Vol. 4, No. 2 • December 1993

EARTH SYSTEM MONITOR

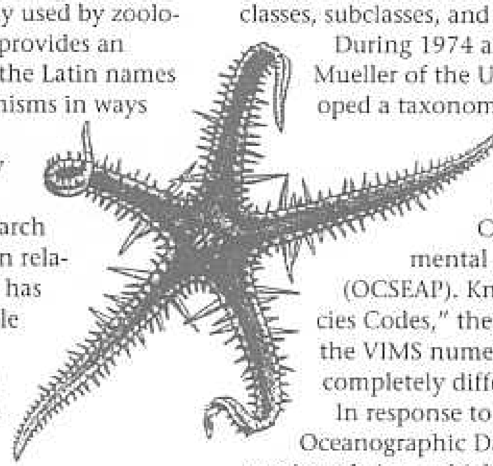
NODC Taxonomic Code links biology and computerized data processing

A guide to NOAA's data and information services

INSIDE

Jerry D. Hardy, Jr.
Data Processing Branch
National Oceanographic Data Center
NOAA/NESDIS

The Linnean system of biological nomenclature, which has been universally used by zoologists and botanists since 1756, provides an excellent method of arranging the Latin names of organisms or groups of organisms in ways that reflect their phylogenetic relationships. It is an extremely flexible system and has served the needs of the biological research community exceedingly well. In relatively recent times, however, it has become a source of considerable frustration to individuals who need to store and retrieve large amounts of biological data in a computerized environment.



During the past two decades, a number of coding systems have been developed in an effort to adapt the Linnean system to modern methods of data storage and retrieval. These systems, some based on simple abbreviations, some on discrete alphanumerical codes, and others on strictly numerical codes, have varied widely in their effectiveness and acceptance. The Taxonomic Code of the National Oceanographic Data Center (NODC) is the most flexible and widely used of these various coding schemes. A new release of the NODC Taxonomic Code (version 7.0), which contains approximately 206,000 records, was recently completed, and now—for the first time—this data product is available on CD-ROM.

History of the NODC Taxonomic Code

In 1972 Richard Swartz, Marvin Wass, and Donald Boesch published *A Taxonomic Code for the Biota of Chesapeake Bay* at the Virginia Institute of Marine Science (VIMS Special Scientific Report No. 62). Their efforts were specifically oriented toward development of a universally acceptable coding system since they felt that "everyone [should] use the same code."

The VIMS codes contained a maximum of 10 digits, with each two digits representing a different level of the systematic hierarchy. The last six digits contained (exclusively) discrete taxonomic levels (families, genera, and species) while the first four digits variously represented phyla, classes, subclasses, and orders.

During 1974 and 1975, Dr. George Mueller of the University of Alaska developed a taxonomic code that enabled him and his colleagues to manage biological data for the Alaskan Outer Continental Shelf Environmental Assessment Program (OCSEAP). Known as the "Alaska Species Codes," these codes were based on the VIMS numeric concept but used a completely different numerical sequence.

In response to a request by the National Oceanographic Data Center for a taxonomic code into which virtually any existing taxon could be placed, Dr. Mueller developed the hierarchical structure on which the present NODC Taxonomic Code is based.

Personnel at the National Oceanographic Data Center, under the leadership of Dr. Elaine

– continued on page 2

New Assistant Administrator named to head NOAA/NESDIS

Robert S. Winokur has been selected as the new NOAA Assistant Administrator for Satellite and Information Services. This position serves as the head of the National Environmental Satellite, Data, and Information Service (NESDIS), one of the five NOAA line offices. NESDIS is responsible for managing and operating NOAA environmental satellites and environmental data centers.

A career Navy employee, Mr. Winokur had served since 1985 as Technical Director in the Office of the Oceanographer of the Navy, Office of the Chief of Naval Operations. He has published numerous papers and reports on underwater acoustics and Naval oceanography. He is also the recipient of many awards, including the Presidential Distinguished Executive and Meritorious Rank Awards for senior executives. Mr. Winokur is currently Vice President for Technical Affairs of the Marine Technology Society, and is a Fellow of the Acoustical Society of America. ■

National Oceanographic Data Center
NOAA/NESDIS E/OC12
1825 Connecticut Avenue, NW
Washington, Dc 20235

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

NODC Taxonomic Code, from page 1

Collins and Mary Hollinger, began adding taxa to this basic framework, and in 1977, published the first edition of the NODC Taxonomic Code. In this edition, which contained approximately 16,000 records, two digits were added to the basic lO-digit format in order to allow inclusion of subspecies or varieties.

A second edition containing approximately 18,000 records was published in 1978, and a third edition containing approximately 25,000 records was published in 1981. The last hard copy edition was published in 1984 and contained approximately 45,000 entries. Subsequent releases have been available only in digital format. The present release, version 7.0, contains approximately 206,000 records.

Structure and content of the NODC codes

The NODC taxonomic codes contain a maximum of 12 digits. The code numbers are partitioned into a series of 2-digit couplets, each of which represents one or more levels of the taxonomic hierarchy (Table 11).

Taxonomic information is contained in the hierarchy of each code. For example, the species 9301010101 is part of the genus 9301010 in the example given in Table 1. Because the taxonomic code has expanded primarily in response to user requests for additional taxon numbers, the code does not reflect a single consistent taxonomy. In the study of taxonomy these classifications can change over time. There are plans to have the taxonomy inherent in the NODC Taxonomic Code peer reviewed by specialists during the next one to two years.

The NODC Taxonomic Code on CD-ROM

To assist users, the data files on the CD-ROM version present the NODC Taxonomic Code in a number of different ways. These files include the NODC Taxonomic Code in different sort orders, subsets of the Taxonomic Code, and other versions modified for special uses or applications.

The CD-ROM includes five independent data files:

1. TAXNUM.DAT. This file presents the NODC Taxonomic Code in phylogenetic (numerical code) order and contains all synonyms, common names, and administrative records (those records with "!" in position 1).

2. TAXALPHA.DAT. This file presents taxonomic information with scientific and common names ordered alphabetically (administrative records found in the file TAXNUM.DAT are omitted).

3. TAXTRIV.DAT. In this file, the order of genus and species names has been reversed (species, genus), as have all common names consisting of more than one word. The information is presented in this way to assist in finding names that may have been misspelled or that were placed in another genus. The data are in alphabetical order.

4. TAXHIER.DAT. This file contains only the major levels of the taxonomic hierarchy (Phylum, Class, Order, Family). The data are presented in phylogenetic order.

5. TAXBRIEF.DAT Records in this file are less than 80 characters to facilitate viewing records on a standard display screen. It includes all records found in TAXNUM.DAT, but the fields "Author Name" and "Code Level" have been omitted.

The CD-ROM also contains a READ.ME file, the file FORMAT.TAX, which presents the record layout of version 7.0 of the NODC Taxonomic Code,

– continued on page 12

Table 1. Example of NODC Taxonomic Code structure*

93	(2 digits)	Subkingdom, Phylum, Subphylum, Class, Superorder, Order
9301	(4 digits)	Superclass, Class, Subclass, Superorder, Order, Suborder, Infraorder, Section, Superfamily
930101	(6 digits)	Class, Order, Suborder, Family, Subfamily
93010101	(8 digits)	Genus
9301010101	(10 digits)	Species
930101010101	(12 digits)	Subspecies

*(*numbers in this example are fictitious)*

EARTH SYSTEM MONITOR

The *Earth System Monitor* (ISSN 1068-2678) is published quarterly by the NOAA Environmental Information Services office. 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:

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

EDITOR
Richard J. Abram

EDITORIAL ASSISTANT
Nancy O'Donnell

U.S. DEPARTMENT OF COMMERCE
Ronald H. Brown, Secretary

National Oceanic and Atmospheric Administration
D. James Baker,
Under Secretary and Administrator

Data products and services

Historical Arctic rawinsonde data now available on CD-ROM

Over 1.2 million vertical soundings of temperature, pressure, humidity, and wind—representing all available rawinsonde ascents from Arctic land stations north of 65°N through 1987—are now available on three CD-ROMs. Most stations begin in 1958, but a few begin in 1947 or 1948. Coverage is relatively uniform, except in the interior of Greenland. Typically, 20-40 levels are available in each sounding.

Sounding data were obtained by the National Snow and Ice Data Center (NSIDC) from the National Center for Atmospheric Research, Boulder, Colorado and the NOAA National Climatic Data Center. Compilation and quality control by J. Kahl and M. Serreze were supported by NOAA, the National Science Foundation, and the Electric Power Research Institute. CD-ROM production was funded by NASA/EOSDIS.

There is no charge for this CD-ROM data product as long as supplies last. Future data distribution may be via anonymous FTP instead of CD-ROM.
Contact: National Snow and Ice Data Center, User Services, University of Colorado, Boulder, CO 80309-0449. Telephone: 303-492- 2366 or 5171. Fax: 303-492-2468. Internet: nsidc@kryos.colorado.edu

CD-ROM version of NOAA Buoy Database completed

The National Oceanographic Data Center (NODC) has completed a set of 14 CD-ROMs containing 7.9 gigabytes of marine meteorological, oceanographic, and wave spectra data collected by moored buoys and C-MAN (Coastal-Marine Automated Network) stations operated by the NOAA National Data Buoy Center (NDBC). NDBC buoys are deployed in waters off the United States, including Alaska and Hawaii, as well as in the Great Lakes. C-MAN stations are located at sites along the U.S. coast and on offshore towers.

The NDBC buoys began reporting in the early 1970s and the NODC archive holds data from October 1972. The first C-MAN stations became operational in March 1983, and the NODC archive of C-MAN data begins in 1985. These discs hold all the NDBC buoy and C-MAN data in the NODC archive through July 1992. Data files for each buoy or C-MAN station are arranged chronologically in directories by their station identifier. Each file contains data for one month for that buoy or station. The set of 14 discs is organized by geographic area:

- Atlantic Ocean - 4 discs
- Gulf of Mexico - 2 discs
- Great Lakes - 2 discs
- Western and Central Pacific Ocean - 1 disc
- North Pacific Ocean (above 50°N) - 1 disc
- Eastern Pacific Ocean - 4 discs

Contact: NODC

Report on snow and ice data workshops

World Data Center A for Glaciology (Snow and Ice) has published a single volume—Glaciological Data, Report GD-25—that contains the proceedings and recommendations of two workshops. The first, Snow Watch '92, Detection Strategies for Snow and Ice, was sponsored by the Canadian Climate Centre, the World Meteorological Organization, and the Institute for Space and Terrestrial Science, University of Waterloo. This third meeting of the informal Snow Watch group convened to review strategies for detecting changes in global snow cover as climate system indicators. It contains 20 papers by scientists from Canada, China, Finland, and the United States on snow cover and lake ice observations, climate change detection, and snow cover modeling, as well as reports of five working groups. These groups considered the progress made in implementing the recommendations of the previous meetings, published in GD-11 (*Snow Watch '80*) and GD-18 (*Snow Watch '85*).

The second report is from a Workshop on Cryospheric Data Rescue and Access supported by NOAA's Environmental Science Data and Information Management program and convened by Dr. Robert Crane, Pennsylvania State University. The workshop participants developed guidelines for selecting data required to support observational studies of modeling the cryosphere, focusing on snow cover, sea ice, ice sheets, and glaciers. Candidate high-priority data sets for these areas are identified in the reports, as well as data at risk and in need of rescue. Copies of GD-25 containing reports of both the workshops are available for $10.
Contact: WDC-A , Glaciology, CIRES, Campus Box 449, University of Colorado, Boulder, CO 80309-0449. Telephone: 303-492-5171. Fax: 303-492-2468. Internet: nsidc@kryos.colorado.edu

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: 704-271-4682, Fax: 704-271-4876
Satellite Services: 301-763-8399, Fax: 301-763-8443

National Geophysical Data Center (NGDC)
303-497-6958
Fax: 303-497-6513

National Oceanographic Data Center (NODC)
202-606-4549
Fax: 202-606-4586

NOAA Environmental Services Data Directory
202-606-5012
(Gerald Barton)
Fax: 202-606-0509

NOAA Central Library
Reference Services:
301-713-2600
Fax: 301-713-4599

Summary reports on major weather events of 1993

The Research Customer Service Group of the National Climatic Data Center has issued three technical reports about significant weather events of 1993:

- *The Big One: A Review of the March 12-14, 1993 Storm of the Century* (Technical Report 93-01)
- *1992-1993 Winter Precipitation in Southwest Arizona* (Technical Report 93-02)
- *The Summer of 1993: Flooding in the Midwest and Drought in the Southeast* (Technical Report 93-04).

These reports are available free of charge while supplies last. Because of their length and attachments that cannot be sent electronically, the reports can be sent only through the mail.
Contact: NCDC. Telephone: 704-271-4800. Internet: orders@ncdc.noaa.gov.

Digitizing Historical Records for the Comprehensive Ocean-Atmosphere Data Set (COADS)

A search and rescue mission for marine weather observations

Joe D. Elms, National Climatic Data Center, NOAA/NESDIS; Scott D. Woodruff, Climate Diagnostics Center, NOAA/ERL; Steven J. Worley, National Center for Atmospheric Research; and Claire S. Hanson, WDC-A Glaciology/ National Snow and Ice Data Center

Digital marine weather records taken over the global ocean, based primarily on merchant ship observations, date back to 1854. These records and related statistics are available in the Comprehensive Ocean-Atmosphere Data Set (COADS). The cooperative effort that produced this widely used resource for climate research is a continuing project that started with COADS Release 1 (Slutz *et al.*, 1985; Woodruff *et al.*, 1987), which covers the period 1854-1979, and the recently completed COADS Release 1a (Woodruff *et al.*, 1993), which extends the Release 1 products through 1992.

There are plans to update the record following World War II (COADS Release 1b) over the next year, with the scope of this update guided partially by the needs of the global atmosphere re-analysis project (Jenne, 1992), and later to update the entire period of record as the goal for Release 2. These efforts are intended to improve coverage during data sparse periods since 1854, extend the period of record back in time prior to 1854, and correct known problems where possible.

The work being accomplished today under the COADS project would not have been possible without the insight and devotion of many earlier mariners and researchers. Among the most important of these are Admiral Francis Beaufort who in 1805—as a Lieutenant in the British Navy—promoted the wind speed scale that now bears his name, and Matthew F. Maury, Superintendent of the U.S. Naval Depot of Charts and Instruments from 1842 to 1861, who helped organize international efforts to systematically collect marine data. Data collected by early mariners proved crucial to improving sailing times based on the knowledge of average weather and current conditions such as provided by Pilot Charts first developed by Maury.

Unfortunately, large quantities of the earliest records as well as numerous records from more recent periods such as the two World Wars, were never saved in digital form. Approximately one million U.S. Merchant Marine and Ocean Station Vessel (1940-1945) observations for the period 1938-1948 were punched on cards from WB (Weather Bureau) Form 1210A-Marine and archived as Deck 115. The cards were discarded in November 1960 because:

- punched cards were not then a recognized archive medium,
- an estimated 40% of the records contained errors (and adequate technology to perhaps correct these errors was not available), and
- the Ocean Station Vessel observations (a large portion of the deck) were maintained elsewhere.

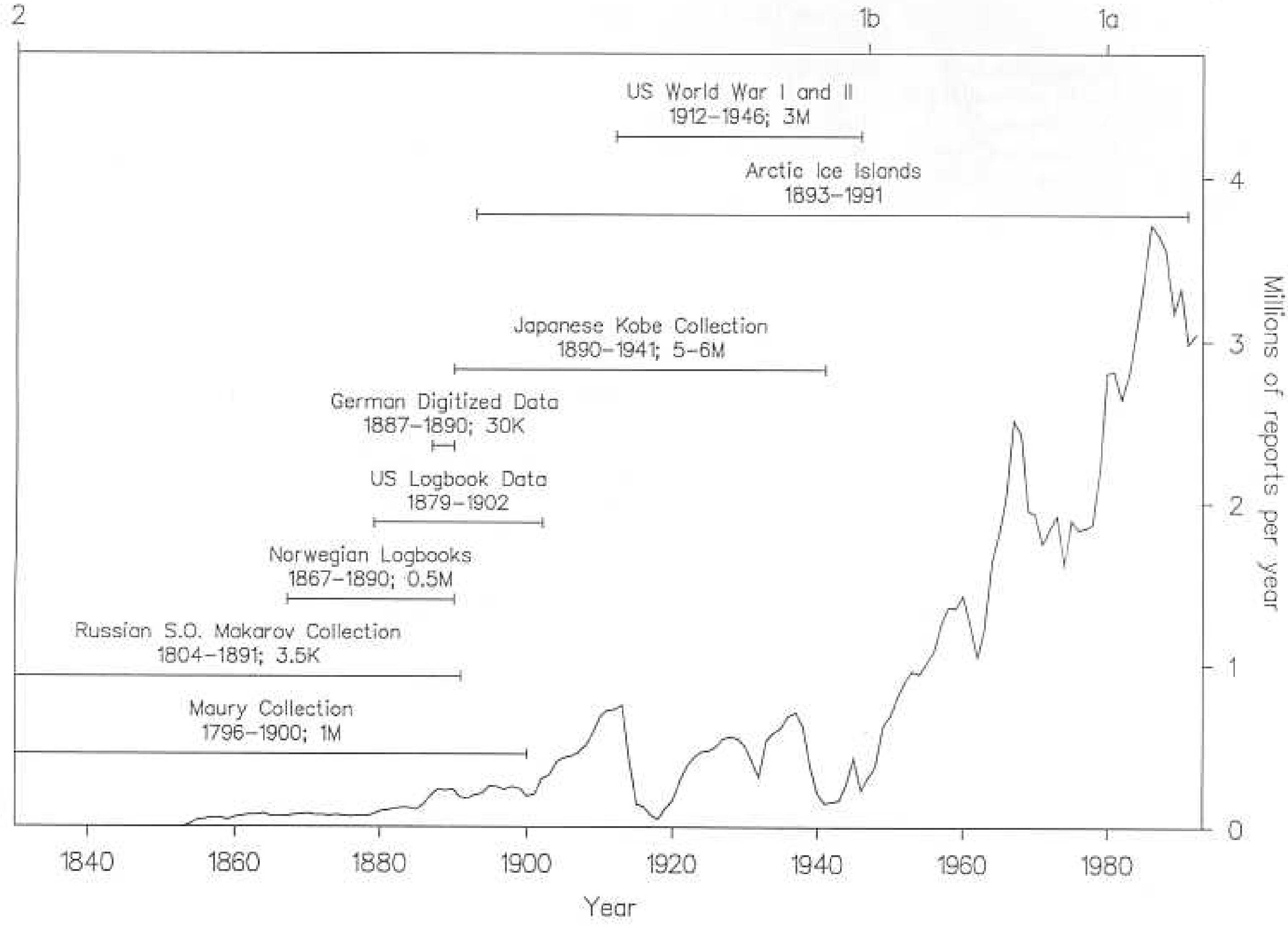
All of the U.S.-recruited merchant marine logbook data prior to 1949, including Maury's own collection of logbooks, have not been digitized. Moreover, in many cases surviving original logbooks are in deteriorating condition. This paper discusses efforts by the COADS project, with the vigorous cooperation of a number of other countries in the international community, to improve data coverage and quality by locating and digitizing as many as possible of these records.

Digitizing records

As part of the COADS project, a major effort has been underway since 1989 at the National Climatic Data Center (NCDC) to digitize U.S.-recruited merchant marine weather observations taken during World Wars I and II and thereby to help fill significant data voids (Fig. 1). The original plan was later expanded to include all available merchant marine data (approximately three million observations) held in the U.S. archives, not in digital form, for the period 1912 to 1946. However, the original estimate provided by NCDC of approximately 18 million undigitized observations taken during 1913-1919 and 1937-1948 (Woodruff *et al.*, 1987) now appears much too high. After actually retrieving the records from the archives, NCDC discovered that the World War II merchant marine records were missing and that a significant number of the Navy records were already digitized and available in COADS. Locating marine weather records for these periods in any of the world's other national archives has proven generally unsuccessful, as the World Wars disrupted routine observing practices and normal government functions such as data archiving.

Only recently, for example, did we discover that the U.S. World War II merchant marine records, as distinct from similar records for other periods, were under the jurisdiction of the U.S. Maritime Commission and War Shipping Administration (later the Maritime Administration). In 1974 the Maritime Administration made the determination to "destroy immediately" all the deck logbooks for the period 1940 through 1947. These were all archived at the New York and San Francisco Federal Archives and Records Centers with the request for destruction approved by the Archivist of the United States. Although the actual destruction had not been unequivocally verified at the time we prepared this article, it is highly probable that the order was carried out because of the large volume of space these records occupied. The justification attached to the disposition form stated: "The Maritime Deck Department Log Books for the WW II period have little if any research value." Unfortunately, it is felt today that these data would have been of significant importance to a

NOAA/NESDIS/NCDC E/CC22, Federal Building, Asheville, NC 28801- 2733; NOAA/ERL R/E/CD, 325 Broadway, Boulder, CO 80303-3328; NCAR, P.O. Box 3000, Boulder, CO 80307-3000; NSIDC, University of Colorado, Boulder, CO 80309-0449.



▲ **Figure 1.** Annual global marine reports after duplicate elimination (curve) for COADS Release 1 through 1979, continued by Release 1a through 1992. Horizontal lines span the time periods for data now being collected and digitized (World War I and II; Arctic) or proposed for future digitization, with the approximate numbers of reports shown in millions (M) or thousands (K). Labelled ticks along the upper horizontal axis mark the starting years for Release 1a, and those planned for Releases 1b (1947) and Release 2 (1854, or earlier).

number of research projects, including the quest to establish the validity of global warming.

At the outset of the project to digitize the World War I and II data, a strategy was developed for digitizing the data, which is proving to be generally applicable for keying historical ship logbook data. The goal was to key as much as possible of the data and metadata (information about the data) contained in the logbooks, even if they were not of immediate use in the COADS project. This would maximize information preservation as the paper forms deteriorate. These forms should still be considered for microfilming, however, because not all information is digitized (e.g., remarks in the daily journal, gale and storm reports, fog reports, and abstract storm log) and because of the possibility of digitizing errors. Moreover, maximizing information preservation in one digitizing pass should prove more cost effective than handling the paper forms multiple times.

Specifically, two types of records were designed for keying. First, since each logbook form generally contains header information pertaining to an individual ship voyage, a single "voyage header record" was constructed to contain information such as the ship's name, captain, departure and destination, and observational metadata such as the method of observing sea surface temperature and any barometer correction. Second, "data records" were constructed to contain the actual observational data at each time and position. Each header record was assigned a unique voyage number, and the voyage number was also entered into each data record, so that the data records could be linked with the appropriate header during later data processing. This record management strategy minimizes keying, because the voyage header record is keyed only once and allows future "track checking" of the data records that compose a voyage.

In order to accommodate the wide variety of original form types, a multitude of formats had to be devised. It will require a large software development effort to convert these digitized data to

– continued on page 6

COADS, from page 5

a common format for inclusion in COADS. It should also be noted that it is quite labor intensive to prepare the original forms for digitizing and to ensure the accuracy of keying through a stringent quality control process.

Some additional unanticipated difficulties arose in development of the digitizing procedures. The digitizing for this project began on a climatological data management system known as CLICOM. This system was developed for XT-class personal computers, and requires that the header information and data be kept in separate files. This file management requirement and the inability of the CLICOM system to be expanded to increase production led to a conversion to the operational system used at NCDC for its routine keying operations. A separate system was developed to manage and quality control the data. NCDC's operational keying system also allowed for the voyage header and data records to be keyed consecutively and maintained in the same file thus simplifying the file management aspects of the process. Each file is processed through a quality control program that flags outliers and invalid codes so personnel with the original observational forms can determine whether the information was either keyed incorrectly or can be corrected using additional information on the form. It is often possible to correct columns that were transposed; miscalculated dates, times, and locations; incorrect barometer adjustments; mislabeled temperature scales; and other elements that were miscoded or miskeyed.

A companion project to digitize data collected at manned stations on ice floes ("ice islands") and ships overwintering in the ice pack in the Arctic Ocean, dating back to 1893 (Fig. 1), began in 1992 at NCDC as part of its contribution to COADS, but also in cooperation with the World Data Center-A Glaciology/National Snow and Ice Data Center (WDC-A/NSIDC), and the Polar Science Center (PSC), University of Washington. Because of their research interests, PSC and WDC-A/NSIDC provided data that they had collected for the missing periods and assisted in establishing the keying priorities.

Funding to start the keying was provided by the National Geophysical Data Center (NGDC). WDC-A/NSIDC is operated by the University of Colorado under contract to NGDC. Table 1 illustrates the available period of record of ice island data digitized or retrieved from several poorly documented digital data files under this initiative. Some data records are still missing.

Most of the ice island records that have been keyed to date came from T-3, often referred to as "Fletcher's Ice Island" after its discoverer, Joseph O. Fletcher, who was also instrumental in launching the COADS project a decade ago. It is appropriate that we will finally be able to add the highly valuable climate data from T-3 to COADS. Figure 2 illustrates recorded positions of T-3 based on the meteorological reports.

Other sources of unique data are being provided directly by various international organizations and governments (see Fig. 1). It is hoped that most of these sets can be completed in time for COADS Release 2 around the mid-1990s. Several nations are contributing to this effort:

- The Arkeologisk Museum in Stavanger, Norway, has obtained a grant to key over 600 late nineteenth century (1867-1890) Norwegian logbooks (approximately 500,000 records) in cooperation with the COADS project which provided keying instructions.
- Germany is keying 30,000 observations for the period 1887-1890.
- The Russian Federation provided approximately 3500 observations recently digitized from the Russian ship *Vitiaz* and other ships for the period 1804-1891 that appeared in the book by S.O. Makarov (1894). Figure 3 presents the geographical distribution of the observations.
- Negotiations led to an agreement between NCDC and the Chinese National Oceanographic Data Center to establish a cooperative keying project to digitize approximately 1 million ship reports in the Maury Collection (primarily between 1820 and 1860).

In addition, other important historical data sets remain undigitized and need to be considered for digitization as time and resources permit. These include:

- Japanese Kobe Collection. In the early 1960s Japan provided 623 rolls of microfilm from the Kobe Observatory in Japan containing merchant marine observations from the period 1890 to 1932 that have not been digitized, as well as those observations previously digitized from 1933 to 1961 (in COADS as decks 118 & 119). The microfilm also contains Japanese Navy Observations from 1903 through 1944, although few observations are available past 1941. The total

Table 1. Dates of observations digitized for T–3 (Fletcher's Ice Island) and AIDJEX. These dates are not all-inclusive as occasionally several observations or days of observations are missing.

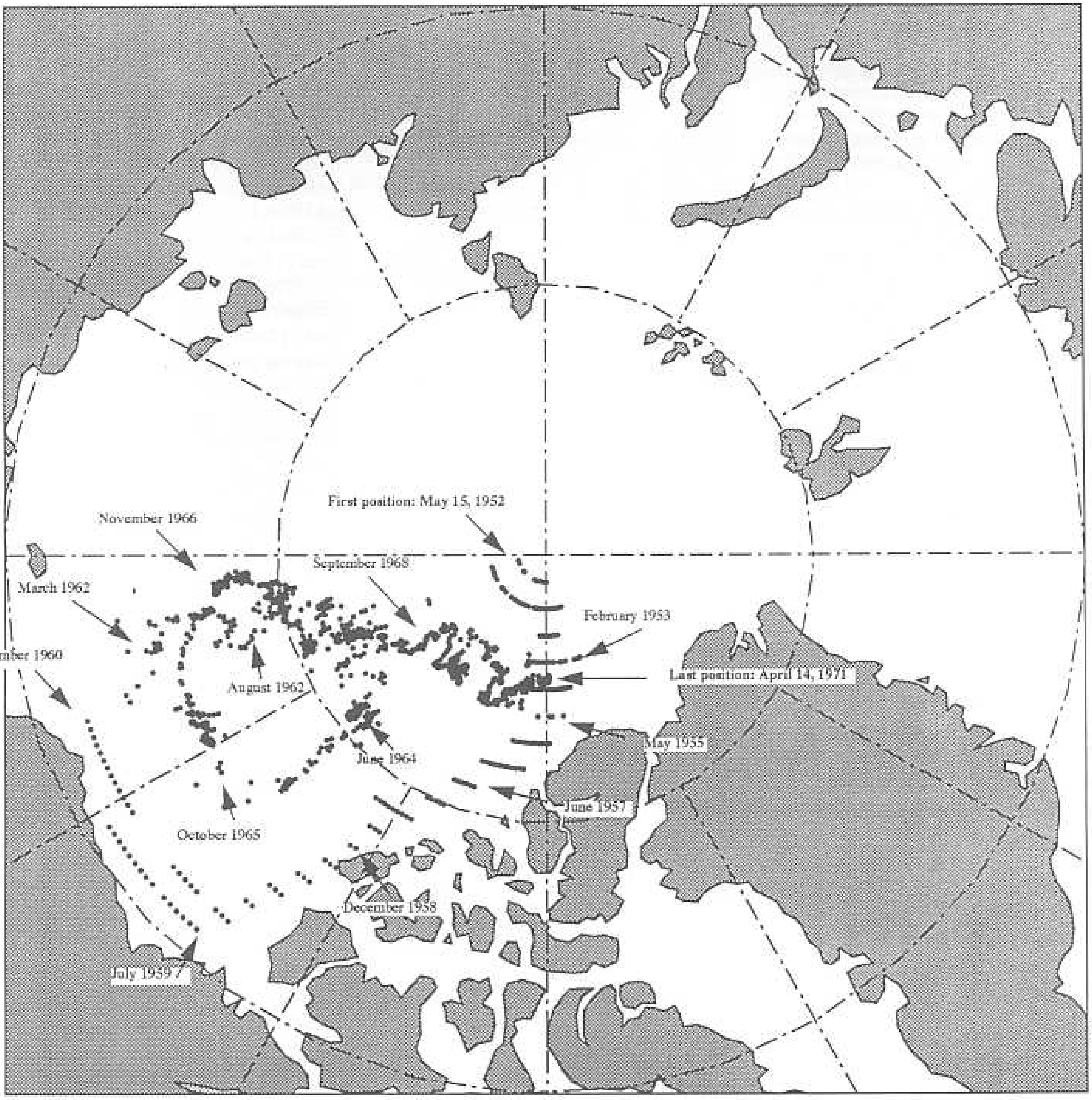
T–3 (Fletcher's Ice Island)

Source	Dates (day/month/year)
Deck 117 (Previously digitized)	15/04/1952 – 14/05/1954 01/05/1955 – 16/09/1955
WBAN Form 610-7	25/05/1957 - 31/03/1958
WBAN Form 10A & 10B	01/04/1958 - 25/10/1961
Summary of the Day WBAN Form 10A & 10B	01/04/1958 - 25/10/1961 *(January 1960 missing)*
Plain Language Teletype Messages	19/02/1962 - 12/06/1966 *(Used to fill in missing records)*
Teletype Observations	12/07/1963 - 30/06/1966
TD 3280 (Previously digitized)	13/06/1966 - 14/04/1971*

**Note: T–3 records not located after 14/04/71. Ice Island abandoned September 1974.*

AIDJEX (Big Bear, Blue Fox, Caribou, Snowbird)

April 1975 – April 1975



▲ **Figure 2.** Fletcher's Ice Island (T3) data available in digital form. Note that data prior to about 1961 had positions keyed only to whole degrees of latitude and longitude. (Figure courtesy of Ignatius Rigor, Polar Science Center, University of Washington.)

amount of undigitized data is estimated between five and six million reports (Uwai *et al.*, 1992; Elms 1992).

- Other undigitized U.S. ship logbooks. In addition to the Maury Collection (1796-1900) at least one other set of nineteenth century merchant marine logbooks housed in the U.S. National Archives needs to be considered for digitizing. Also, the Archives may possess undigitized U.S. Navy data.
- East India Company logbooks (located in the India Office Collection of the British Museum). The East India Company operated uninterrupted from 1599 to 1834, with its ships collecting a "wealth of information about the wind and weather" (Smith, 1925).

There is no doubt that other valuable marine data sets not listed here remain undigitized. The COADS project welcomes all participants who have the resources to provide additional data to join the effort to produce a more complete data set for use by the scientific community.

Impacts of changes in coding and observational procedures

Changes in coding and observational procedures require that data adjustments be made to assure data

– *continued on page 8*

COADS, from page 7

continuity. Marine weather observations over time have been recorded on a number of different form types as communications technology, the science of meteorology, and subsequent coding practices evolved.

The earliest observations from the East India logbooks, for example, predate establishment of the Beaufort Wind Scale and even the invention of the barometer in 1643. This means much of the information from this early period would be incompatible with current codes and methods of measurement. Similarly, the earliest reports in the Maury Collection contain wind direction, but no wind speed. Nevertheless, these data sets represent a unique resource of weather information having great value.

Unforeseen data continuity problems have been encountered in our endeavor to collect and digitize U.S. merchant (1912-1946) and Arctic ice island data. Some problems surfacing in the retrospective data are the result of established observing practices in effect at the time, while others result from observer error due to carelessness or lack of procedural knowledge/training. Date and time are recurring problems throughout the original records because observers often miscoded the date and time when converting from local ship watch to local time to Greenwich time and date. Also, the dry bulb and wet bulb temperatures occasionally seem reversed on the observational form. In most of these cases the data are corrected during the QC process.

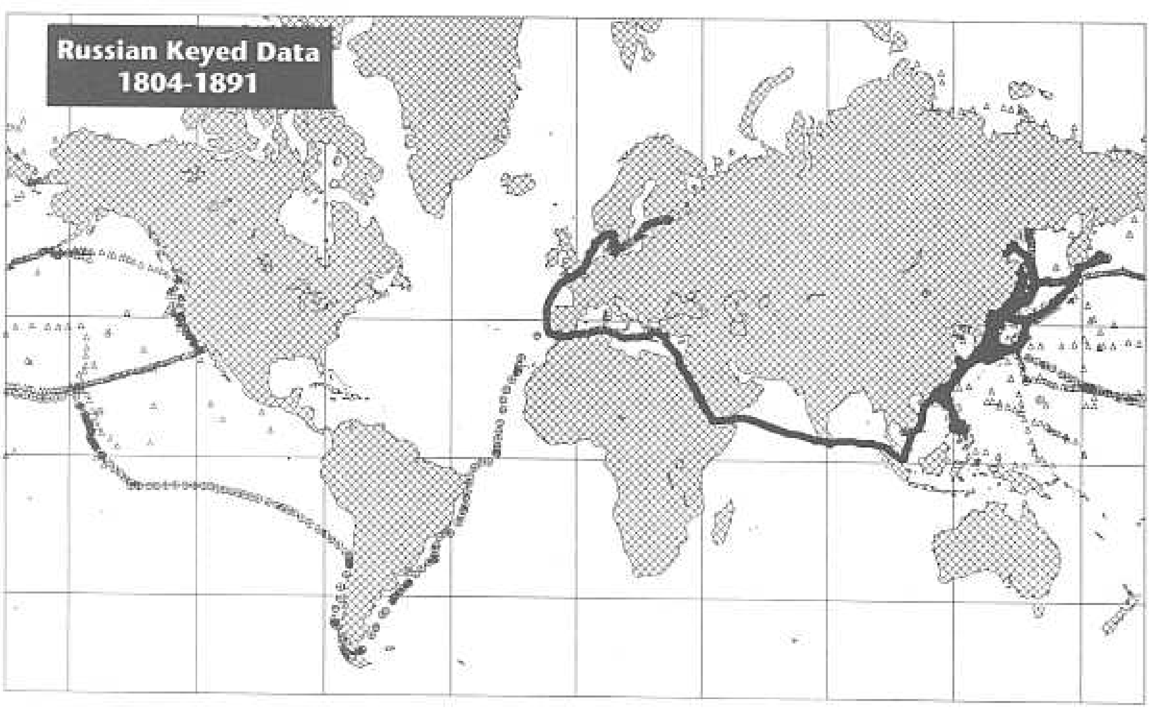
Changes in instructions to U.S.-recruited observers generally resulted in the issuance of a new edition of the standard Weather Bureau Instructions to the Marine Meteorological Observer, which later became known as Circular M (for "marine handbook"). Table 2 gives examples of such changes encompassing the period of the 1912-1946 merchant data. We believe the instructions prior to the first edition in 1906 were typically attached to the logbooks. The instructions remained fairly consistent from 1906 until the introduction of major international code changes in 1949. Even the international code changes of 1929—when great progress was made in standardizing methods of reporting weather observations, especially by radio—did not substantially alter U.S. coding practices.

Unfortunately, observers frequently did not adhere to the observing instructions. One example of a problem associated with the 1912-1946 merchant data is that wind directions were sometimes not coded in accordance with the U.S. instructions (32 point scale), occasionally appearing as 3 digits (e.g. 240). It is interesting to note that in the instructions in the 1938 and 1941 editions, the Weather Bureau broke somewhat with tradition by allowing the winds to be reported on the marine weather log form (1210A) in two acceptable codes:

"The direction of the wind may be entered in the appropriate column on Form 1210A either directly in terms of compass points or in code, according to the scale 01-32, in which 08=E, 16=S, etc. However, inasmuch as the wind direction must be coded in figures whenever an observation is transmitted by radio, it is customary for most observers to enter the code number, and this procedure is preferred by the Weather Bureau. Therefore, observers who have been accustomed to recording the wind direction directly in terms of compass points are urged to make a practice of using the numeric scale instead."

This practice required that different form types be developed before keying the data, but did not intrinsically lead to any data biases in the wind directions. However, this does possibly lead to biases in other observed elements such as cloud amounts by inadvertently establishing a precedence of substituting the radio code for the established code to be used on the observational form.

For example, the "total cloud amount" entered on the logbook forms appears to have come from two different codes, one developed for the log form entry in tenths (0-10) and the other a single digit code (0-9) adopted for radio transmission. Apparently, the observers, for convenience, often used the radio code instead of the form code to make their entry on



▲ **Figure 3** Russian S.O. Makarov Collection of data, 1804-1891, showing the observation locations of the *Vitiaz* (circles) and other ships (triangles) in the Collection.

Table 2. U.S. Weather Bureau Instructions to Marine Meteorological Observers.

		EXAMPLES OF CHANGES IN CODES AND OBSERVING PRACTICES					
EDITION	YEAR	Clouds	Wind Direction	Wind Speed	Weather	Sea Temperature	Mercury Barometer
Instructions attached to forms	Late 1800s	Proportion of clear sky in tenths	Mean magnetic direction	Beaufort force	Beaufort weather code	Bucket temperature	No corrections as read*
First	1906	Amount of clouds in tenths	True wind direction 32 point				
Second	1908						
Third	1910						
Fourth	1925						No corrections as read**
Fifth	1929						
Sixth	1938				00-99 present weather code	Bucket/ injection	
Seventh	1941						
Provisional	1949	Eighths	36 point	Knots			Corrected to sea level

** Indications are that this was probably the practice, however, it is not clearly evident from reading the instructions.*
***Corrections to be added for radio transmissions.*

NOTE: Updates after 1941 include Circular M (editions 8–12) in 1950, 1954, 1959, 1963, 1964, and NWSOH No. 1 editions in 1971, 1982, and 1992. A new update will be required for the November 2, 1994 approved WMO code changes.

the log form. Since there is generally no way of distinguishing which code was used, this innocent practice introduced an observational bias (e.g., overcast skies coded as 10 in the form code and as 8 in the radio code) that can only be corrected statistically after making certain assumptions regarding overall cloud distributions. (Table 3).

Careful checks will be required for conversion of digitized sea level pressure values into a common COADS format because of the many changes in instrumental or reporting procedures. In a number of cases all the necessary information needed to make these checks is not available on the original form. Before the time of radio transmissions, observers using mercurial barometers were instructed not to make the corrections for temperature and gravity, but simply enter the value as read; the Weather Bureau made the necessary corrections upon receiving the forms.

Some of the early aneroid barometers issued did have an attached thermometer for convenience in reading the ambient air temperature (dry bulb). It is difficult to imagine, however, that the mariner would have mounted the valuable barometer in the open air exposed to all the weather elements and not in the protection of the cabin. There is also a large number of attached thermometer entries indicated to have come from an aneroid barometer that have the same value as the dry bulb temperature. This can be corrected by careful QC.

Many of the U.S. merchant marine logbook forms located in the archives for the period 1912-1946 contain only one observation per day at 0000 UTC. This is in contrast both to the mid- to late 1800s when U.S. observations were reported every two hours and to contemporary international ship observations, which are generally reported every six hours (0000, 0600, 1200, 1800 UTC).

Starting sometime before 1906, the Weather Bureau required that radio reports be sent twice or even four times a day but requested only the 0000 UTC observation be sent by mail. The logic behind the practice of reporting only the 0000 UTC observation on the forms

– continued on page 10

Table 3. Total Cloud Amount

Radio transmission code figures	Proportion of sky covered (in tenths)	Code on observation form
0	0	0
1	Less than 0.1	*
2	0.1	1
3	0.2 to 0.3	2,3
4	0.4 to 0.6	4,5, 6
5	0.7 to 0.8	7,8
6	0.9	9
7	More than 0.9 but with openings	*
8	Sky completely covered with clouds	10
9	Sky obscured by fog, dust storm, or other phenomenon	*

*No corresponding code.

COADS, from page 9

is explained in the 1906 edition. The Weather Bureau felt that with the advent of weather forecasting as a science, mariners could best determine which route to take based on conditions actually encountered (referencing the daily synoptic charts) rather than on average conditions.

The 0000 UTC observational practice can bias the digital database, however. For example, at certain longitudes all reports are observed near the average diurnal cycle for maximum heating, while at other longitudes reports are observed near maximum cooling, with reports from intervening longitudes falling somewhere in between. These biases can be statistically corrected.

One partial solution to correcting the bias of once-daily logbook observations would be to supplement them with radio messages. However, few of these radio messages (often garbled) are available in their original format today. Reasons for this include:

- lack of storage space,
- lack of economical or viable technology at the time for archiving the information on film or digital media, and
- deterioration over time of the teletype paper.

A large number of these messages were plotted on the Northern Hemisphere Charts and are available dating back to 1899. These observations would be very time consuming and expensive to digitize, and a number of the elements would have to be estimated because of the coarseness of the plotting code. For example, wind speeds were plotted only to the nearest 5 knots, wind direction would have to be based on direction of the plotted wind shaft, and ship position estimated from the location of the plotted station model.

Remaining work

Large tasks remain before COADS Release 2 can be completed. One of the largest will be the conversion of all the additional data sets to the common format used for COADS. The conversion programs must be designed so that data elements are preserved for future research even though they may not be compatible with the COADS Release 2 data format and statistics. To provide adequate metadata will also be a sizeable undertaking. Lastly, all the data sources will have to be merged, duplicate observations eliminated, and quality control applied to the dataset with erroneous or suspect entries flagged. Those elements passing the quality control will then be used to compute the COADS Release 2 statistics which, along with the observations, are planned for general availability in the mid-1990s.

Acknowledgements

COADS is the result of a continuing cooperative project between the National Oceanic and Atmospheric Administration (NOAA)— specifically its Environmental Research Laboratories (ERL), National Climatic Data Center (NCDC), and Cooperative Institute for Research in Environmental Sciences (CIRES, conducted jointly with the University of Colorado)—and the National Science Foundation's National Center for Atmospheric Research (NCAR). The NOAA portion of COADS is currently supported by the NOAA Climate and Global Change Program and the NOAA Environmental Services Data and Information Management (ESDIM) Program.

We also acknowledge the many individuals who have worked over the years through their national meteorological services and the World Meteorological Organization (WMO) and its predecessors to see that global scale marine observations were collected, preserved, and distributed for the use and benefit of all.

References

Elms, J.D., 1992: Status of NCDC Keying of Historical Marine Data. *Proceedings of the International COADS Workshop, Boulder, Colorado, 13-15 January 1992.* H.F. Diaz, K. Wolter, and S.D. Woodruff, Eds., NOAA Environmental Research Laboratories, Boulder, Colo., 37-45.

Jenne, R.L., 1992: The Importance of COADS for Global Reanalysis. *Proceedings of the International COADS Workshop, Boulder, Colorado, 13-15 January 1992.* H.F. Diaz, K. Wolter, and S.D. Woodruff, Eds., NOAA Environmental Research Laboratories, Boulder, Colo., 9-15.

Makarov, S.O., 1894: *Vitiaz in Pacific Ocean,* Volume I, St. Petersburg.

Slutz, R.J., S.J. Lubker, J.D. Hiscox, S.D. Woodruff, R.L. Jenne, D.H. Joseph, P.M. Steurer, and J.D. Elms, 1985: *Comprehensive Ocean-Atmosphere Data Set; Release 1.* NOAA Environmental Research Laboratories, Boulder, Colo., 268 pp. (NTIS PB86-105723).

Smith, H.T., 1925: Marine Meteorology, History and Progress. *The Marine Observer.*,Vol.II, No. 15, 33-35.

Uwai, T. and K. Komura, 1992: The Collection of Historical Ships' Data in Kobe Marine Observatory. *Proceedings of the International COADS Workshop, Boulder, Colorado, 13-15 January 1992.* H.F. Diaz, K. Wolter, and S.D. Woodruff, Eds., NOAA Environmental Research Laboratories, Boulder, Colo., 47-59.

Woodruff,S.D., R.J. Slutz, R.L. Jenne, and P.M. Steurer, 1987: A Comprehensive Ocean-Atmosphere Data Set. *Bull. Amer. Meteor. Soc.*, 68, 1239-1250.

Woodruff, S.D., S.J. Lubker, K. Wolter, S.J. Worley and J.D. Elms, 1993: Comprehensive Ocean-Atmosphere Data Set (COADS) Release 1a: 1980-92, *Earth System Monitor,*Vol 4, No. 1. ■

CD-ROM Data Products of the National Geophysical Data Center

This summary from the NGDC is the second in a series listing CD-ROMs currently available from the NOAA national data centers. They are presented as a sequel to the article on CD-ROM activities at the centers presented in the June 1993 issue of the Earth System Monitor.

MARINE GEOLOGY/GEOPHYSICS DATA

- **Deep Sea Drilling Project (DSDP)**
  This 2-volume set contains all digital marine geological and geophysical data compiled by the DSDP. Volume I holds sediment/hardrock data and reference files. Volume II holds downhole logs and underway geophysics data. NGDC-developed access software is available for PC and Macintosh platforms. For more information, please call 303-497-6339.

- **Ocean Drilling Program, Version 2.0**
  This 2-volume set contains all digital marine geological and geophysical data (exclusive of downhole logging) available for Legs 101 through 129 of the ongoing ODP. Both PC and Macintosh software are included. For more information, please call 303-497-6339.

- **Gulf of Mexico GLORIA Data**
  This disc of marine data contains long-range sidescan sonar records including three sets of imagery data for the Gulf of Mexico Exclusive Economic Zone. The data were collected by the U.S. Geological Survey, which also contributed to the development of the CD-ROM. For more information, please call 303-497-6338.

- **NOAA & MMS Marine Minerals**
  This CD-ROM contains the results of a 9-year cooperative effort between NGDC, NOAA's National Ocean Service, and the Minerals Management Service to compile a comprehensive, computerized bibliography and geochemical database on offshore hard mineral resources. Both PC and Macintosh software are included. For more information, please call 303-497-6339.

- **GEODAS CD-ROM**
  This data set consists of two CD-ROM discs and includes custom menu-driven, access software developed by NGDC. Included on the discs are "MGD77" header and data records from 3501 cruises stored in a compacted binary format. The software provides access to over 4 gigabytes of worldwide marine geophysical trackline data including bathymetry, magnetics, gravity, and seismic shot-point navigation. Access programs allow users to search the data by a combination of the criteria: data type, institution, platform, year of cruise, cruise identifier, or date the cruise was added to the system. The software produces trackline plots on the user's computer screen or as a PostScript file for printing. For more information, please call 303-497-6338.

- **Global Relief Data**
  This disc contains several digital gridded databases of global topography and bathymetry at various scales, digitized continental coastlines, gridded oceanic gravity anomalies, ocean floor gazetteers, and digital topographic relief images in a wide variety of formats. Software or program code for PC-DOS, Macintosh, and UNIX workstations is provided to access the major topographic and geographic databases. Macintosh users will find a special "Mac only" area of the disc devoted to topographic imaging, including software to generate full-color relief images. For more information, please call 303-497-6338.

Ordering Information

To receive further information about CD-ROM data products of the National Geophysical Data Center or information on availability and pricing, please contact:

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

Telephone: See listed numbers
Fax: 303-497-6513
Internet: info@mail.ngdc.noaa.gov
Voice/TDD: 303-497-6958

SOLAR/TERRESTRIAL DATA

- **Flares in Hydrogen-Alpha**
  This is a "write-once" CD-ROM holding ASCII text files of solar flare records from ground based observations grouped by year. Data include reports (1938-present), events (1955-present), patrol (1955-present), and indices (1955-1980). For more information, please call 303-497-6761.

- **GOES Space Environment Data**
  This CD-ROM contains X-ray, magnetometer, and energetic particle data measured by NOAA's Geosynchronous Operational Environmental Satellites (GOES) at 1- and 5-minute resolution. These data are an excellent resource for studying solar activity and its effects in near-earth space. For more information, please call 303-497-6761.

- **Neutron Monitor Data**
  This "write-once" CD-ROM contains ASCII text files of cosmic ray data from neutron monitors operating in many parts of the world. Some data goes back to 1953. For more information, please call 303-497-6761.

- **Regions of Solar Activity**
  This "write-once" CD-ROM contains ASCII text files of records of calcium plage regions and sunspot groups. Records are grouped by year and by station. Plage observations are from Big Bear (1981-1987), Hale (1979-1981), McMath (1942-1979), AGD Table (1969- 1982), and Indices (1970-1987). Sunspot observations are from Catania (1978-1987), Greenwich (1874-1982), Mt. Wilson (1962-1991), Rome (1958-1989), and U.S. Air Force (1981-1991). For more information, please call 303-497-6346.

- **Solar Radio Bursts**
  This "write-once" CD-ROM holds ASCII text files of solar emissions recorded since 1960 at ground observation stations which are available at discrete and sweep frequencies. For more information, please call 303-497-6761.

– continued on page 12

NGDC CD-ROMs, from page 11

SOLID EARTH AND OTHER DATA

- **Gravity CD-ROM**

This data set includes station data, gravity networks, regional surveys, and gravity anomaly grids (both regional and global). The boundary data include World Vector Shoreline and World Data Bank II. Access software accompanies this CD-ROM. For more information, please call 303-497-6120.

- **Minnesota Aeromagnetics**

This 2-volume CD-ROM set contains Minnesota aeromagnetics data contributed by the Minnesota Geological Survey, U.S. Geological Survey, U.S. Department of Energy, and Canadian Geological Survey. Software to access the trackline data (surveys or individual tracks) or grids (whole or as subsets) is provided. For more information, please call 303-497-6128.

- **Experimental Calibrated GVI**

This CD-ROM presents two versions of an experimental Normalized Difference Vegetation Index (NDVI) based on the AVHRR (Advanced Very High Resolution Radiometer) data. The biweekly data are successive weeks of screened NDVI data, which is temporally composited, based on maximum NDVI. The monthly data are two biweekly periods that best fit a given calendar month and are reprojected to latitude-longitude. For more information, please call 303-497-6125.

- **Geophysics of North America**

The majority of this land and marine geophysical data is from the Geological Society of America's "Decade of North American Geology."It includes magnetics, gravity, earthquake seismology, thermal aspects, and stress data. Also available are satellite imagery data, topography, and additional grids of magnetics and gravity. A special feature of this product is its color graphics capability. Access software is furnished on a floppy disk. For more information, please call 303-497-6591.

- **Global Ecosystems Data**

This CD-ROM includes selected data on global ecosystems, vegetation (including vegetation index), climate, topography, soils, and other data. These data are on several different grids, from 2 degrees to 2 minutes. Vector data for coastlines and other features are also provided. For more information, please call 303-497-6125. ■

NODC Taxonomic Code, from page 2

and the file TAXPARTS.TXT, which describes the "segmented" version of the NODC Taxonomic Code

The segmented version of the NODC Taxonomic Code presents the entire code as a series of 41 separate files, each of which is small enough to be copied to a high density diskette. Each file contains members of a single major group of organisms (e.g., protozoans, fishes, birds) or several different groups of more-or-less closely related organisms. Some of the larger taxonomic groups are divided among several files:

- Magnoliophyta 3 files
- Mollusca 2 files
- Crustaceans 2 files
- Insects 12 files
- Fishes 3 files

To order and for further information

Version 7.0 of the NODC Taxonomic Code is available on CD-ROM, as well as on magnetic tape. Individual sections of the code can also be provided on high density diskettes. For information on costs and ordering procedures, please contact:

National Oceanographic Data Center
User Services Branch
NOAA/NESDIS E/OC21
1825 Connecticut Avenue, NW
Washington, DC 20235
Telephone: 202-606-4549
Fax: 202-606-4586
Omnet: NODC.WDCA
Internet: services@nodc.noaa.gov

Requests for technical details about the NODC Taxonomic Code, as well as requests for new codes or suggested modifications of the code should be submitted to:

J.D. Hardy, Jr.
National Oceanographic Data Center
NOAA/NESDIS E/OC12
1825 Connecticut Avenue, NW
Washington, DC 20235
Telephone: 202-606-4636
Fax: 202-606-4586
Internet: dave@nodc.noaa.gov ■

U.S. DEPARTMENT OF COMMERCE
National Oceanic and Atmospheric Administration
Publication Distribution Facility
1315 East-West Highway
Silver Spring, MD 20910
ATTN: Earth System Monitor

Address Correction Requested
OFFICIAL BUSINESS
Penalty for Private Use $300

Third Class Rate

40