NOAA’s Climate Database Modernization Program - More than 10 Years of Data Rescue Activities

ACRE Meeting 3-5 November 2010 -- Tom Ross, CDMP Program Manager
International Partners

- Atmospheric Circulation
  Reconstructions over the Earth

- IEDRO
  International Environmental Data Rescue Organization

- Climate Database Modernization Program (CDMP)
  Major Historic International Data Rescue Projects
Climate Database Modernization Program
The Climate Database Modernization Program (CDMP) goal is to preserve and make major climate and environmental data available via the World Wide Web.

The CDMP supports NOAA’s data stewardship through all of its activities.
Data Received from Many Sources

Forecast Warning Analysis
Voluntary U.S. Observers
Global Weather Reports
NCEP Weather Charts & Models
Ship, Buoy Reports
Rocketsonde
Weather Balloons
Storm Data
Doppler Radar
(GOES, POES, NPOESS, and many other Satellites
Aircraft Observations
Wind Profiler
Airport Weather Reports (ASOS)
U.S. Climate Reference Network
Climate Models
Climate Database Modernization Program: 2000-2010

Over 14 terabytes of climate data now digitized

- 54 million weather and environmental images online
- Hundreds of millions of records digitized now online
- International data access and rescue activities
- 97 current NOAA climate/environmental rescue projects

July 1st, 1842 hourly weather data from Washington, DC, imaged and digitized through the CDMP Program
Climate Database Modernization Program

Highlights/Past Successes

2010 Major NOAA CDMP Data Recovery Tasks

35 NCDC Tasks

Over 410 million surface observations will be added to NCDC’s historic synoptic database by the end of 2010

67,022 missing images 1950-1980 are now on-line on the IPS system

Working with the NC state climatologist office to add 43,000 additional monthly cooperative station data

- Rescue collections of surface, upper air data, cooperative station data, and historical diaries and journals. Additional projects include “Forts” (mid 19th century) data, NCDC historic publications, various national and international daily and monthly data for inclusion into research datasets (GHCN, IGRA, ICOADS, etc.)
Climate Database Modernization Program

Highlights/Past Successes

2010 Major NOAA CDMP Data Recovery Tasks

5 NODC Tasks

- Rescue collections of ecosystem surveys along California coast
  - Beach Watch Program
  - Common Murre Restoration Project
  - California Kelp Resources Project
- Scan ~110K survey slides and log sheets

Highlighted tasks include: California marine ecosystem surveys, NODC metadata, World Ocean Database project, NOAA library and film transfer tasks.
Climate Database Modernization Program

Highlights/Past Successes

2010 Major NOAA CDMP Data Recovery Tasks

16 NOS tasks

Highlighted tasks: Image nautical charts & historical coast pilots, vectorize and geo-reference shoreline charts and photographs, image and key water level gauge and tidal benchmark records, environmental sensitivity maps, water and tide level data, fishery management and catch tracking tasks

Same area on a NOAA nautical chart. Comparing these sources, created at different times, provides information on the rate of change in the coastal zone, which aids in the design of coastal zone mapping projects.

High-altitude aerial photography, Hudson River, NY.
Climate Database Modernization Program

Highlights/Past Successes

2010 Major NOAA CDMP Data Recovery Tasks

4 OAR/AOML Tasks

- Surface observations have been taken at Puerto Baquerizo Moreno (9°S, 89.6°W) on Isla San Cristobal in Ecuador’s Galapagos Islands since 1967. There are no stations within 1,000 km which regularly report surface data for this period. This station is located in the heart of the East Pacific Cold Tongue.

The East Indiaman Warley (1795), as depicted by Robert Salmon. Courtesy of the National Maritime Museum, London

Image and key historic European ship logbooks for incorporation into ICOADS, hurricane reconnaissance information, San Cristobal keying and various research and metadata publications.
Climate Database Modernization Program

Highlights/Past Successes

2010 Major NOAA CDMP Data Recovery Tasks

13 NGDC Tasks

- Rescue collections of ionospheric data, various marine geophysical rescue datasets, glacier and tsunami tide data, and solar data rescue projects, cosmic ray and DMSP film data rescue projects. Additional tasks include marine and Lacustrine records of Climate change, glass plate data rescue, solar observations.
Climate Database Modernization Program

Highlights/Past Successes

2010 Major NOAA CDMP Data Recovery Tasks

15 NMFS Tasks

Sample tasks include: Imaging and keying data on cetaceans, fish eggs/population, coral reefs, plankton data recovery, ecosystem surveys, and Turtle Exclusion data.

Bycatch Reduction Engineering Program to “provide information and outreach”…that will encourage adoption and use of technologies.
Climate Database Modernization Program

Highlights/Past Successes

2010 Major NOAA CDMP Data Recovery Tasks

Government Joint - Sponsored Tasks

• Major rescue projects include: USDA Bird Phenology dataset of 1-3 million cards documenting bird migrations; over 250,000 historic snow survey data records from many U.S. western states; keying data from 23 Automated Weather Stations (varying periods from 2-13 years) in the Navajo Nation in Arizona (Hourly and daily) from 1986-1999 - original computer data were lost; Hawaiian island precipitation data project.
NCDC Non-Digital Data Archive

**Manuscript / Autograph***
103 Million Pages stored in 120,000 boxes

* Located at Asheville; additional paper records located at the Federal Records Center in Georgia that will be inventoried and prioritized for digitization

**35mm & 16mm Film**
125,129 Rolls

**Microfiche**
860K fiche containing 51 million pages

---

**Percent digitized**
(Keyed or imaged)

- 72%
  - (78 million)
  - 1,500 forms/day

- 1.7%
  - (2,105 reels)
  - sporadic

- 10.0%
  - (86,000 fiche)
  - 3,372 forms/day

---

* Located at Asheville; additional paper records located at the Federal Records Center in Georgia that will be inventoried and prioritized for digitization
Long-term Archive Tasks

**Satellite Data**
TIROS, NIMBUS, ITOS, ATS, ESSA, GOES, more
Mainly 1964-1990
± 1,300 boxes of raw images in many film formats
  - Film ranges 16mm-70mm
  - Black & white images, negatives
The many formats makes this a challenging conversion task

Scientists in RSAD need to access these holdings for use
Climate Database Modernization Program

Data Integration Projects

NCDC’s digital data for hourly surface observations generally began around 1948. CDMP has imaged over 2 million forms and keyed over 400 million records as part of this keying project. This has added over 50 years of hourly/synoptic data for over 225 stations.
The CDMP “Forts” project is extending daily data records from the beginning of the Weather Bureau era (circa 1892) as far back as the 1780’s. Over 390 stations’ data have already been keyed. CDMP’s Forts team will prepare approximately 75 more stations for keying in the upcoming year.

Keyed data for all the CDMP Forts stations are available online through the Midwestern Regional Climate Center (MRCC). In addition, close to 200 stations’ data have passed Quality Control tests applied at MRCC. Over 175 of these stations are available from NCDC and MRCC via an FTP download.
International Simultaneous Marine Observations-  
Multi year Task - Start  2011

• Created by the International Meteorological Congress, Vienna 1873  
  • Published by the U.S. National Weather Service from 1875-1887  
  • At peak ~ 450 US land stations and ~ 600 vessels reported  
  • ~ 30 countries included, from Algeria to the U.S.  
  • ~ 5,000,000 daily land and marine obs.  
• Maritime Observations POR 1/1877-6/1884  
  • Beginning with a few Naval reports to ending with over 600  
    vessels reporting, average of 75/day  
  • ~ 205,000 daily observations  
• Elements  
  • Locations, pressures, temperatures, winds, clouds, precipitation,  
    sea and weather conditions.  
• Caveats  
  • Various types of elements, corrections, reductions, scales, etc.
Climate Database Modernization Program
Sailing Ahead with Historical Marine Observations
Data Integration Projects

Observations digitized by CDMP since 2007 (~2.27 million) and current projects to help fill temporal and spatial gaps to better enhance marine climatologies and reanalysis efforts. Red arrows represent the temporal range of each collection.

**English East India Company Logbooks**
- Period of Record: 1790-1834
- 285K Observations
- Digitization goal: 2009

**US Lightship Observations**
- Period of Record: 1891-1982
- 430K Observations
- Digitization goal: 2009

**Greenwich Mean Noon / Simultaneous Ship Observations**
- Period of Record: 1874-1947
- 2.6 Million Observations
- Digitization goal: 2011


R.M.S. Laconia (1912), courtesy of the Steamship Historical Society of America, Inc.

Diamond Shoals Lightship LV 71, US Coast Guard

Pollock Rip Lightship LV 114, US Coast Guard
Data from these international rescue projects will be added to various NOAA/NCDC surface, marine and upper air databases to aid in research and applied climatological studies and applications.
Climate Database Modernization Program

Data Integration Projects

58 of 95 Observatories keyed (29 different formats)
- 35 coop stations yet to be keyed. Total keying of 13.1 million daily observations (430,000 forms) should be completed by end of 2012 - Data period roughly 1876-1972
- approximately 55% of the data has been keyed. Data will used in international North American Drought monitoring activities with Canada and Mexico and incorporated into Global Historical Climatological Network (GHCN) at NCDC

Daily data keyed for 30 stations covering 1890’s -1920’s - total of 6092 station months. Data incorporated into GHCN

Another source of data 1903-1978 Mexican Daily Weather Maps- Daily pressure data in tables

Hourly data for 16 stations in the 1950-1954 period. Data will be incorporated into the (GHCN) and Integrated Surface Hourly (ISH) database Network at NCDC

http://docs.lib.noaa.gov/rescue/Foreign_climate/FCD_006_025.pdf/Mexico_carta_del_tiempo/
### Climate Database Modernization Program

#### Data Integration Projects

- **Over 150,000 images of pibal (upper air wind) records from the 1940’s to 2003 received to date from 7 African countries (Kenya, Malawi, Mozambique, Niger, Senegal, Tanzania and Zambia)**
- **Most data have been keyed and after passing quality control checks will be integrated into NOAA’s Integrated Global Radiosonde Archive Database (IGRA)**
- **Digital data files were provided to the host countries that imaged the data. Keyed data files also hyperlinked to the actual images providing an easy access to the original record**
- **Ongoing pibal tasks with Tanzania and Mozambique – may expand to surface data in 2011**
- **NWS and IEDRO (non-profit) provides some technical, administrative and logistical support**
- **Possible Namibia and ACMAD Microfiche Recovery Project**

<table>
<thead>
<tr>
<th>Country</th>
<th>Sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kenya</td>
<td>5 sites</td>
</tr>
<tr>
<td>Senegal</td>
<td>3 sites</td>
</tr>
<tr>
<td>Malawi</td>
<td>7 sites</td>
</tr>
<tr>
<td>Mozambique</td>
<td>2 sites</td>
</tr>
<tr>
<td>Tanzania</td>
<td>2 sites</td>
</tr>
<tr>
<td>Zambia</td>
<td>2 sites</td>
</tr>
<tr>
<td>Niger</td>
<td>2 sites</td>
</tr>
</tbody>
</table>
Making Data User Friendly and Accessible

The index on the CD-ROM will be the main point for searching for related data. The index page provides links to all available data types (both images of recorded data and actual keyed data) associated with the CD-ROM.

- Station: 67423
- Day/Mo/Year Time: 07/11/1970 04Z
- Keyed Image ID: CDMP06MA/MA0001/0009.jpg
- Source Image ID: .MA0001\1\P6160006.jpg

<table>
<thead>
<tr>
<th>COMPUTED</th>
<th>OBSERVED</th>
</tr>
</thead>
<tbody>
<tr>
<td>TIME</td>
<td>HGHT Azimuth Elev DIR SPD DIR SPD MINS FEET DEG DEG DEG KTS DEG KTS</td>
</tr>
<tr>
<td>&gt; 0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>500</td>
</tr>
<tr>
<td>2</td>
<td>1000</td>
</tr>
<tr>
<td>3</td>
<td>1500</td>
</tr>
<tr>
<td>4</td>
<td>2000</td>
</tr>
<tr>
<td>5</td>
<td>2500</td>
</tr>
</tbody>
</table>

Note: > Marks the Beginning of each Record in the Launch

The data above is an upper air pibal, due to the limited wind data available a keying format was developed to key the azimuth/elevation data along with the surface wind data, found in the header section. The wind data, to the middle right of the image, we found, was often determined by the average of multiple levels of wind data.
The Foreign Data Library (FDL)

The best index we have for our “vast unknown” of foreign data:

2071 boxes of data in the FDL

CDMP multi-year project to inventory data suitable for keying in GHCN daily and monthly, IGRA upper air and ISD surface

Web interface inventory will be developed after inventory is complete
The images are provided in PDF and multi-page TIFF formats.
Fiji

List of publications containing climatological data for Fiji

The publications shown below contain climatological and weather data for Fiji. The description under which images are available:

Meteorological observations taken at Suva during the year...

Annual Meteorological Report for the year...

General Note: Meteorological observations taken at the Office of the Harbour Master, Suva. Description under which images are available:

Volumes available:
- 1910-1912 [PDF format, PDF format]
- 1913-1914 [PDF format, PDF format]
- 1915-1916 [PDF format, PDF format]
- 1917-1918 [PDF format, PDF format]
- 1919-1920 [PDF format, PDF format]
- 1921-1922 [PDF format, PDF format]
- 1923-1924 [PDF format, PDF format]
- 1925-1926 [PDF format, PDF format]
- 1927-1928 [PDF format, PDF format]
- 1929-1930 [PDF format, PDF format]
- 1931-1932 [PDF format, PDF format]
- 1933-1934 [PDF format, PDF format]
- 1935-1936 [PDF format, PDF format]
- 1937-1938 [PDF format, PDF format]
- 1939-1940 [PDF format, PDF format]
- 1941-1942 [PDF format, PDF format]