, ensuring continuity across space and time;
- To review and copy (or migrate to different media) 20,000 archive magnetic tapes per year, starting in FY 1991, performing remedial quality assurance in the process;
- To promote adherence to standards and development of advanced technologies for environmental data and information management; and
- To position the agency centrally and responsively in the Earth systems science research community.

Program Vision

The foundation of the ESDIM Program is a network of revitalized Data Centers, which are fully automated and distributed across the agency. Built on the model of a multi-campus academic library system, there will be several "main" Data Centers (developed from the existing NESDIS National Data Centers), and numerous, semi-autonomous "branch" Data Centers supporting NOAA's specific Line Organizations, Program Offices, laboratories and workgroups. Personal or "office" data holdings will also be recognized. Integrating the whole system will be the top-level NOAA Earth System Data Directory--itself automated and distributed--identifying all data and information holdings in NOAA. The data system architecture is diagrammed below (Fig. 1).

In general, a user will gain access to the NOAA Data Centers through a personal computer system attached to public and/or private telecommunications networks. Using specially developed software, supplied by NOAA, the user's computer will be conversant with the NOAA directory, and so be able to assist the user in browsing the Data Centers' holdings interactively--an "electronic librarian." This software will also understand the network topologies and protocols (in NOAA and other organizations), and the associated storage/retrieval facilities, so it can make arrangements for delivery of local copies of requested data and information, "on-line" or "off-line," automatically for the user.

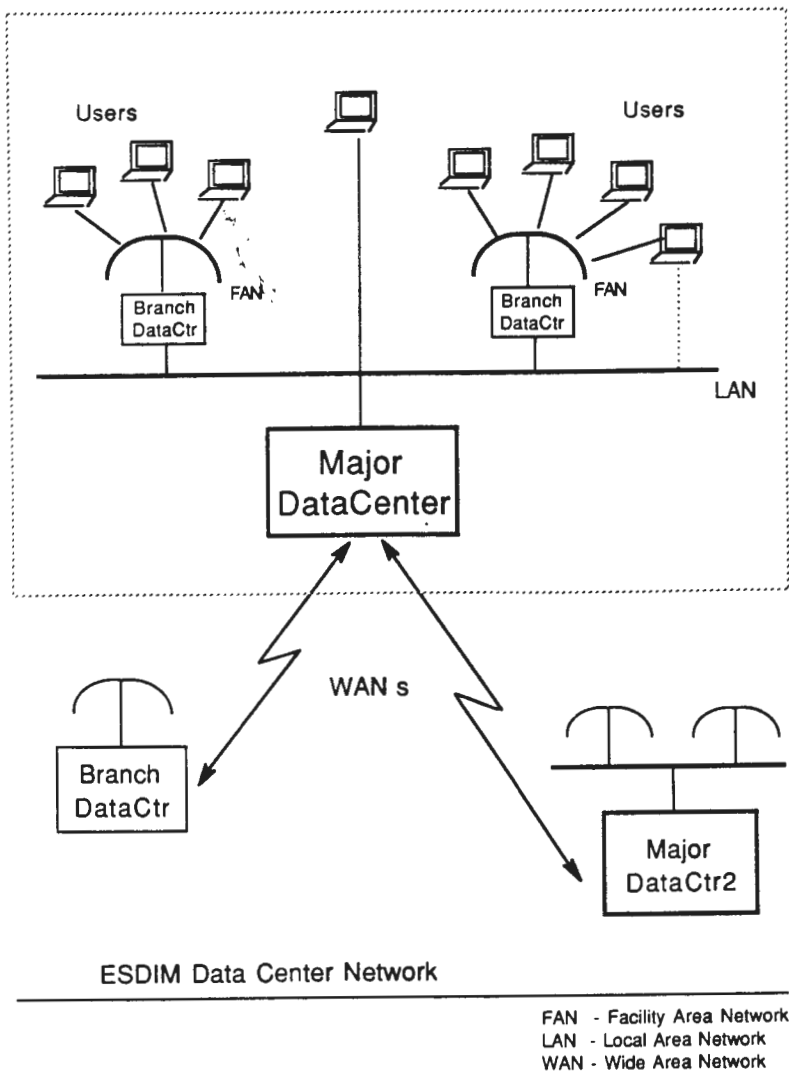


Fig. 1 Data Center Network Architecture

Full distribution of both the holdings and the access procedures is critical, too, so that Data Center resources can be dynamically deployed and reconfigured over time, in response to system loading, maintenance operations, and changing organizational needs. NOAA's Data Centers will aggressively compile and validate new data and information from sources inside and outside NOAA to support present and future NOAA programs. On-going "curatorial" operations in the Data Centers will continuously review holdings for accuracy, completeness, quality, relevance, etc. User-generated corrections and comments will also be merged back into the holdings. Actively used data and information will be automatically updated and migrated forward onto new, higher performance media.

Program Benefits

The ESDIM Program will be beneficial to both NOAA's line office operations and its interdisciplinary programs in the 1990s and beyond; it will be of value as well to NOAA's sister agencies--notably DOE, DOI, EPA, NASA, NSF, and USDA--and the many universities actively involved in the nation's environmental affairs.

The two primary benefits of the ESDIM Program will be:

1) increased scientific productivity and improved scientific judgment, based on more accessible and more reliable data and information, and 2) increased operating efficiencies in data centers, utilizing up-to-date equipment and standardized technologies.

NOAA's public activity may also spark development of major new enterprises in the private sector, concerned with storage/retrieval, processing, distribution and display of environmental data and information.

II. BACKGROUND/RATIONALE

Management of Earth systems data and information resources is becoming an increasingly visible and important issue for the scientific community. This matter was noted succinctly in a recent speech by D. Allan Bromley, Director of the President's Office of Science and Technology Policy:

In the Earth sciences alone, there are more than 100 Terabytes [equivalent to the contents of the entire Library of Congress, 10 times over]^{*} of numerical and graphical data stored in [computer] systems around the country. Within the next eight years the National Oceanic and Atmospheric Administration will add another 200 Terabytes to its files . . . as more and more information is stored on large systems, the ability to use these systems and to search them becomes critical.

Data and information management is of particular importance to the National Oceanic and Atmospheric Administration (NOAA). As an integral part of its data management mission (Table 1) to observe, analyze, and predict Earth system behavior, NOAA routinely acquires large amounts of environmental data and information. Much of this

<p>The National Oceanic and Atmospheric Administration shall . . .</p> <ul style="list-style-type: none">· 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; <p style="text-align: right;">Department of Commerce Organization Order #25-5 (1989)</p>
--

Table 1 NOAA's Data Management Mission

^{*} 1 Terabyte = 10^{12} bytes or characters = 1,000,000,000,000 bytes or characters

material is used on a real-time basis; some of it is saved for retrospective research purposes. Over time, the agency has built up a real treasury of Earth systems data and information—literally the most comprehensive, long-term and up-to-date description of the planet that exists today.

Existing Facilities

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 (see Appendix B). In addition, NESDIS operates the agency's Main Library, in Washington, DC, and several Branch Libraries at NOAA sites around the country.

Holdings of the Data Centers comprise approximately 200,000 magnetic tapes of digital data, and over 90,000,000 page-equivalents of original hardcopy manuscript and analog records and over 1,000,000 microfiche.

Historically, resources appropriated to the NOAA Data Centers have been insufficient for maintaining this volume of data and information. In fact, their operating budgets have been declining in absolute terms for the past decade, as shown in Table 2. This persistent decline in funding has imposed severe physical constraints on the Data Centers, resulting in a growing backlog of work.

FY	Millions of Dollars		
	Appropriated Dollars	Inflation Multiplier	FY 1991 Dollars
80	24.6	1.87	46.0
81	24.5	1.65	40.4
82	26.0	1.49	38.7
83	27.3	1.37	37.4
84	25.2	1.33	33.5
85	24.5	1.28	31.4
86	22.2	1.23	27.8
87	21.5	1.18	25.4
88	19.6	1.13	22.1
89	21.2	1.08	22.9
90	21.5	1.04	22.4

Table 2. NESDIS Data Management Funding 1980-1990

Present Deficiencies

The present state of management of environmental data and information in NOAA--and in the Earth system science community generally--is not adequate to the tasks at hand and expected in the future.

Much of the raw data in NOAA's possession have been collected piecemeal, to meet requirements of specific projects and individual researchers, rather than for long-term benefit. Documentation and quality control of these data are often inadequate. Information products subsequently derived from these data without appropriate regard for their origins and limitations, are similarly flawed.

Compounding these scientific concerns are a number of purely technical problems. There is no effective NOAA-wide data exchange, so data and information tend to be passed hand-to-hand around the agency and saved eclectically by individuals. Many

duplicative sub rosa collections exist. Even the so-called "permanent" archives are kept on a frightening variety of incompatible, obsolescent storage media; all of these are deteriorating with time, and some are already unreadable. Transliterations between media types are not always reliable.

In fact, eighty percent of NOAA's magnetic tapes have never been backed up, and some 40% exceed the typical life expectancy of such tapes (7 to 10 years). Actual tape damage has been recently detected. The data sets at risk if lost will severely hamper NOAA's ability to produce important environmental products as shown in Table 3. Overall, more than half of the data and information in NOAA's treasury are inaccessible, suspect, deteriorating, damaged, or otherwise at risk. In default of the ESDIM program, it is estimated that 40% of NOAA's data and information holdings will be lost, owing to lack of modern computer processing and storage facilities.

Research scientists, in NOAA and elsewhere, are tremendously frustrated and handicapped by this situation. Policy-makers and the general public really have no hope of learning--from NOAA or anyone else--the full state and trend of the environment. The situation is regrettable from all perspectives, economic, political, and scientific.

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

Table 3 Implications if Satellite and In Situ Data are Lost

Future Expectations

At the same time, NOAA and the world research community are embarking on an array of massive new environmental programs which are inherently interdisciplinary, interagency, and international in scope. Again NOAA is the designated archivist for these programs. The complexity and volume of the data and information to be collected, together with demands for improved quality control and continuity over space and time, place unprecedented demands on NOAA management and technology. The expected growth of NOAA's archives is shown in Figure 2. In FY 1990, the growth was nearly 20,000,000,000,000 (20 terabytes).

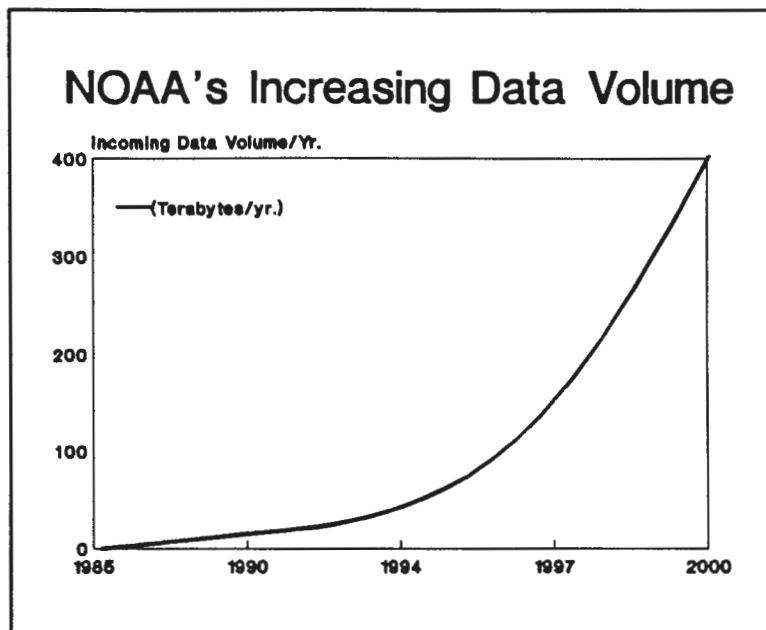


Figure 2 NOAA's Increasing Data Volume

III. PROGRAM DESCRIPTION

The NOAA Earth System Data and Information (ESDIM) Program is an urgent scientific effort which will benefit both operational and research users throughout the agency. The ESDIM Program focuses on two elements: 1) Rescue of critical environmental data and information currently at risk; and 2) Improved Access to environmental data and information, both on-hand and those in the process of being rescued. Specific activities within these elements are described below.

A. FY 1991 Reprogrammed Base Activities

In Fiscal Year 1991, the following activities--many of which have been piloted in Fiscal Years 1989 and 1990--will be undertaken immediately with resources developed by a reprogramming of base funds within the agency.

1. Data Rescue Activities

Planning and Prioritization

A scientific advisory panel, involving both NOAA-internal and external experts, will be convened to develop recommendations for the specific collections and priorities of NOAA data and information to be rescued. This panel may also consider and make recommendations on particular rescue techniques (rekeying, optical character recognition, graphical interpretation) and storage media (micrographics, magnetic tapes, cartridge tapes, CD-ROM, WORM, etc.)

Rescue of Digital Data

High-priority NOAA satellite data, including AVHRR, GOES, and TOVS data, will be immediately migrated from deteriorating magnetic tapes to more stable media. Currently, the medium of choice is the high-density tape cartridge. In addition, backup copies of these data will be made to protect against catastrophic loss.

Digitization of Endangered Data

Also based on the recommendations of the scientific advisory panel, digitization of endangered analog and tabular NOAA data will be initiated. Some of these data are on deteriorating media such as paper, or are on media that is becoming obsolete. These data include 76.5 million tabular atmospheric records (e.g., hourly and daily surface meteorological observations), 7.5 million tabular geophysical records (e.g., scaled ionospheric observations), and 6.0 million tabular oceanographic records (e.g., subsurface temperature and salinity measurements). In addition, 250 thousand analog bathythermograph records (glass slides) and a great number of analog marigrams (tide observations) need to be scanned and digitized.

System Support

Upgrades to the computer hardware, software and networking systems at the Data Centers will be procured and installed. These enhancements are necessitated by the increased computer system capabilities required for the Data Rescue activities.

2. Data Access Activities

Environmental Directory

The existing NOAA Earth System Data Directory (NESDD) prototype will be extended and networked across the Data Centers; subsequently it will be populated with descriptions of the NOAA data holdings according to the priorities established by the scientific and policy advisory panels.

CD-ROM Publication

Selected collections of NOAA's baseline environmental data and information, which are of general interest to the scientific research community, will be published on CD-ROM media. Companion software distributed on the CD-ROM will allow these collections to be conveniently accessed and manipulated, on machines as small as personal computers.

Data Exchange

A policy advisory panel, involving both NOAA-internal and external experts, will be convened to develop recommendations and specific funding

level and mechanisms to provide NOAA-supported researchers free access to environmental data and information streams and holdings; both national, interagency, and international data exchange arrangements will be considered.

Planning and Integration

An information technology study will be subcontracted to define capability and capacity requirements for the Data Centers, both short-term and long-term; and to develop an integrated approach to meeting these requirements over time. This effort was begun in FY 1989 and continued in FY 1990

B. FY 1992 Program Enhancement Activities

In Fiscal Year 1992, and out-years, the reprogrammed base activities described above will be continued; they will be spurred by additional resources, requested with this initiative, which accelerate and augment most activities and introduce several new ones.

1. Data Rescue Activities

Rescue

Augmenting the AVHRR/GOES/TOVS data (5-year project) will be the rescue of other existing atmospheric, geophysical, and oceanographic holdings in the Data Centers. Target data sets are space environment monitor, solar observatory, and seismic exploratory records.

Digitization

Similarly, digitization of other existing atmospheric, geophysical, marine fisheries, and oceanographic records in the Data Centers will be undertaken. Target data sets are historical atmospheric data tables, analog marigrams, ionospheric scaled values, geomagnetic hourly values, bathythermograph slides, and historical oceanographic tables.

Rescue New Data Streams

NOAA is implementing new environmental observation systems such as ASOS (Automated Surface Observation System), NEXRAD (Next Generation Meteorological Radar) and PROFILER (Remote Vertical Profiler of the Atmosphere) for its operational missions, but will not have the capability to save these data for archival. Many of these systems will operate at very high data rates generating massive quantities of data.

There are also new sources of geophysical and oceanographic data (e.g., snow and ice cover data, space environment data, digital hydrographic data, and acoustic doppler current meter data) that need to be captured or the data will not be saved during its operational use and thus will be lost for future scientific use.

The ESDIM Program will provide the means to save these data and make them readily accessible.

2. Data Access Activities

Data Exchange

In addition, NOAA needs data from non-NOAA satellites such as ERS-1 but does not have the capability presently to save these data streams.

On-Line Browsing

The NOAA Directory will be enhanced to include both summary "snapshot" products and detailed inventories of the NOAA data and information holdings; complementary software will be developed and distributed to permit users to "browse" the holdings directly and interactively (via electronic networks)

Random-Access Media

Techniques developed for the CD-ROM Publication activity will be extended to support preparation and distribution of customized collections of NOAA data and information on other advanced magneto-optical media.

IV. PROGRAM IMPLEMENTATION

The ESDIM program is a large-scale, long-term effort for NOAA. The work is both technically demanding and organizationally sensitive. Further, because of the time constraints--data and information at risk, coordination with ongoing programs in NOAA and other organizations, etc.--an ambitious implementation is required. The entire scientific research community, across Government, industry, and university sectors, must be entrained in the process.

The ESDIM Program will be implemented in two phases as follows: 1) reprogrammed resources will be used in FY 1991 to immediately attend to NOAA's most critical data and information management concerns--rescue of critical data and improved access to data and information; and 2) an enhanced program, beginning in FY 1992, that will expedite actions begun in FY 1991 and begin to capture new data streams at risk of being lost such as ASOS and NEXRAD.

Activities to Date

To this point, preliminary organization of the ESDIM program has been handled through a special Data and Information Management Program Office (DIMPO), reporting through the Office of the Chief Scientist to the Administrator of NOAA. DIMPO personnel presently consist of a Director and a Deputy, a Chief of Staff and a senior Information Technologist. Companion to an implementation document, a formal Program Management System will be set up to track data management assignments and performance according to the Program Implementation Plan.

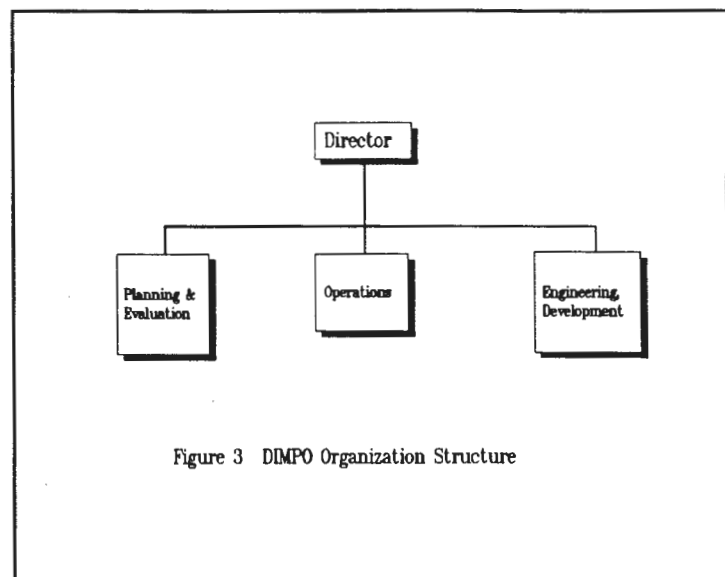
Pilot projects in Data Rescue and Data Access have also been initiated. Two large data rescue prototypes involving NOAA's satellite data were started in Fiscal Year 1990 cooperatively with NASA. The objective of these prototypes is to explore technologies required to migrate the critical AVHRR and GOES data from magnetic tape to high-density tape cartridge. In addition, a directory prototype, the NOAA Earth System Data

Directory (NESDD), has been under development since Fiscal Year 1989. This, too, will be continued, and fully implemented, during Fiscal Year 1991.

The adherence to technological standards and documentation standards is critical to the success of the ESDIM program. A major standards evaluation and implementation activity is underway. Through the Data Centers, NESDIS has been actively involved with the National Institutes of Standards and Technology (NIST) in the development and promulgation of standards relevant to its work since Fiscal Year 1988. All ESDIM-related standards will comply with national and international standards to the fullest extent possible.

Management Structure

The ESDIM Program will continue to be managed through the Data and Information Management Office (DIMPO). DIMPO staffing will still be kept intentionally small: a Director and a Deputy, supported by an Administrative Secretary. As appropriate, senior technicians and other staff from each of the existing Data Centers may also be seconded to work with DIMPO. The planned DIMPO organization structure is shown (Fig. 3).



The primary functions of the Data and Information Management Office (DIMPO) are:

- **Planning and Evaluation** - Working with representatives of the Line Offices and Program Offices, DIMPO will prepare program development plans, implementation plans, and supporting documentation describing the NOAA-wide Data and Information Management Program. Goals will be set and implementation will be monitored. DIMPO will also evaluate NOAA's data and information management activities, enlisting external expert assistance, as required.
- **Operations Oversight** - DIMPO will monitor data and information management operations throughout NOAA for adherence to the guidelines of the ESDIM program. Operations deficiencies will be identified and correctional measures will be implemented, as required.
- **Engineering Development** - Implementation of the ESDIM program will require the development of improved data and information management systems to replace NOAA's existing outmoded capabilities. DIMPO will work with NOAA data management system specialists, enlisting external engineering advisement, to develop these systems. Various design alternatives will be evaluated and proposed applications of new technology will be considered. Concept studies and system design studies will be prepared, as required.
- **Policy Development, National and International** - DIMPO will work with NOAA's Line Offices and Program Offices to develop 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 information policies will also address national and international agreements.
- **Standard Guideline Implementation** - Standards provide a common foundation of data and information management, especially in a distributed data management system such as NOAA's. Standards must be established for all facets of data and information management, including original environmental observations, quality control procedures, data handling facilities, and communication protocols. DIMPO will coordinate efforts to establish standards throughout NOAA and has

established a working group to evaluate alternatives and recommend specific data and information management standards for NOAA. This working group will be assisted by representatives of the National Institutes of Science and Technology.

- Data and Information Tracking - DIMPO will implement a tracking system to monitor NOAA's data and information management activities. The system will track data and information through all phases of data handling, storage, and retrieval, and document user requests and responses. DIMPO will evaluate tracking system information to identify data and information management deficiencies and plan improvements.

DIMPO will be formally advised by, and coordinate its affairs with, NOAA's line and program offices through the existing Data Management Council (DMaC). DIMPO Director will serve as Chair of DMaC. Other members of DMaC are the NOAA Assistant Administrators (or their designees) and a second representative from each of the five Line Offices, plus the Program Managers (or their designees) from each of the three Program Offices. DMaC meetings will be held on a quarterly basis.

Outside NOAA, DIMPO will coordinate its activities through the Computer Research and Applications Committee (CRAC) of the Federal Coordinating Council on Science, Engineering and Technology (FCCSET), in the President's Office of Science and Technology Policy (OSTP). The DIMPO Director will be appointed as NOAA's representative to FCCSET/CRAC, other members of which include representatives from DOE, DOI, EPA, NASA, NSF and USDA. The charter of FCCSET/CRAC includes all aspects of information science and technology—particularly supercomputing, networking, and visualization, all of which are important to successful management of the very large databases with which NOAA is involved.

Another important external committee is the Interagency Working Group on Data Management for Global Change, which has already established important mechanisms for interagency coordination on such matters as data policy and standards compliance.

Work Plan

Specific oversight tasks planned or in process are:

- Building a top-level consensus within NOAA on data and information topics; establishing related policies and procedures both internally and toward outside organizations
- Developing a comprehensive Program Implementation Plan addressing both technological and organizational issues
- Establishing a formal project management system, to coordinate and monitor the extensive anticipated activities of the Program
- Documenting and publishing NOAA data management plans via newsletters, journal articles, seminars and workshops, etc.
- Participating in national and international standards bodies on data base, data communications, and data management issues

The proposed ESDIM Program budget, which reflects both the requested reprogramming in Fiscal Year 1991, and supplemental funding beginning in Fiscal Year 1992, is shown in Table 4.

Table 4 ESDIM Program Budget

**Earth System Data and Information Management
Program Element Description***
(dollars in millions)

	Fiscal Year					
	91	92	93	94	95	96
I. Data Rescue	4.9 (*)	9.4 (*)	10.0	10.6	10.6	10.6
• Prioritization requirements for digitization/migration	0.2 (2)	0.1 (4)	0.1	0.1	0.1	0.1
• Rescue of endangered magnetic tape NOAA data	3.4 (1)	4.5 (1)	4.9	5.1	5.1	5.1
• Digitization of endangered analog and tabular data	1.0 (3)	2.0 (3)	2.2	2.6	2.6	2.6
• New NOAA data streams at risk	0.3 (4)	2.8 (2)	2.8	2.8	2.8	2.8
II. Access to NOAA Data and Information	2.1	4.6	4.0	3.4	3.4	3.4
• Data exchange and directory	0.8 (1)	1.3 (1)	1.5	1.5	1.5	1.5
• Planning and integration	0.5 (2)	0.5 (4)	0.4	0.1	0.1	0.1
• On-line inventories and data browse	0.0	0.9 (3)	0.5	0.3	0.3	0.3
• Data on random access media	0.8 (3)	1.9 (2)	1.6	1.5	1.5	1.5
GRAND TOTAL	7.0	14.0	14.0	14.0	14.0	14.0

* Parenthetical Numbers Represent Priorities

I. Data Rescue Element

	Fiscal Year					
	91	92	93	94	95	96
• Prioritization requirements for digitization/migration	0.2	0.1	0.1	0.1	0.1	0.1

		Fiscal Year					
		91	92	93	94	95	96
•	Rescue of Endangered Magnetic Tape NOAA Data	3.4	4.5	4.9	5.1	5.1	5.1
--	AVHRR/GOES/TOVS	2.3	2.0	2.0	2.0	2.0	2.0
--	Atmospheric data rescue . solar backscatter ultraviolet radiometer data	0.2	1.1	1.7	1.7	1.7	1.7
--	Geophysical data rescue . space environment monitor . solar observatory	0.2	0.5	0.7	0.8	0.8	0.8
--	Oceanographic data rescue . ocean station data	0.1	0.3	0.3	0.4	0.4	0.4
--	Data rescue system support	0.6	0.6	0.2	0.2	0.2	0.2

		Fiscal Year					
		91	92	93	94	95	96
•	Digitization of endangered analog and tabular data	1.0	2.0	2.2	2.6	2.6	2.6
--	Analog atmospheric data . 76.5 million tabular records	0.4	0.7	0.8	1.0	1.0	1.0
--	Analog geophysical data . analog marigrams . ionosphere scaled values . geomagnetic hourly values . 7.5 million records	0.3	0.7	0.7	0.9	0.9	0.9
--	Analog ocean data . bathythermograph slides . 6 million tabular records . tide data . fisheries data	0.3	0.6	0.7	0.7	0.7	0.7

		Fiscal Year					
		91	92	93	94	95	96
•	New NOAA Data Streams at Risk	0.3	2.8	2.8	2.8	2.8	2.8
--	Implementation planning and prototyping	0.3	0.0	0.0	0.0	0.0	0.0
--	Atmospheric data . ASOS . NEXRAD . PROFILER	0.0	1.5	1.5	1.5	1.5	1.5
--	Geophysical data . snow and ice cover data . space environment data . digital hydrographic data	0.0	0.8	0.8	0.8	0.8	0.8
--	Oceanographic data . acoustic doppler . satellite ocean products	0.0	0.5	0.5	0.5	0.5	0.5

II. Access to NOAA Data and Information Element

		Fiscal Year					
		91	92	93	94	95	96
•	Data Exchange and Directory	0.8	1.25	1.5	1.5	1.5	1.5
--	Directory	0.5	0.5	0.5	0.5	0.5	0.5
--	NOAA & national	0.3	0.5	0.75	0.75	0.75	0.75
--	International	0.0	0.25	0.25	0.25	0.25	0.25

		Fiscal Year					
		91	92	93	94	95	96
•	Planning and Integration	0.5	0.5	0.4	0.1	0.1	0.1

		Fiscal Year					
		91	92	93	94	95	96
•	On-Line Inventories and Data Browse	0.0	0.9	0.5	0.3	0.3	0.3
--	On-line inventory software development (including links to directory)	0.0	0.5	0.2	0.0	0.0	0.0
--	On-line inventory implementation (over 1,000 data sets at Data Centers)	0.0	0.0	0.1	0.15	0.15	0.15
--	Data browse software development (including pilot projects)	0.0	0.4	0.1	0.0	0.0	0.0
--	Data browse implementation (for highest-priority data sets)	0.0	0.0	0.1	0.15	0.15	0.15

		Fiscal Year					
		91	92	93	94	95	96
•	Data on Random Access Media	0.8	1.9	1.6	1.5	1.5	1.5
--	Atmospheric data on magnetic/optical disk (includes 1 million microfiche)	0.0	0.8	0.8	0.7	0.7	0.7
--	Geophysical data on magnetic/optical disk	0.0	0.15	0.15	0.15	0.15	0.15
--	Oceanographic data on magnetic/optical disk	0.0	0.15	0.15	0.15	0.15	0.15
--	Data available on low-cost, high-density media (CD-ROMs)	0.8	0.8	0.5	0.5	0.5	0.5

In the early years, as previously discussed, the majority of the requested ESDIM Program resources will be spent on time-critical activities which can no longer be postponed (e.g. satellite data rescue), and improved research-service activities (e.g. data access), for which planning is already in process. A small fraction, estimated at 5%, of ESDIM Resources will be assigned to administrative activities within DIMPO itself.

Wherever possible, efforts will be made to utilize the NOAA Cooperative Institutes, which offer both flexible access to superior talent, and a established mechanism for transferring technology from the Universities back into NOAA.

Performance Measures

DIMPO will prepare an Annual Report on the ESDIM Program for the NOAA Administrator. This report will summarize both internal accomplishments of the Program, and specific measures of services provided to its external user communities.

Performance of the ESDIM Program will be measured on four specific points:

- **Data Rescue**
Quantity of existing NOAA environmental data and information digitized and/or migrated forward from original records and aging magnetic tapes.
Target: 20,000 tape-equivalents per year

- **Data Access**
Percentage of high-priority NOAA environmental data and information which is easily located and retrieved on-line, utilizing the environmental directory
Target: >90% in 6 years

- **Data Acquisition**
Rate at which new domestic and foreign environmental data and information is received, processed, stored and catalogued in the NOAA Data Centers. (This includes increases realized through the Word Data Center system improvements)
Target: 3-fold increase over 6 years

V. RELATIONSHIP TO OTHER PROGRAMS

The responsibilities of the NOAA-wide ESDIM program are distinctly different from the responsibilities of programmatic data management. Unfortunately the same term, "data management," is often used to describe the data handling functions in both. The intended purpose of data management needs to be considered in the context of program objectives to define whether it is programmatic or agency-wide.

Program Data Management

Programmatic data management, including both research and operational programs, refers to data management that is required to carry out program objectives that are not data management services. Examples of scientific programs are the world-wide monitoring of atmospheric quality, ocean studies, and episodic mesoscale weather experiments. Examples of operational programs include weather forecasting, marine charting, and fisheries assessments.

The NOAA programs that take their own data want control of these data, but they also want access to all other data of interest. They do not want to be burdened providing data sets to secondary users. After operational use of these data and analysis, these programs generally send them to the appropriate data center for maintenance and distribution to secondary users.

Many NOAA programs use data sets that are already assembled at National Data Centers or similar capabilities of the ESDIM program. These programs access these assembled data and produce subsets thereof and products that fulfill program objectives, including analysis, prediction, and assessment. Completing this cycle, these programs archive subsequently archive these subsets and results in an ESDIM program facility for maintenance and distribution to other users.

NOAA-wide Data and Information Management Program

The NOAA-wide ESDIM program is responsible for the stewardship of all NOAA's data and information. It must provide long-term care, with continuing improvement, and provide easy access for all users. In this regard, the ESDIM program must maintain awareness of all of NOAA's environmental data and information, build directories that will guide users to these data, provide easy access, maintain data and rescue data at risk of being lost, assure proper documentation, provide end-to-end data quality accountability, set data management standards as guidelines for all of NOAA, and implement technological enhancements of data management systems as needed to improve capabilities. The ESDIM program relieves the operational and research programs of long-term data management responsibilities.

The ESDIM program is intended to be responsive to all NOAA Line Offices and Programs, as shown in Figure 4. That is, it supports NOAA's programs that in turn exist to directly support scientific missions. As infrastructure, the ESDIM program will be supportive--not directive and certainly not restrictive--of any and all NOAA activities. To accomplish its mission, the ESDIM program must develop, deploy, and maintain standards for data management throughout NOAA. Line Offices and Program Offices will obtain maximum benefit from the ESDIM program by participating fully in the specification of these standards, and subsequently adopting them for use wherever possible.

Where programmatic users reprocess and rearchive data and information, the distinction between the ESDIM program and programmatic data management is more subtle, but still essential. In these cases, a simple test should apply: if data and information are to be shared between programs or required to support a secondary user community, then their management is an agency-wide function.

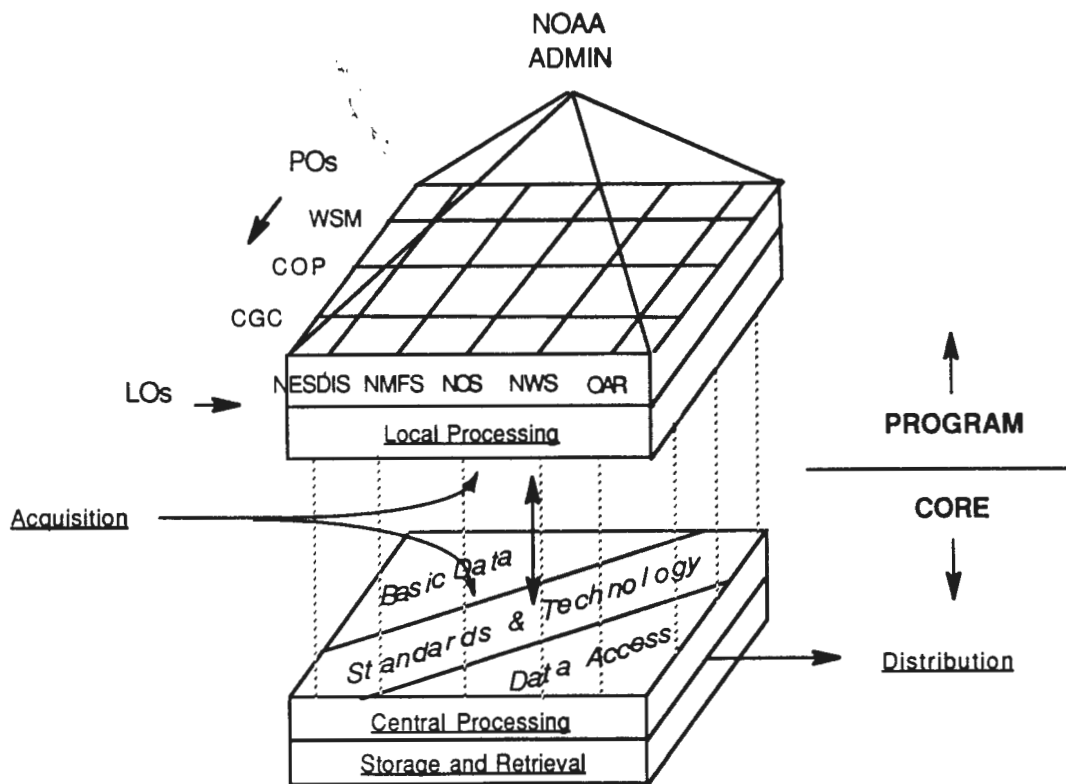


Figure 4 Relationship of ESDIM Program and Other NOAA Programs

Figure 5 provides another view of the relationship between the ESDIM program and other NOAA programs. The ESDIM program is shown to be at the core supporting NOAA programs, the programs of other agencies, and international programs. Moving outward on the figure, the ESDIM program is shown to be providing basic data and information to programs that will prepare subsets for their own programmatic purposes, and receiving data from programs that look to the ESDIM program to provide stewardship for their data and make it accessible to all users.

A business-related analogy also might prove useful to explain the difference between programmatic data management and a core or agency-wide data management function. In this case, each business [program] gathers employment data and maintains records on

its employees for payroll purposes, social security, and tax information. This data is passed on to central repositories [NOAA-wide Program], including the Social Security Administration and Internal Revenue Service where data from all sources is compiled and maintained under strict stewardship. In the programmatic sense, each business will use its basic data in management of its resources, such as organization, planning, manpower allocation, etc. As such, each business will maintain data files on its data management system and will produce subsets and products as company management requires.

Analogous to an agency-wide program, individual business data sets are assembled into nation-wide data bases at the central repositories for use by national programs such as social security and income taxes. The stewardship of these data and the requirement to provide access to users are national responsibilities. The final step in this analogy is that the central repositories are able to provide feedback of assembled demographic and business statistical information to the very businesses that provided the original data. These data and information are valuable to businesses preparing market strategies and making management decisions.

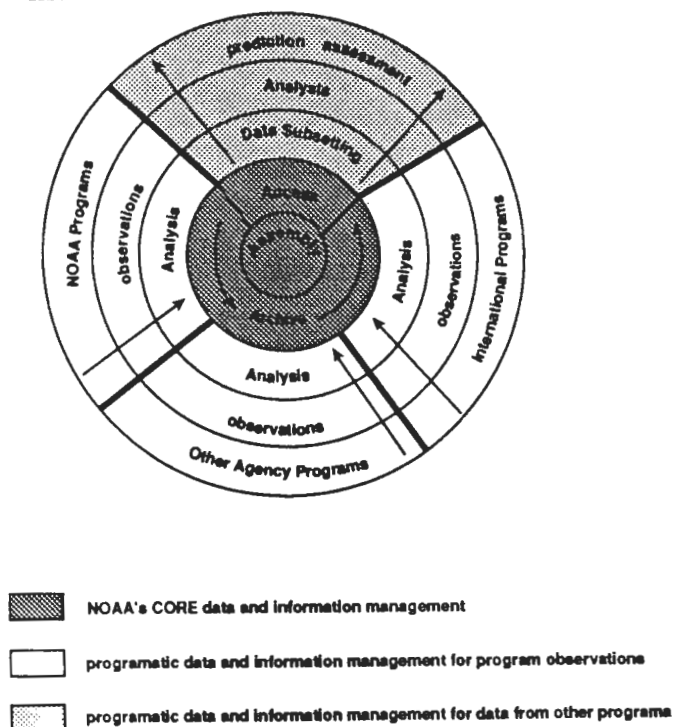


Figure 5 Core Program for Data and Information Management

its employees for payroll purposes, social security, and tax information. This data is passed on to central repositories [NOAA-wide Program], including the Social Security Administration and Internal Revenue Service where data from all sources is compiled and maintained under strict stewardship. In the programmatic sense, each business will use its basic data in management of its resources, such as organization, planning, manpower allocation, etc. As such, each business will maintain data files on its data management system and will produce subsets and products as company management requires.

Analogous to an agency-wide program, individual business data sets are assembled into nation-wide data bases at the central repositories for use by national programs such as social security and income taxes. The stewardship of these data and the requirement to provide access to users are national responsibilities. The final step in this analogy is that the central repositories are able to provide feedback of assembled demographic and business statistical information to the very businesses that provided the original data. These data and information are valuable to businesses preparing market strategies and making management decisions.

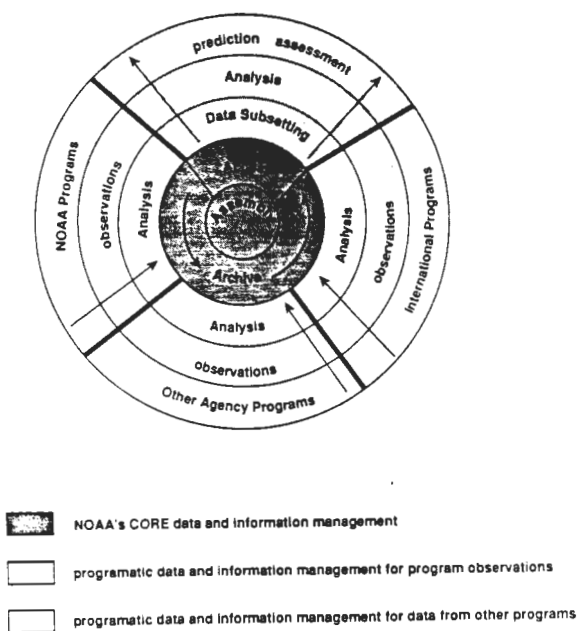


Figure 5 Core Program for Data and Information Management

NOA CENTRAL LIBRARY
CIRC DE47A1 E37 1990
Earth System Data and Info



3 8398 0005 9740 5