Vol. 3, No. 1 ● September 1992

EARTH SYSTEM MONITOR

A guide to NOAA's data and information services

INSIDE

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

Planning begins for major upgrade of NOAA's data management system

Donald Cote
Office of the Deputy Director
Environmental Research Laboratories
NOAA/OAR

During the week of April 27, 1992, NOAA's Environmental Research Laboratories (ERL) hosted an agency-wide data management workshop in Boulder, Colorado, to start laying the groundwork for a major upgrade of NOAA's data management system. Senior representatives from all of NOAA's line offices and major crosscutting programs met to discuss present and future environmental data management issues facing NOAA.

Supported by NOAA's Earth System Data and Information Management (ESDIM) Program, the data management workshop represented a major step in the development of requirements for a NOAA-wide data system modernization program. The primary goal of the workshop was to synthesize from the disparate views of the participants a NOAA prospectus for scientific data management. Over the five days the discussions touched upon all aspects of the modernization program.

Over 70 scientists, administrators, and data managers joined in the discussions. They wrestled with deficiencies in current systems and programs that make it difficult to access and use NOAA's data resources. They expressed their individual organization visions of NOAA data management for the decade of the nineties. A common goal is to implement a NOAA data management system that will transcend current barriers and provide retrospective data services that are fully responsive to the nation's data and information needs in research, industry, and governmental policy making.

The workshop opened with remarks by Gregory Withee, Deputy Assistant Administrator for Environmental Information Services, Vernon Derr from the Office of the Chief Scientist, and the workshop chairman, Robert Mahler, Deputy Director of ERL. These speakers stressed the importance of the workshop to NOAA and to the scientific community in general. The current NOAA retrospective data handling system is dangerously close to being overwhelmed by the flood of new environmental measurements and the increase in ever more sophisticated requests for environmental data and information. A robust and viable earth data and information system is becoming increasingly important to NOAA's mission. Such a system is needed to support the research being conducted by NOAA's scientists and programs and to meet the agency's commitment to the earth science community throughout the nation and the world.

The workshop continued with a review of the extent and breadth of NOAA's current data holdings. This information had been gathered from earlier ESDIM reports on the National Data Centers and new NOAA observing systems such as NEXRAD, ASOS, GOES-NEXT, and Profiler, and from an ESDIM-sponsored survey of the data holdings and data management facilities of NOAA line and program offices.

The survey revealed a broad range of data bases within the line offices. Nearly one hundred NOAA facilities provide retrospective data services from thousands of data sets covering every aspect of the agency's mission. These data, with different levels of quality and documentation, are stored on many different media at numerous

– continued on page 2

Environmental Research Laboratories
NOAA/OAR R/Ex1
325 Broadway
Boulder, CO 80303

A new look

This redesigned issue marks the beginning of the third year of publication of the *Earth System Monitor*. When it was first introduced, the *Monitor* was published as a prototype to see if it would be well-received by its intended audience of scientists, program managers, and others involved in environmental data management. Judging from the many favorable comments we have received and the willingness of authors to devote their time to writing articles, we believe the *Monitor* is a success.

We hope that the new design, typography, and physical format make the *Monitor* easier and more enjoyable to read. We look forward to continuing to provide our readers with information about data and information management activities and programs within NOAA. ■

NOAA Data Managment, from page 1

locations across the country. Data and information products derived from analysis of the raw data are provided to individual line office scientists, other government agencies, universities and industry. NOAA data holdings collected over the past century currently measure over 100 terabytes. By the middle of the decade it is projected that these data holdings will increase by more than that amount each year.

Workshop participants divided into smaller groups to discuss current data management system deficiencies in detail, to define requirements for responsive environmental data management for the decade of the nineties, and to propose concept demonstrations needed to validate new and innovative approaches to data management during the decade. Five groups were formed to address these issues from the following perspectives:

- National Data Centers,
- NOAA Centers of Data,
- Research Support Service Facilities,
- Environmental Information Analysis Center, and
- Connectivity/Communications.

Each group included representatives from each line office and program area to ensure a universal rather than parochial view in the discussions. Assisted by a chairperson and a scribe, the groups produced interim presentations and written reports. The presentations were made to the workshop as a whole. Notebooks kept in each working group to capture the flow of the discussion were used to generate written reports, which were collected at the end of the workshop. The reports will be incorporated into a system analysis of the agency-wide requirements for the end-to-end data management system upgrade for NOAA.

These sessions were followed by another series of working groups addressing agency-wide topics of importance to the data management process. These groups were formed to discuss:

- Standards,
- Science Quality Data Sets,
- User Interface, and
- Data Centers/Centers of Data Interactions.

The Working Group on Standards defined areas of interest for NOAA and developed recommendations on how best to approach standards for such elements as data formats, data documentation, and media.

The second working group consisted of scientists from across the agency who discussed the need for new integrated multiparameter data sets, such as COADS. They discussed what the attributes of these data sets should be and defined a process to create and monitor them.

The third working group shared their experiences and frustrations in attempting to access and use NOAA's myriad incompatible systems. They described the nuances of the user interfaces of the various data systems employed throughout NOAA. They expressed the need for a uniform level of access to all NOAA data resources and suggested developing of specialized tools to support data queries.

The fourth working group addressed the complex interactions of the different types of data services, and the need for coordination, common working procedures, and joint planning.

All the information from the working groups will become an important part of the process to upgrade NOAA's data management capability. The reports from the working groups have been provided to all workshop participants for review and comment. These reports will form the foundation for the system engineering analysis of NOAA's data management infrastructure. The thoughts and ideas of NOAA's data managers and scientists provide the best starting point to describe both the current capabilities and the future needs of the agency.

In FY 1993 NOAA will begin planning a major overhaul of its data management system infrastructure. The current outmoded system has evolved piecemeal over the life of the agency and its components. This evolution has created a disjointed and uncommunicative environment that is detrimental to providing the data and information required by the urgent environmental problems facing the nation. The present system will not be able to respond to the environmental demands of the next century. Through cooperative planning and initiatives on the part of all NOAA elements, however, it will be possible to prepare NOAA to carry out its responsibilities as the Earth System Agency of the 21st century.

To begin preparing NOAA's data management capability to meet this responsibility, the ESDIM Program Office has requested that ERL provide preliminary systems engineering support for the data system modernization effort. As the first step in this process, ERL has initiated requirements analysis and system planning. The results will be compiled in a four- volume set. The

– continued on page 12

EARTH SYSTEM MONITOR

The *Earth System Monitor* is published quarterly by the NOAA Office of Environmental Information Services. If you have any questions, comments, or recommended articles, or if you would like to be placed on the mailing list, please call Richard Abram at 202-606-4561 or write:

Office of Environmental Information Services
NOAA/NESDIS Ex2
Universal Building, Room 506
1825 Connecticut Avenue, NW
Washington, DC 20235

EDITOR
Richard J. Abram

U.S. DEPARTMENT OF COMMERCE
Barbara H. Franklin, Secretary

National Oceanic and Atmospheric Administration
John A. Knauss, Under Secretary and Administrator

NCDC breaks ground for new building

On June 29, ground was broken for a new facility to house NOAA's National Climatic Data Center (NCDC), which is located in Asheville, North Carolina. Construction is expected to take two years. When the new building is completed, NCDC offices will move from their present location in the Federal Building.

The world's largest active archive of weather data, the NCDC holds weather satellite images dating back to 1960; 210 million paper records; a million microfiche records; and 200,000 magnetic tapes. "We have more than 100 years of data on hand with 55 gigabytes (billion bytes) of new information added each day—that's equivalent to 18-million pages a day," said Kenneth Hadeen, the center's director

Additional Geosat data released for distribution

The U.S. Navy has announced that it will declassify and release for distribution altimeter data south of 30°S latitude from the Geodetic Mission of its Geodetic Satellite (Geosat). Previously, only the data south of 60°S had been available to researchers. The entire mission of the U.S. Navy Geodetic Satellite (Geosat) lasted for nearly 5 years, from April 1985 to January 1990. During its first 18 months of operation, known as the Geodetic Mission, Geosat collected data for military applications. In October 1986, Geosat was placed in a new orbit and until its power supply failed it collected data as part of the unclassifed Exact Repeat Mission.

A group within NOAA's National Ocean Service has processed Geosat data and provided the data to the National Oceanographic Data Center (NODC) for public dissemination. The NODC is in the midst of a project to make all the Geosat data sets available on CD-ROM. Geophysical data records from the Geosat Exact Repeat Mission are already available on a set of six CD-ROMs. The Geosat altimeter crossover differences and other Geosat data sets—including the newly declassified Geodetic Mission data—will be released on CD-ROM over the next year.

Workshop looks at NOAA networking

On August 11-13, 1992, a small group of NOAA managers and technical staff participated in a Network Integration Strategy Workshop held in Bethesda, Maryland. The workshop was held to:

- delineate NOAA's network integration goals for the next five to seven years,
- define the scope and nature of integration and interoperation requirements among NOAA networks,
- establish a schedule to achieve the goals of the networking strategy, and
- estimate a budget for the identified tasks.

The workshop will produce a summary of strategic goals and objectives and a strategic plan. Proceedings of the workshop will be published. A report about the workshop will be presented in the next issue of the *Earth System Monitor*.

News briefs

New division directors appointed at NODC

Vacancies in two of the four Divisions at the National Oceanographic Data Center have recently been filled. Dr. Parmesh Dwivedi, a computer specialist who holds a doctorate in physics from the University of Oklahoma, now heads the NODC ADP Support Division. Taking over the NODC Information Services Division is Bobby Gill, who previously served as Chief of the Systems Development Branch, National Geodetic Survey Division, within NOAA's National Ocean Service.

NGDC hosts visiting scientist from China

In June, Wen Gang of the Institute of Atmospheric Physics, Chinese Academy of Sciences, arrived at the National Geophysical Data Center, Boulder, Colorado, to work with NGDC scientists on the Global Change Data Base. He brought with him monthly mean temperature and accumulated precipitation data from several hundred Chinese meteorological stations for the period 1950-1988, and a new digital vegetation map of China. His Institute will also send to NGDC digital soils and elevation maps. During his four-month stay, Mr. Wen and NGDC researchers will work on a regional enhancement of the Global Change Data Base for China and, in collaboration with scientists at the National Climatic Data Center, will experiment with ways to optimize the spatial representation of climate data.

Global Change data managment policy issued

Official "Policy Statements on Data Management for Global Change Research" have been issued by the U.S. Global Change Research Program of the National Science Foundation. The purpose of these policy statements is to facilitate full and open access to quality data for global change research. They were prepared in consonance with the goal of the U.S. Global Change Research Program and represent the U.S. Government's position on access to global change research data. Each of the agencies within the Federal Coordinating Council for Science, Engineering and Technology reviewed and agreed to the statements.

The policy statements are published in a brochure available from the U.S. Global Change Research Program, National Science Foundation, Washington, DC 20550. Telephone: (202) 357-9715.

First phase of Pathfinder Data Project completed

In the first phase of a cooperative project between NOAA and NASA, AVHRR satellite data for 1985 through 1987 has been transcribed from approximately 2,500 magnetic tape reels and 2,600 cartridge tapes to 105 12-inch optical platters. Each optical platter holds about 6.4 gigabytes of data, or about 10 days worth of data from one satellite. The transcribed data are AVHRR (Advanced Very High Resolution Radiometer) 4 km resolution data from afternoon satellites. The data are of particular interest to scientists studying global sea surface temperature and vegetation index.

The Pathfinder Data Project was initiated by NOAA and NASA in October 1990 to improve access to operational satellite sensor data. The data transciption process was carried out at the National Oceanographic Data Center using an optical disc drive on its VAX computer system and interface software developed jointly by NASA and the University of Miami. Copies of the completed platters will be maintained at NOAA's National Climatic Data Center and at the EOSDIS Distributed Active Archive Center, NASA Goddard Space Flight Center. The Earth Observing System (EOS) Data and Information System (EOSDIS) is being developed to manage data from NASA's planned series of EOS satellites.

Early climate records in the United States

NOAA comes to the rescue to preserve 19th century Army weather data

Richard R. Heim, Jr., Global Climate Laboratory, and
John P. Hughes, Project Coordinator
National Climatic Data Center
NOAA/NESDIS

The 19th century was a time of great territorial expansion for the United States. As new settlements pushed into the vast interior, military protection was provided by the U.S. Army. In addition to its policing responsibilities, however, the United States Army also carried out a scientific task: taking weather observations at the military posts.

This responsibility fell upon the Surgeon General's office. Meteorological observations had been taken by various private observers and state networks for some time, but the Surgeon General's directive set a precedent by inaugurating the first meteorological organization in the country that was financed by the federal government. The nation's weather program was later transferred to the Department of the Army Signal Corps. In 1891, the program became a civilian operation with the establishment of the Weather Bureau under the Department of Agriculture. (In 1940, the Weather Bureau was transferred to the U.S. Department of Commerce where it remains today as the National Weather Service within the National Oceanic and Atmospheric Administration.)

The weather observations for the early years were summarized in a series of *Meteorological Registers*—books published by the Department of the Army. The data were published in Annual Reports beginning in the 1860's. Yellowed and brittle, these books are now suffering from deterioration. To preserve the data published in these early volumes, NOAA's National Climatic Data Center (NCDC) has been systematically digitizing selected elements from the monthly tables.

In these computerized files, monthly mean temperature data begin in 1822. Rain gauges were not provided to the army posts until much later, explaining why monthly precipitation data did not begin until 1837. The stations are generally Department of the Army fort stations, but in the later decades reports from voluntary observers and railroad companies were included in the published tables. The files also include data from 1854-59 published by the U.S. Patent Office. Data through 1900 have been key entered into digital form..

So far all work has been done outside the auspices of a formal project. Efforts are underway to obtain funding, as considerable work remains to be done. The following is a brief summary of the remaining work that needs to be accomplished:

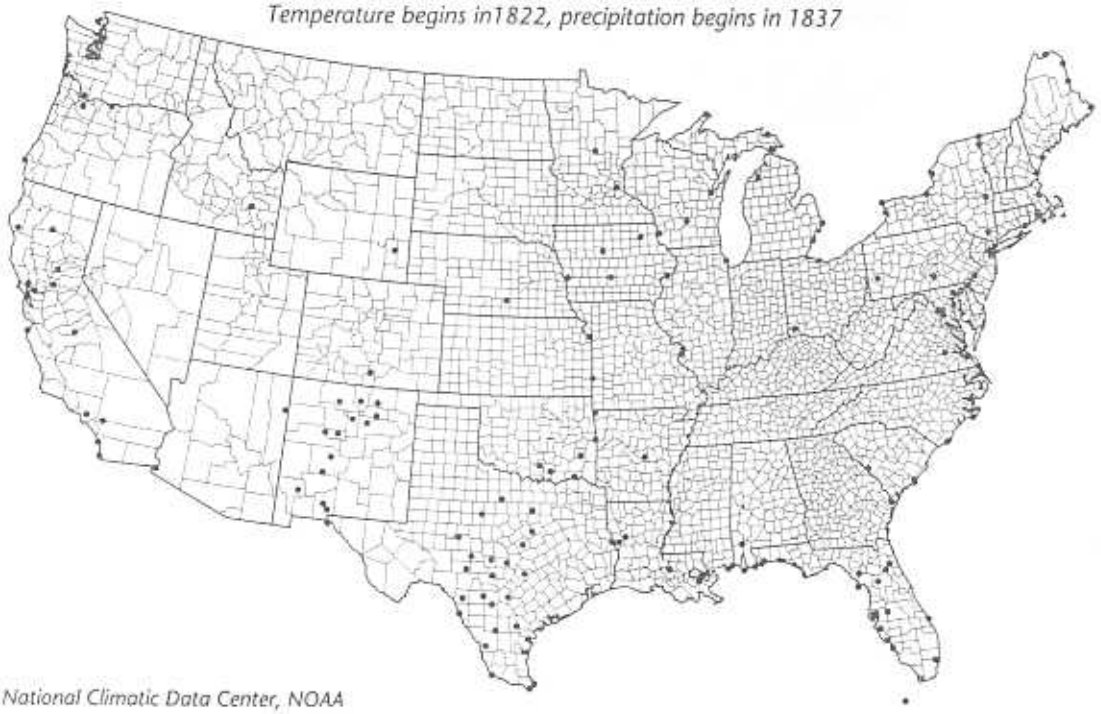
1. **Merge the digital files into a common data format.** The data were keyed in the format in which they were published to streamline the digitizing process and minimize keying errors. Unfortunately, the published tables changed format frequently. As a result there are approximately two dozen data files. The very first step, therefore, is to merge the data files into one master file with a standard project format.

2. **Perform quality control and eliminate duplicates.** Some of the published tables covered calendar years, while some covered nonstandard (e.g., fiscal) years, so there is an overlap of data in some cases. When the data files are merged, these duplicate data will need to be compared. If duplicate observa-

National Climatic Data Center
NOAA/NESDIS E/CC
Federal Building
Asheville, NC 28801

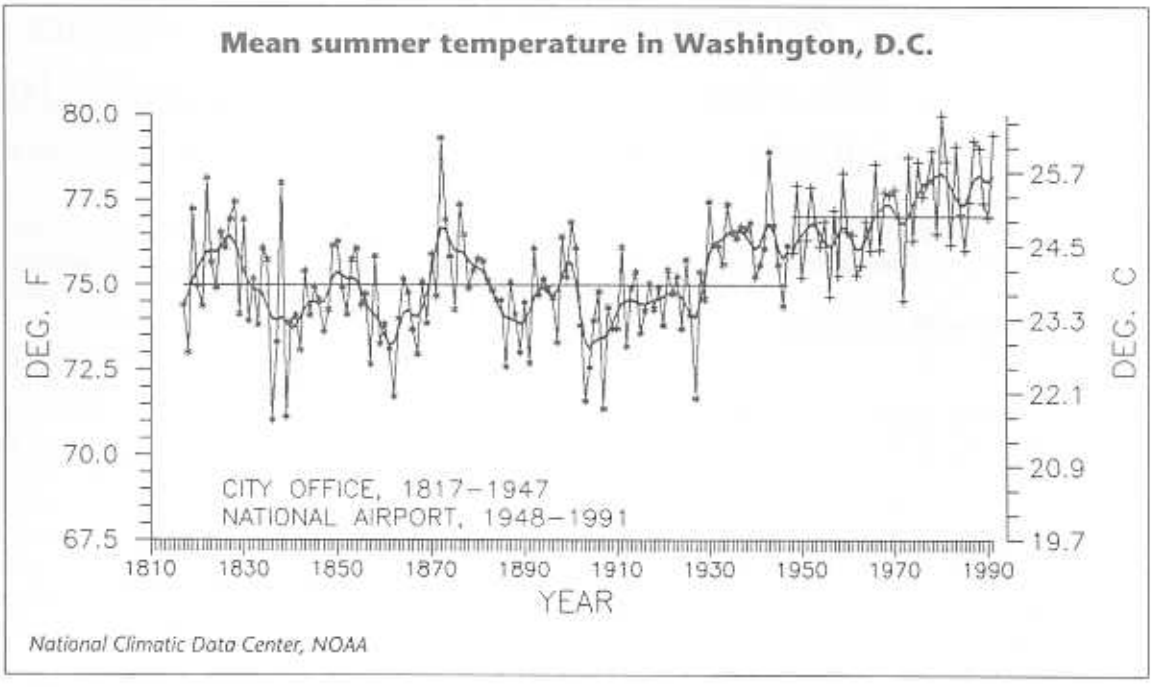
Signal Corps fort stations (as of 1852)

Temperature begins in1822, precipitation begins in 1837



National Climatic Data Center, NOAA

tions have the same value, the merge can be completed without incident. If they are different values, then they will be flagged and a decision will need to be made about which value to use. In addition, QC checks will need to be made for keying errors (these should have been minimized—though not eliminated—by the double-keying process employed) and for unreasonable values that were published (which will need to be flagged).



▲ Example of an inhomogeneity problem: Average summer (June-August) temperature for Washington, D.C., from the early 1800s to the present. The data plotted are the published values and have not been adjusted for inhomogeneities. Any long-term trends and variations (as shown by the darker smooth line), therefore, may be due to changing climate, or they may reflect non-climatic changes (i.e., instrumentation, observing practices, urban warming) that occurred over the years.

3. Inventory the data. The approximately two dozen data files comprise about 115,000 records. The data must be inventoried to determine the total number of stations having data (temperature, precipitation, and both temperature and precipitation), the number of stations per year, the number of months with missing data (by station, by year, and by element), the number of stations with complete records, how complete each station is, and its period of record. An inventory by region (i.e., by state) and year is also planned.

4. Assign station numbers and climate division numbers. The stations are identified in the files by station name and state name. A station number and climate division number will be assigned to each station. The NCDC standard station history data base will prove invaluable in determining which forts stations are already in the system (in which case the existing station number will be used) and which are new to the system.

5. Develop metadata. For some years the Signal Corps publications included latitude, longitude, and elevation tables for the stations. The metadata tables that were published have been digitized. For most years, however, the publications included only the climate data and not the metadata. An inventory of the metadata will be made to determine if metadata are available for all stations. Missing metadata (at least latitude and longitude) will be filled in. Other metadata (e.g., instrumentation, observation practices, exposure, probable station moves) will need to be determined for each station from other sources. This information can be estimated in general terms from what is known of the instrumentation and practices of the time.

6. Adjust the data. In the early 1800s, before the implementation of maximum/minimum thermometers and Stevenson shelters, thermometers were typically attached to the north wall of buildings and readings were taken three times a day. These and other factors introduce a bias into the data that needs to be compensated for before the data can be compared to modern observations. Other researchers have determined independently that some 19th century data may have a warm/dry bias. To eliminate such biases, adjustment algorithms will need to be derived and applied to the data.

Several federal agencies and individual researchers have expressed interest in the "1800s Forts" data base, as it has come to be called. By the end of the year, it is anticipated that the data will be merged into a common format, and an inventory will be produced. Only after this is accomplished, can station numbers be assigned and the biases removed. Then the data can be integrated into other climate data bases to increase the time span of the historical climate record for the United States. For example, this will allow the national drought time series (which currently begins in 1895) to be extended back several decades.

The Forts data base will enable NOAA to gain a better understanding of climate variability in the United States during the last one and one-half centuries. For example, it will help provide a more complete answer to the question of how temperatures and the frequency of droughts during the 20th century compare to those of the 19th century. This increased knowledge will have a direct bearing on national policy decisions, responses to media questions, and NOAA's contribution to research in global climate change. ■

Marine data management and research with institutions of the former Soviet Union

Political and economic disruptions put ocean data collections at risk

Ronald L. Fauquet, Deputy Director, and Robert C. Lockerman, Chief, User Services Branch
National Oceanographic Data Center NOAA/NESDIS

The political dissolution of the Soviet Union has placed extreme stress on the economies of its component nations, now newly independent. Most government programs, including environmental sciences, are undergoing severe funding cutbacks while the previously controlled economy is allowed to float to find a new equilibrium. As a result, large volumes of irreplaceable historical marine environmental data held by institutions in the former Soviet Union (FSU) are at serious risk of being lost forever to the international research community.

This situation presents the international marine science community with unique opportunities to assist in merging the previously separate data bases and cooperating with colleagues in the FSU to reanalyze the resulting enhanced global data base.

Under the leadership of Dr. Ned Ostenso, NOAA Assistant Administrator for Oceanic and Atmospheric Research, and Gregory Withee, Deputy Assistant Administrator for Environmental Information Services, an *ad hoc* interagency working group was formed last spring to determine joint US-FSU environmental science projects that could be accomplished with existing resources. The main membership of the interagency group is made up of representatives from NOAA, the National Science Foundation, and the U.S. Navy. In the three meetings held since February, the working group determined the scope of initial efforts, lined up funds, and allocated them to the implementing activities. The group also decided to focus initial efforts on three discipline areas: physical oceanography, Arctic bathymetry, and sea ice data.

National Oceanographic Data Center
NOAA/NESDIS E/OC
1825 Connecticut Ave, NW
Washington, DC 20235

Known FSU Data Holdings

Oceanographic data holdings in nations of the former Soviet Union are widely distributed both geographically and organizationally. Oceanographic research was carried out at institutions across the former Soviet Union under the sponsorship of four major organizations:

- State Hydrometeorological Committee at Vladivostok, Murmansk, Odessa, St. Petersburg, and Moscow;
- Soviet Academy of Sciences at Moscow, St. Petersburg, Vladivostok, Kaliningrad, and the Baltic States;
- Soviet Navy Ministry at Murmansk, Vladivostok, Sevastopol, and St. Petersburg (where the Navy data center is located); and
- Soviet Fisheries Ministry at Vladivostok, Murmansk, St. Petersburg, and Sevastopol.

During joint research meetings in the Washington, D.C. area in January-February 1992, discussions were held with Russian colleagues to assess the magnitude of the data risk problem and estimate volumes of data involved. Table 1 summarizes results of these discussions for data types under consideration for initial data rescue efforts. For comparison, the estimated volumes of bathythermographs and hydrocasts held by the Russian NODC are 30% and 50%, respectively, of the volumes of those data types held by the U.S. NODC.

Through discussions with Russian colleagues, Sydney Levitus of the U.S. NODC determined that many of the identified data supplement those held by U.S. national data centers. Approximately 65% of the water column data held by the Russian National Oceanographic Data Center, for example, is open ocean data from areas where little western oceanographic data exists, namely, former Soviet Union peripheral seas, the Arctic Ocean, the Indian Ocean, and the Southern Hemisphere. Similarly, bathymetric data in the high Arctic Ocean areas held by VNII Okeanologiya, St. Petersburg, complement those now held by the U.S. Navy. Therefore, the availablility of these data offers unique opportunites for cooperative work with Russian colleagues in previously unstudied areas.

Sea ice data at the Russian Arctic and Antarctic Research Institute are especially valuable. The Russian landmass forms over half of the coastline of the Arctic Ocean, and the Northern Sea Route has been the only shipping route available for Siberian development. The former Soviet Union dedicated enormous resources in ice-breakers, aerial ice reconnaissance, ice observers, and ice island camps to define, describe, and forecast sea ice behavior in the arctic. Joint research work with Russian sea ice data would promote significant advances in understanding sea ice behavior and its relation to climate change.

Soviet oceanographic data are highly desirable because of their areas of coverage and high quality. Previous comparisons of Soviet and U.S. oceanographic temperature and salinity data conducted by the U.S. NODC show close agreement in temperature measurements. Although Soviet ocean salinity observations show greater variance than U.S. data, there is close agreement between average values, and the data are considered to be of good quality. Similarly, U.S. Navy examination of narrow-beam bathymetric data collected by the former Soviet Union in the Arctic Ocean indicates that the quality and resolution of these data are at least as high as those collected by the U.S. Navy.

Program Plan

The interagency working group has endorsed a proposal for a five-year, NOAA-led program of collaboration with marine science institutions of the

FSU. The primary objectives of this collaboration are to:

- digitize large volumes of manuscript and atlas data into common formats,
- perform automated, standardized data quality control,
- perform data continuity analyses,
- merge the U.S. and FSU data to create a global data base available without restriction to the international research community, and
- reanalyze the merged data base and recompute ocean climatologies.

The mechanism proposed to carry out these efforts is an organization-to-organization collaboration with one U.S. and one former Soviet state center designated for each of the three discipline areas covered by the scope of planned data rescue work. Recommended centers by discipline area are:

- **Physical oceanographic data** - The U.S. National Oceanographic Data Center and the Russian National Oceanographic Data Center at Obninsk.
- **Bathymetric data** - The U.S. National Geophysical Data Center and the Geological Institute of the Academy of Sciences at Moscow.
- **Sea ice data** - The U.S. NOAA/Navy Joint Ice Center and the Arctic/Antarctic Research Institute at St. Petersburg.

To achieve its objectives this program will foster scientific exchange visits between U.S. personnel and their counterparts at the FSU marine science institutions. As this is being written, two oceanographers from the U.S. NODC are visiting various data centers to help initiate the first phases of this program.

– continued on page 8

Table 1. Estimated Oceanographic Data Holdings in the Former Soviet Union

INSTITUTE	DATA TYPE/ TIME COVERAGE	AREA OF COVERAGE	DATA VOLUME	ESTIMATED RISK LEVEL
Physical Oceanographic Data				
Russian NODC, Obninsk	Hydrocasts (1900-1990)	Global	220,000 digitized 260,000 undigitized	High to very high
	MBTs (1946-1991)	Global	180,000 digitized 160,000 undigitized	
Murmansk Marine Biological Institute	Hydrocasts	Baltic, North Atlantic north of England	100,00 undigitized	High
Navy Ministry, St. Petersburg	MBTs (1946-1991)	Global	150,000 digitized 200,000 undigitized	Unknown
Arctic and Antarctic Research Institute, Vladivostok and Murmansk	Hydrocasts	Polar regions, Sea of Okhotsk, Bering Sea	5,000 undigitized	Moderate
Fisheries Ministry, Vladivostok	Hydrocasts MBTs (1946-1991)	North Pacific, Sea of Japan, Yellow Sea	100,000 undigitized 50,000 undigitized	High
Riga Institute, National Academy of Science, Latvia	Hydrocasts	Global	140,000 undigitized	Moderate
Bathymetric Data				
VNII Okeanologiya, St. Petersburg	Single beam echosounder (1946-1989)	High Arctic	100,000 analog records 15,000 analog stations	High
	Multibeam echosounder (1986-1991)	North Atlantic	Unknown	High
Geologic Institute, Academy of Science, Moscow	Single beam echosounder (1946-1989)	Global, except high arctic	2 million miles analog tracklines	High
	Multibeam echosounder (1986-1991)	Global, except high arctic	20 million digital records Unknown analog records	Very high
Sea Ice Data				
Arctic and Antarctic Research Institute, Vladivostok and Murmansk	Sea ice analysis charts (1990-1991)	Polar regions	Unknown	Unknown

NOAA Earth System Data Directory provides link to worldwide information resources

The NOAA Earth System Data Directory (NOAADIR) provides a key to both managing and accessing global change data in NOAA. The Directory is being used to document NOAA data sets and data systems that support global change and other earth science studies.

Users can access the Directory at no cost through national telecommunication systems such as the Internet and NASA's NSI-DECnet or through dial-in telephone lines including an 800 number. Interconnections enable users to link from a data description in the NOAA Directory to NOAA data systems such as the STORM System and the Climatic Analysis on-line data service. The NOAA Directory uses the Oracle DBMS software and is currently installed on a DEC VAX 11/785 computer at NOAA's National Oceanographic Data Center.

The NOAA Directory is a one of the directories in the Global Change Master Directory System. This national and international system of directories uses the Directory Interchange Format (DIF) for the exchange of data descriptions between directories. The international Committee on Earth Observation Satellites sponsors the International Directory Network (IDN) that has three Coordinating Nodes: (1) the NASA Master Directory; (2) the European Space Agency Directory in Frascati, Italy; and (3) the Japanese Directory at the National Space Development Agency.

All NOAA data descriptions entered into the NOAA Directory are transferred to the Global Change Master Directory maintained at the National Space Science Data Center, NASA Goddard Space Flight Center, Greenbelt, Maryland. Through interconnections to the three IDN Coordinating Nodes, users of the NOAA Directory can gain access to worldwide data directory information.

The NOAA Directory continues to grow and now contains descriptions of over 1,000 NOAA data sets. Recent additions include about 200 descriptions documenting over 900 individual data sets held by the National Oceanographic Data Center and about 100 descriptions of foreign meteorological data reports in the collections of the the NOAA Central Library, Rockville, Maryland.

In FY 1993 the NOAA Directory will be transferred to a new workstation environment that will provide improved access to users. WAIS, the Wide Area Information Server, will be installed on the workstation to provide Internet users with access to a full text version of the NOAA Directory.

– Gerald Barton
Environmental Information Services
NOAA/NESDIS Ex2
1825 Connecticut Avenue, NW
Washington, DC 20235 ■

Access to worldwide data directory information

NOAA EARTH SYSTEM DATA DIRECTORY

The NOAA Directory is your tool to locate NOAA data sets. You can access the Directory using your PC.

Terminal settings:
Full duplex, 8 bits, no parity, one stop bit, 1200 baud
Terminal type: VT-100 (preferred)

Via NASA's NSI-DECnet:
At the $ prompt, enter: SET HOST NODC
At the prompt USERNAME:, enter: NOAADIR

Via Internet
At the $ prompt, enter: TELNET NODC.NODC.NOAA.GOV
or: TELNET 140.90.235.10
At the prompt USERNAME:, enter: NOAADIR

Via direct dial (1200 baud):
Dial 202-606-4662 or 202-606-4666
At the prompt XT_COMMAND, enter: C NODC
At the prompt USERNAME:, enter: NOAADIR

When you end the session, enter the BREAK key.
At the XT_COMMAND, enter: D
This disconnects the link to the NODC VAX, and you can hang up.

The NOAA Directory has a system of menus and prompts to lead you through your search session. There is HELP available at any point in your session, or call Gerry Barton at the number below.

If you know of NOAA data sets that should be described in the Directory, please contact your NOAA Directory LO Team Member or Gerry Barton at 202-606-4548.

Marine Data and the FSU, *from page 7*

Through this program we hope to assist the FSU institutions in maintaining the continuity of their data management and research operations, to preserve ocean data at risk of being lost, to combine the U.S. and FSU data holdings into a more complete and comprehensive global ocean data base, and to make this valuable resource available to the climate and global change research community. ■

Geographic information systems: More than just mapping

Lessons learned at the National Geophysical Data Center

David A. Hastings
National Geophysical Data Center
NOAA/NESDIS

Among the many buzz words floating around these days is the term geographic information systems, or GIS. Many people think that GIS is for geographers. Others associate GIS only with mapmaking. These are only parts of a field in computer science for managing and analyzing data located in a geographic (x,y,z) frame of reference on Earth or another celestial body.

Introduction: What is a GIS?

GIS can be defined in three ways that relate to different user interests and applications:

1. *The basic definition:* A GIS is hardware/software for developing, storing, manipulating, and selectively retrieving digital spatial data.
2. *The analogy:* A GIS can be considered a spatial Data Base Management System. Input and output include maps as well as tables.
3. *The scientific GIS:* A system that meets definitions 1 and 2 and helps the user to assess, document, and improve data quality, as well as to analyze and interpret multivariate spatial scientific data bases. It also helps make sophisticated environmental models from data that describe various aspects of the overall phenomena being modeled.

There are two families of GIS. *Vector GIS* was originally based on computer cartographic technology. It stores data as points, lines, or polygons. Vector GIS is often optimized for cartographic input and output but not for sophisticated scientific analysis.

Raster GIS, on the other hand, originated with image processing technology. Like satellite images or digital elevation models, it stores data as raster grid cells. Raster GIS was considered weak in producing high quality cartographic output, but it has recently become more capable in this area, particularly with the increased availability of advanced electrostatic, laser, ink jet, and other raster output devices. Raster systems are sometimes optimized for scientific analysis.

Today the marketplace presents users with more than a straightforward choice between vector and raster systems for scientific applications. GIS users now also have the option of hybrid systems that incoporate substantial scientific analysis from the raster side, substantial vector digitizing, and adequate cartographic output from vector and/or raster sides.

GIS applications at the NGDC

NOAA's National Geophysical Data Center (NGDC), Boulder, Colorado, has used GIS for about eight years. It first experimented with the widely used, public domain GIS called the Map Overlay and Statistical System, or MOSS. One of the early applications was prototype inventories of data, for example, location plots of aeromagnetic and gravity survey data. This GIS could do the job easily and could store metadata as well as actual scientific data in a single system that included analysis functions. For simple retrievals, however, this general purpose system was slower than NGDC's special purpose geographic inventory system, GEODAS.

Currently, NGDC uses modern scientific GIS to integrate individual global environmental data sets into a coordinated Global Change Data Base. GIS has also revealed some of the problems associated with using geographic data sets. NGDC scientists have discovered, for example, that several global data sets are inconsistently registered to the globe, several have legend categories that are globally inconsistent or inappropriate for rigorous digital processing, and others have simple data errors. All of these problems can adversely affect traditional models such as Global Circulation Models. Developers of these models usually assume that such errors are minor compared with other approximations and error propagation throughout such models. The implication sometimes seems to be that it is easier to improve modeling algorithms than to evaluate or improve data. This would not be true if they had a modest scientific GIS capability to complement their supercomputer.

NGDC has also begun to investigate the ability of existing data sets to characterize the global environment. Many of these individual data sets, however, were developed without digital access to related data. This can introduce significant errors in models. For example, assume that change in soil type in a particular area is associated with a change in vegetation type. If the data sets for soils and vegetation are developed separately with different boundaries for soil and vegetation, the data may show spurious "associations" between various soil and vegetation types (figure 1). When this can happen for several hundred combinations of vegetation and soils, models become less accurate than they can be after the input data are improved. NGDC researchers are working to enhance GIS through the application of statistical and artificial intelligence methods to improve data quality control and to assess relationships between environmental data/phenomena.

Preliminary work with statistical techniques shows that incongruous relationships often occur between environmental data in existing data sets. In one case, an area labeled a desert had a consistently low NOAA Global Vegetation Index but over 50 mm precipitation for 10 months each year. Inspection showed the area to be an icefield in the Himalayas. The developer of the ecosystem data set had lumped deserts and icefields together. Although this had no effect on the developer's project

– *continued on page 10*

National Geophysical Data Center
NOAA/NESDIS
325 Broadway
Boulder CO 80303

GIS, from page 9

to trace carbon dioxide, it may not be consistent with other research.

In another case, global vegetation indices from NOAA's Advanced Very High Resolution Radiometer and chlorophyll concentrations from phytoplankton from NASA's Coastal Zone Color showed misregistered coastlines. This hampered research into possible relationships between climate, growth conditions on land, and offshore phytoplankton concentrations. These data are in the process of being corrected to make such studies more feasible.

In its initial experiments with analysis and modeling of environmental phenomena, NGDC has found that errors in models can be attributed to:

- imperfect functionality of modeling algorithms (for example, large error propagation),
- incomplete understanding of relationships between environmental phenomena,
- inappropriate representation of phenomena with current data sets, and
- simple data errors.

NGDC scientists believe that these problems merit attention and that improvements in any combination of these areas can markedly improve Earth system science.

NGDC has been using these techniques to improve the integration and documentation of the Global Change Data Base and to work with developers to improve the design and quality of their environmental data sets. The long-term benefit of such work should be markedly improved abilities to analyze the Earth as a system, using data optimized for multivariate digital analysis. NGDC has also been adding functions to increase the ability of GIS to perform such analysis and modeling. These functions are currently prototypes, but NGDC hopes to offer these to the entire scientific community in the future.

Getting started with GIS

Researchers interested in scientific—rather than simply cartographic—aspects of GIS should approach the technology with some care. There are numerous competing packages in the marketplace, all making claims that can be difficult for the novice to evaluate. Before investing in an expensive GIS package, it is best to gain some experience with one of the low cost or public domain packages. It is helpful to start with a low cost but flexible, capable GIS that emphasizes education and scientific research, as opposed to cartographic production.

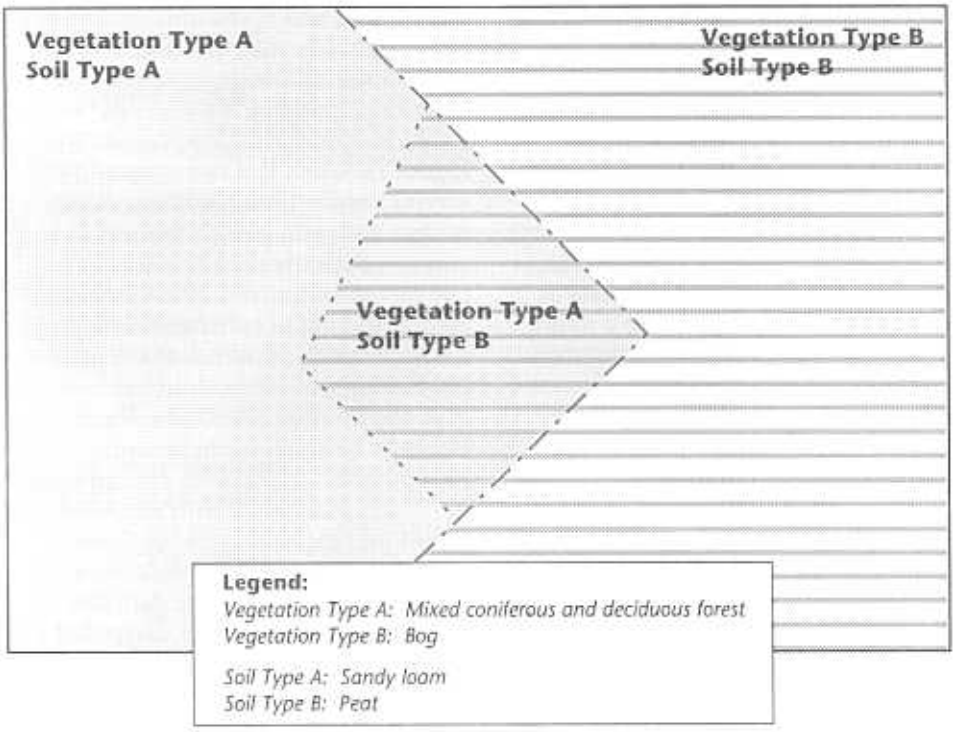
Typically, a GIS package will do much of what a user wants, but will need to be customized to meet certain specific needs. Therefore, it is smart to get a highly modular system, preferably with source code, designed for easy modification and enhancement. The user—or a contractor—can then customize as required to meet specific applications. Buying a low cost system makes it possible to afford such customization, which will be required for virtually any GIS, even the expensive ones.

Among the low cost or public domain GIS software available are two systems that serve as good starter packages:

1. The *Geographic Resources Analysis Support System (GRASS)* is a public domain GIS with strong raster and increasing vector capabilities that operates on a variety of UNIX workstations. Developed by the U.S. Army Corps of Engineers, Construction Engineering Research Center, Champaign, Illinois, GRASS may be the first GIS designed specifically for environmental analysis. It is certainly the most open system designed for this purpose, as you can obtain source code. Private companies have devoted considerable effort to support the distribution and customization of GRASS. Commercial versions for PCs and Apple Macintosh computers running UNIX are available. A copy of the general release of GRASS is available free over the Internet. NGDC can help NOAA offices evaluate and obtain a copy of GRASS.

2. *IDRISI* is a GIS software package for education and research developed at the Graduate School of Geography, Clark University, Worcester, Massachusetts. IDRISI is designed to operate in the DOS environment, with most standard PC displays supported. Although it is copyrighted, and without source

– continued on page 12



▲ Figure 1. Example of one type of problem encountered in using GIS. In this illustration, the vegetation and soil maps were developed independently and have different boundaries (at least one of which is probably erroneous). Combining the maps in a GIS produces a spurious middle zone with forest vegetation but peat soils. If the combined map were used to derive soil moisture and surface albedo (e.g., as input to an atmospheric circulation model) these parameters would also be incorrect for this zone.

Gravity CD-ROM

A new CD-ROM produced by the National Geophysical Data Center makes 600 megabytes of gravity data readily available to researchers. Accompanying the CD-ROM is access software that allows users to extract data from the disc and to view associated documentation. Among the types of data on the CD-ROM are:

- U.S. gravity station data and source documentation (including the Defense Mapping Agency and National Geodetic Survey data bases),
- data from gravity networks and from regional surveys,
- gravity anomaly grids, and
- satellite measurements.

Contact: NGDC

NODC Users Guide update

Revised pages for the *National Oceanographic Data Center Users Guide* are now available. The *Users Guide* provides detailed information about NODC data holdings, products, and services. It is published in a three-ring binder format to enable holders to keep it current by addition or exchange of revised pages. Pages for the annual update are sent automatically to registered holders of the *NODC Users Guide*. Copies of the complete *Users Guide*, including the update, are available on request.

Contact: NODC

Internet access to the NCDC and the NODC

The National Climatic Data Center recently added Internet access to make it easier for researchers to retrieve customized data. Users can also post messages and queries about climatological and meteorological data holdings at the Center. Internet users can connect via FTP to a Sun workstation and then proceed to a subdirectory to download data. For information on using the Internet to access the NCDC, send a message using Telnet to "tross@ncdc.noaa.gov" or to "nlott@ncdc.noaa.gov".

The installation of a domain server at the National Oceanographic Data Center means that it is now easier to access the NODC via the Internet. Instead of using the NODC's IP number, Internet users can now contact individual nodes at the NODC using the host name "nodc2.nodc.noaa.gov" Customers who wish to place orders or to receive information about

Data products and services

NODC data holdings, products, and services should send their inquiries to the NODC User Services Branch at "services@nodc2.nodc.noaa.gov".

Contacts: NCDC (Climate Services) and NODC

Bibliography on ecosystems of the Florida Keys

The NOAA Central Library in Rockville, Maryland, has issued *Ecosystems of the Florida Keys: A Bibliography* (Current References 92-1). Originally prepared for the Florida Keys National Marine Sanctuary located at Key Largo, Florida, by the NOAA Regional Library in Miami, the bibiography was compiled from searches of over two dozen relevant data bases. Most of the citations are to works about marine invertebrates and flora, but a substantial number cover oceanography, geology, meteorology, marine mammals and fishes, and terrestrial flora and fauna. The bibliography is available in paper copy or in digital form as a Wordperfect file.

Contact: NOAA Central Library

CONTACT POINTS

For further details and ordering information about any of the NOAA products or services listed here or elsewhere in this issue of the Earth System Monitor, *please contact the appropriate source listed below.*

National Climatic Data Center (NCDC)
Climate Services: 704-259-0682
Satellite Services: 301-763-8399

National Geophysical Data Center (NGDC)
303-497-6958

National Oceanographic Data Center (NODC)
202-606-4549

NOAA Earth System Data Directory
202-606-4548
(Gerald Barton)

NOAA Central Library
Reference Services:
301-443-8330

Marine Climatic Atlas and International Climate Summary CD-ROMs

The National Climatic Data Center has produced a new CD-ROM entitled Marine Climatic Atlas of the World. The disc contains 1-degree and 5-degree grid point monthly means and standard deviations of global historical marine data including air and sea temperature, dewpoint temperature, sea level pressure, wind speed, wave height, and the probability of superstructure icing and gale force winds. Wind and current roses are also provided. The disc enables users to either display or contour the listed data for user-defined regions. Grid point data can also be exported to a user-defined file. The disc is self-contained with access/display software and help text included.

Another recent CD-ROM product of the NCDC is version 2.0 of the International Station Meteorological Climate Summary. The revised disc contains detailed climate summaries for 980 global sites together with more limited data for 5,000 additional locations worldwide. Surface elements are summarized in up to 40 specific tables with one to 108 subtables. Individual stations are selected by a geographic pan and zoom system. All access/display software and help text are included on the disc.

Contact: NCDC (Climate Services)

Ocean Time-Series Data Set

A set of tapes containing oceanographic data taken at the same locations for long time periods is now available from the National Oceanographic Data Center. These time-series data were selected from the NODC Oceanographic Station Data File as of May 1992 and include temperature, salinity, density, and nutrients. The data set is global in coverage and contains data from 27 North Pacific sections, 56 North Atlantic sections, and 19 sections from the remainder of the world ocean (including sections from the Mediterranean, North Barents, and Tasman Seas), plus data from the 10 Ocean Weather Stations. The data are held on seven magnetic tapes: two for the North Pacific data (57,766 stations), three for the North Atlantic data (74,309 stations), and two for the remainder of the world ocean and the 10 Ocean Weather Stations (31,158 and 25,472 stations respectively).

Contact: NODC

GIS, from page 10

code, IDRISI is inexpensive, highly modular, and easily modified to add functions. IDRISI's extensive manuals include a primer on GIS, along with sample data, so it can be used "out of the box" as a self-training manual.

A package such as GRASS or IDRISI may serve all your GIS needs at little or no cost. Alternatively, you might find that modest enhancements would improve GRASS or IDRISI for your specific applications. In this case, it might be more practical to contract out the development of such enhancements than to buy a commercial package. GRASS offers many benefits in particular to users who want to install a GIS package on several computers or who require source code to review or modify,

NGDC serves as the NOAA representative on the GRASS Interagency Steering Committee. Among its functions the steering committee helps to design enhancements to GRASS and run annual GRASS users meetings. NGDC is also a major beta test site for IDRISI. In this role it has offered suggestions for fixing bugs and for enhancing capabilities for Earth System applications.

In addition to GRASS and IDRISI, there are several other low cost GIS packages available. With a steady stream of new releases coming out, the commercial marketplace offers a wide variety of GIS packages ranging up to very expensive high-end products. Yet, no GIS is generally accepted as a complete solution for all applications. Each system has its relative strengths and weaknesses. Becoming familiar with GIS capabilities and limitations through use of one of the low cost packages will prepare potential GIS users to determine their own needs and to evaluate best how to meet them.

After using a GIS such as GRASS or IDRISI, many users will want to explore other options. If so, they will face a rich, possibly bewildering, variety of commercial packages. Because this field is developing so rapidly, it is best to assume that last year's top choice is not necessarily the best choice this year. One way to start surveying what is available is to look at the trade magazines, for example, the annual survey of GIS packages published in the newsletter *GIS World*. Another way is to attend the annual GIS/LIS conference. This meeting is cosponsored by several professional societies including the American Congress on Surveying and Mapping and the American Society of Photogrammetry and Remote Sensing. The 1992 meeting is scheduled for November 8-12 in San Jose, California.

The NGDC welcomes inquiries from researchers interested in further details about its work with GIS. For further information, please contact David Hastings at the National Geophysical Data Center. Telephone: 303-497-6729. E-mail (Omnet): D.Hastings. Internet: dah@mail.ngdc.noaa.gov. ■

NOAA Data Managment, from page 1

workshop results will contribute significantly to the first two volumes, *NOAA's Current Retrospective Data Management Capability* and *The NOAA Year 2000 Data Management Vision*.

The last two volumes, on system alternatives and implementation planning, will complete the first phase of the requirements documentation. Over the next few months drafts of these documents will be given to the workshop participants for review and comments. The final drafts of the documents will be reviewed and updated at a workshop in late summer.

The ESDIM data management surveys on NOAA data sets and facilities have produced a valuable summary of the total agency data management capability. It will be used to support engineering trade-off and system sizing studies for the NOAA-wide data system modernization. It will be maintained and updated periodically to assist in the performance measurements of the infrastructure upgrades.

Workshop participants will continue their involvement in the requirements development process. Specific recommendations have been made for the working group chairpersons to become the review group for concept demonstration proposals and for the workshop to be reconvened to review planning and documentation of the data system modernization program. ■

U.S. DEPARTMENT OF COMMERCE
National Oceanic and Atmospheric Administration
Publication Distribution Facility
Rockville, MD 20852-3809
ATTN: Earth System Monitor

Address Correction Requested
OFFICIAL BUSINESS
Penalty for Private Use $300

40

Third Class Rate