2007

# NOAA's Climate Database Modernization Program: Rescuing, Archiving, and Digitizing History

Lesley-Ann Dupigny-Giroux
*University of Vermont*, ldupigny@uvm.edu

Thomas F. Ross
*NOAA/NESDIS/NCDC*

Joe D. Elms
*NOAA/NESDIS/NCDC*

Raymond Truesdell
*Information Manufacturing Corporation*

Stephen R. Doty
*Doty Data Services*

# NOAA'S CLIMATE DATABASE MODERNIZATION PROGRAM

## Rescuing, Archiving, and Digitizing History

BY LESLEY-ANN DUPIGNY-GIROUX, THOMAS F. ROSS, JOE D. ELMS, RAYMOND TRUESDELL, AND STEPHEN R. DOTY

Historic weather, climate, and ocean observations from as far back as the mid-1700s are being made easily available on the Internet for use in studying global climate variability and change and for helping to improve mitigation and response. The National Oceanic and Atmospheric Administration (NOAA) Climate Database Modernization Program (CDMP) began in 2000 with a major emphasis on imaging and keying worldwide climate and environmental records from the eighteenth through the twentieth centuries. This multimillion dollar program is an ongoing effort to process data from the United States and elsewhere, improve its access, and maintain a permanent data archive. The CDMP, now in its eighth year, is a product of the extraordinary efforts of both NOAA personnel and the private sector. Initiated by the U.S. Congress, the program is managed by NOAA's National Climatic Data Center (NCDC). The keying, imaging, and database development needed for the CDMP projects have created new private-sector jobs in several economically challenged areas in West Virginia, Kentucky, and Maryland. An invaluable NCDC on-site contractor prepares much of the data for shipment and performs extensive quality control on returning data products.

**AFFILIATIONS:** DUPIGNY-GIROUX—University of Vermont, Burlington, Vermont; ROSS AND ELMS—NOAA/NESDIS/NCDC, Asheville, North Carolina; TRUESDELL—Information Manufacturing Corporation, Fairfax, Virginia; DOTY—Doty Data Services, Arden, North Carolina
**CORRESPONDING AUTHOR:** Lesley-Ann Dupigny-Giroux, University of Vermont, Department of Geography, 200 Old Mill Building, 94 University Place, Burlington, VT 05405-0114
E-mail: ldupigny@uvm.edu

DOI:10.1175/BAMS-88-7-1015



The CDMP acquires, digitizes, and provides access to the climate and environmental data held in various national and international archives. In earlier data rescue efforts of U.S. weather and climate records, the National Archives and Records Administration (NARA) microfilmed and catalogued their 1819–92 climatological records in 1952. These data were recorded using the many different formats existing at that time, including monthly and quarterly reports received by the U.S. surgeon general from 1819 to 1859; monthly reports of volunteer observers for the Smithsonian Institution from 1849 to 1859; weekly and monthly reports of Signal Office and Weather Bureau stations from 1870 to 1892; and monthly reports from volunteer observers from the Signal Office and Weather Bureau from 1874 to 1892. All of these records have been scanned and are available as part of the CDMP, with image access available via the Internet. Selected records from this collection are being keyed and made available in digital format in various NCDC databases. CDMP supports NOAA's core mission to archive, store, and manage environmental data and information under data stewardship for the United States.

**DATA ACQUISITION, QUALITY CONTROL, AND ARCHIVING.** Data acquisition begins with original or microform images from either national or foreign archives. CDMP contractors scan these images using various methods to ensure the best possible image quality. These images are then indexed so they can be retrieved through a series of inquiries based on the various index fields available. Then, depending on the goals of a data acquisition project, the imaged data may also be keyed by contractors, with the keyed data integrated into NCDC's digital database.

Once a database or data series has been imaged, the data are made available online. The indexed images are loaded to the Web Search Store Retrieve Display

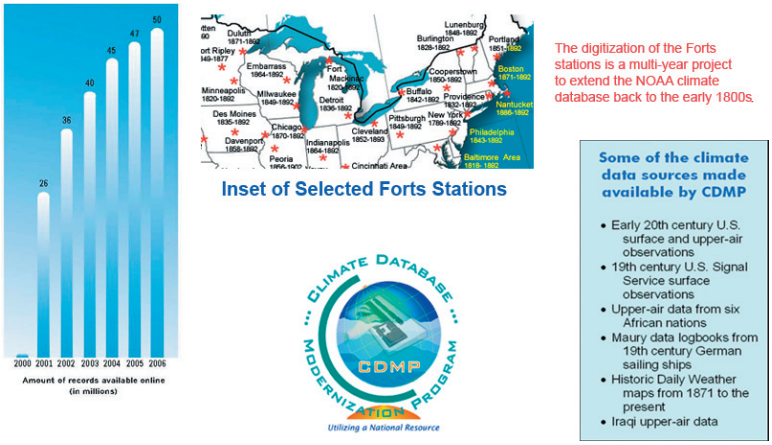


**Fig. 1. Summary of the CDMP, showing annual totals of keyed records, the spatial distribution of a portion of the FORTS project, and selected examples of data resources.**

(WSSRD®) system developed by one of the CDMP prime contractors. The WSSRD system is an online, password-protected, Web-based image-retrieval tool. It is a searchable system for displaying images, text, PDF files, and other documents that are indexed and stored in the database, available at www.ncdc.noaa.gov/oa/climate/cdmp/wssrd.html. Interested persons should contact the CDMP project team to request access to WSSRD (e-mail: *Cynthia.B.Karl@noaa.gov*).

All climate and environmental data deemed important for the climate record by the NCDC are candidates for this program. The length of the records imaged by CDMP depends upon the data type and availability. Stations are tested for meteorological consistency before their data are imaged and keyed. Thus, some data are available for discrete periods only, while others extend back to the eighteenth century. For example, daily records for Cambridge, Massachusetts, begin in 1742 and are part of the "Historic Climate Diaries and Journals" cabinet on WSSRD.

Numerical data and metadata extracted from these imaged forms through the CDMP will become part of NOAA climate and environmental databases. Current data-rescue projects involve all five NOAA Operational Line offices, Regional Climate Centers, State Climatologists, the U.S. Air Force and U.S. Navy, the World Meteorological Organization, and international meteorological services. NOAA's global databases have been extended backward, and data gaps filled where possible. Major projects include imaging and keying daily cooperative and hourly weather-station data, daily weather maps back to 1871, upper-air observations and forms, NCDC serial publications, station-history reports, hurricane reconnaissance information, sea and subsurface temperature information, ship logbook entries, ionospheric observations, arctic sea-ice charts, glacier photos and volcano information, and imaging Defense Meteorological Satellite Program filmstrips. More than 32 million hourly surface observations have been keyed, extending the database to the late 1920s, when approximately 150 U.S. airport stations were established. Daily synoptic surface observations made at city Weather Bureau offices and Army Signal Service sites in the mid-1800s have also been added (Fig. 1). These historical, keyed, surface hourly and daily data undergo extensive data conversion at the Northeast and Midwest Regional Climate Centers prior to being quality-controlled and merged with existing databases (e.g., the DS3206 Integrated Surface Hourly database) at NCDC. Other keyed data are processed via existing quality control and quality-assessment routines at NCDC before being made available. Merged NCDC datasets are available through NCDC's online store, with free access from *.edu* and *.gov* domains.

Additional CDMP projects include imaging and keying data on cetaceans, marine fish species, shoreline and nautical charts, water level, coast pilot charts, and tide and prediction tables. International projects include imaging and keying upper-air data from six African countries, ship observations from Canada, and surface observational records from Uruguay, Nicaragua, and Mexico.

**APPLICATIONS AND POSSIBILITIES.** The CDMP has improved access to valuable historical data sources, whose analogue storage on such media as paper or film was in danger of being lost forever without proper preservation. The keying of the data and digital preservation of the historic record will benefit not only NOAA, but also researchers and data users in government, commerce, industry, science, education, engineering, and national defense. The ability to access station inventory data at the page level will facilitate better comparisons with proxy data and the reanalyses of standard variables.

Potential applications of these data are unlimited. Ionospheric data represent the best long-term mea-

surements of the critical regions that are important to the operation of radio communications, surveillance radars, satellite-to-ground communications, and data-driven space environmental models. The digitizing of mechanical bathythermograph data measurements of water temperatures at various ocean depths will help to fill important gaps in the sea temperature database.

The proper usage of the newly digitized data depends upon an assessment of the history and the quality of the observations. The CDMP has undertaken the documentation of the weather-observing history at many long-term sites (FORTS data) around the country. Existing gaps and long-standing discrepancies have been revealed in the metadata in the station history records. For example, the addition of Zadock Thompson (1832–56) (Fig. 2) and McKendree Petty's (1858–64) records in Burlington, Vermont, corrected the entry on the Weather Bureau Station History form of 1 December 1953, which had indicated an unknown location for the Burlington observations during this period. Similarly, the address of Walter Benton Gates, a volunteer observer in Burlington, Vermont, has been corrected from City Hall Park to 55 Elmwood Avenue (a few blocks north of City Hall Park). Finally, previously unknown observers have also been identified.

Another benefit of the historical investigation is the uncovering of additional observational data previously unavailable to the climate community. This opportunity is especially enhanced when local climatologists assist in the investigation. Finding data from a data-sparse era, such as the earliest days of organized observing in the 1820s and for the Civil War years, will add greatly to the national climate database. For example, in Mobile, Alabama, several

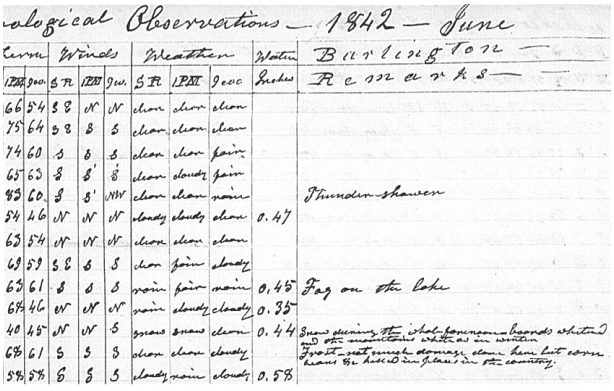

**Fig. 2. A June 1842 excerpt from the Zadock Thompson diary, "Meteorological Observation Made Mostly at Burlington Vermont in Latitude 44°28' North and Longitude 73°11' West from Greenwich 256 feet above Lake."**

additional military locations were found to have records in the 1820s, and observations taken in the 1860s were found in the local newspaper.

**CONCLUDING REMARKS.** The CDMP supports NOAA's effort to provide access to its vast archive of climate and environmental data. It is an example of a successful government program working with the private sector to recover valuable climate and environmental data and to create jobs in various sectors of the national economy. In this effort, CDMP serves society's need for weather and water information. CDMP projects also help improve the understanding of climate variability and change and, thus, our ability for mitigation and response.

**ACKNOWLEDGMENTS.** This collaboration was facilitated by the NCDC State Climatologist Exchange Program.