

## 5.6 SAVING A NATIONAL TREASURE — CDMP “FORTS” PROJECT: ADDING HISTORIC PRE-1893 WEATHER OBSERVATIONS TO NOAA’S DIGITAL DATABASE

Raymond T. Truesdell \*  
Information Manufacturing Corporation, Asheville, North Carolina

Thomas F. Ross  
NOAA National Climatic Data Center, Asheville, North Carolina

### 1. PROJECT GOALS AND BACKGROUND

Prior to the the United States Weather Bureau’s founding in the 1890’s, volunteer civilian and military weather observers throughout the growing Nation observed and documented weather data. Now, meteorologists at NOAA’s National Climatic Data Center (NCDC), through NOAA’s Climate Database Modernization Program (CDMP), are working to convert these historic climate data to a digital file format. These data are being keyed through the CDMP “Forts” program, so-called because many of the observations were made at U.S. Army posts throughout the country. These Forts data will be incorporated into NOAA’s digital database, extending the database’s period of record back close to 100 years for some stations, and providing an invaluable resource for researchers performing climate studies.

CDMP meteorologists face unique challenges in preparing these forms for keying. A pre-keying review process has been developed uniquely for these Forts data. These steps, detailed below, enable keying contractors to efficiently convert these data from the forms to digital files with a high degree of accuracy.

### 2. FORTS PROGRAM & PARTNERS

The CDMP Forts program is a successful example of a cooperative effort between government agencies and the private sector. CDMP meteorologists at NCDC in Asheville, NC, including contractors employed by CDMP partners Information Manufacturing Corporation and STG, Inc., work closely with the Midwestern Regional Climate Center (MRCC) in Champaign, IL, and with Kentucky-based keying and imaging contractors SourceCorp. While this document will focus primarily on the CDMP role in the project, each group contributes vitally to the overall Forts effort.

MRCC has inventoried and selected approximately 160 first-priority stations, all with data collected prior to 1893’s start of U.S. Weather Bureau observations, for keying into digital data files. These stations include U.S. Army Signal Service locations (many of which were the immediate predecessors of the Weather Bureau

stations), and volunteer observers, including Army surgeons, noted scientists, and other professionals.

These original observers’ forms, which extend as far back as the 1780’s and vary widely in quality, content, and format, were originally microfilmed at NCDC in the 1950’s. For ease of keying and to preserve the images (as the original film reels were eventually destroyed), these microfilm reels were converted by SourceCorp to a JPEG image format.

Once CDMP meteorologists prepare the data for keying using MRCC-developed custom web-based tools, keying professionals at SourceCorp’s Jenkins, KY location carefully key each form, following the instructions provided by CDMP. When keying is completed for a station, its raw keyed data are then subjected to rigorous Quality Control (QC) at MRCC to meet NCDC standards, and are converted to standard digital formats. Once QC is completed on the data, they will be ingested into NCDC’s digital database and made available to the research community.

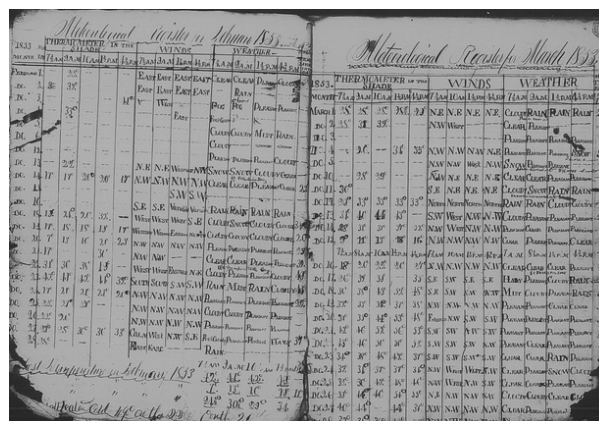


Figure 1. Hand-drawn journals are just one type of weather observation form keyed through Forts.

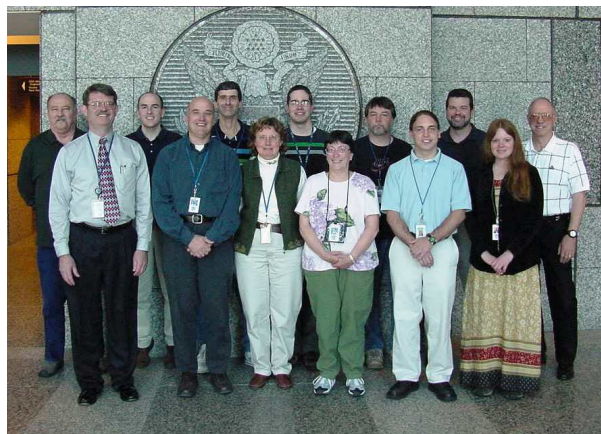
### 3. THE “PICS” PROCESS

All climate data keying projects overseen by CDMP require some level of pre-keying preparation to reduce keying errors and make the keying process easier. The varied nature of the Forts-era forms and non-standard observational practices of the day, however, make the Forts keying project unique. In addition, the forms themselves often contain precious nuggets of information about the weather, observing standards,

\* Corresponding author address: Raymond T. Truesdell, NOAA/NCDC, 151 Patton Avenue, Asheville NC 28801; e-mail: [Raymond.Truesdell@noaa.gov](mailto:Raymond.Truesdell@noaa.gov).

and instrumentation of the time. Thus, the steps taken at CDMP before actual keying of the data can begin are an integral part of the process for Forts.

To meet these unique needs, a special set of pre-keying steps was established for the Forts project. This, so-called "PICS" (Pre-keying Inventory and Comments Summary) review has been designed to accomplish parallel goals: to identify the meteorological elements for keying on each data form and supply special keying instructions when needed, and record metadata elements for inclusion into station history records.



**Figure 2. CDMP team members (above) spend hours preparing data for keying.**

### 3.1 Element Keying

Unlike most meteorological data keying projects, the Forts collection consists of numerous disparate form types. Since the project's inception in the spring of 2003, CDMP metadata keyers have encountered more than 30 different form types used by Forts-era observers, including hand-drawn personal journals (Figure 1). Thus, it is essentially impossible to create any general element format for keyers to follow.

The solution, as implemented in the Forts project, is a unique keying format created for each data form found for a particular station. Web-based tools created at MRCC are used by CDMP's meteorologists performing the pre-keying analysis (aka "metadata keyers", Figure 2) to enter the form-specific element list for keying.

The Forts data are typically recorded on monthly forms, with each day's observations written on a single row of data (Figure 3). CDMP's metadata keyers label each column of data from these daily data forms with a four-letter code identifying the column's weather element, and note whether that column's data should be keyed. On any particular form, numerous daily data elements may be selected for keying. These include values of temperature, pressure, humidity, wind direction and speed, cloud data, precipitation totals with start and stop times, state of the weather, character of

the day, and other elements; mean, at-hour, and daily values may be keyed depending on the element type.

Once the element keying for a station is completed, this information is extracted by MRCC and forwarded to SourceCorp's daily data keyers. The keyers then use the form-specific keying format much as they would use a general format in other keying projects.

### 3.2 Metadata Keying

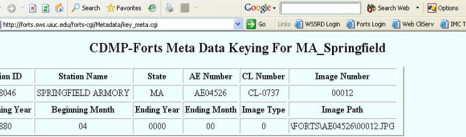
In addition to identifying data elements for keying, the Forts project affords the singular opportunity to record metadata for each station. CDMP metadata keyers are able to record this information individually for each data form (Figure 4). Metadata keyed include the collecting organization (e.g. Signal Service or Smithsonian Institution), station location, observers, instrumentation, and noteworthy observer comments.

Perhaps the most valuable piece of metadata from each form is the station location. Anything noted on the observer's form containing information about this location, including addresses, building names, elevation, and latitude/longitude, is recorded. Also keyed are any notes on station moves. This location information will be integrated with NCDC's station history and metadata inventory.

For some stations, detailed instrumentation data are recorded on the forms. These too are recorded by the metadata keyers, including instrument manufacturer, model, and type.

STATION, <u>Los Angeles Lead</u> .....; month, <u>June</u>																	
		THERMOMETER, (a)															
DATE	1896	EXPOSED								WET BULB							
		THERMOMETER TIME															
		7 a.m.	9 p.m.	11 p.m.	Mean.	7 a.m.	9 p.m.	11 p.m.	Mean.	7 a.m.	9 p.m.	11 p.m.	Mean.	7 a.m.	9 p.m.	11 p.m.	Mean.
1		49.5	55.0	61.3	61.6	49.0	55.0	61.3	61.6	49.0	55.0	61.3	61.6	49.0	55.0	61.3	61.6
2		52.2	57.0	62.5	63.0	52.5	57.5	63.0	63.5	52.5	57.5	63.0	63.5	52.5	57.5	63.0	63.5
3		60.0	67.9	62.8	63.6	60.0	67.9	62.8	63.6	60.0	67.9	62.8	63.6	60.0	67.9	62.8	63.6
4		53.0	54.0	61.2	61.1	53.0	54.0	61.2	61.1	53.0	54.0	61.2	61.1	53.0	54.0	61.2	61.1
5		53.0	52.5	61.5	61.7	53.0	52.5	61.5	61.7	53.0	52.5	61.5	61.7	53.0	52.5	61.5	61.7
6		53.5	55.0	62.2	64.6	53.5	55.0	62.2	64.6	53.5	55.0	62.2	64.6	53.5	55.0	62.2	64.6
7		53.5	57.5	61.7	61.8	53.5	57.5	61.7	61.8	53.5	57.5	61.7	61.8	53.5	57.5	61.7	61.8
8		53.5	57.5	61.7	61.8	53.5	57.5	61.7	61.8	53.5	57.5	61.7	61.8	53.5	57.5	61.7	61.8
9		53.5	57.5	61.7	61.8	53.5	57.5	61.7	61.8	53.5	57.5	61.7	61.8	53.5	57.5	61.7	61.8
10		53.5	57.5	61.7	61.8	53.5	57.5	61.7	61.8	53.5	57.5	61.7	61.8	53.5	57.5	61.7	61.8
11		53.5	57.5	61.7	61.8	53.5	57.5	61.7	61.8	53.5	57.5	61.7	61.8	53.5	57.5	61.7	61.8
12		53.5	57.5	61.7	61.8	53.5	57.5	61.7	61.8	53.5	57.5	61.7	61.8	53.5	57.5	61.7	61.8
13		53.5	57.5	61.7	61.8	53.5	57.5	61.7	61.8	53.5	57.5	61.7	61.8	53.5	57.5	61.7	61.8
14		53.5	57.5	61.7	61.8	53.5	57.5	61.7	61.8	53.5	57.5	61.7	61.8	53.5	57.5	61.7	61.8
15		53.5	57.5	61.7	61.8	53.5	57.5	61.7	61.8	53.5	57.5	61.7	61.8	53.5	57.5	61.7	61.8
16		53.5	57.5	61.7	61.8	53.5	57.5	61.7	61.8	53.5	57.5	61.7	61.8	53.5	57.5	61.7	61.8

**Figure 3. Element column headers are extracted in the PICS process for use by daily data keyers.**



CDMP-Forts Meta Data Keying for Community Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Google Search Web Options

http://fortis.usc.edu/ftrts/col/headtablekey.asp

## CDMP-Forts Meta Data Keying for MA\_Springfield

Station ID	Station Name	State	AE Number	CL Number	Image Number
190046	SPRINGFIELD ARMORY	MA	AE04526	CL-0737	00012

Beginning Year	Beginning Month	Ending Year	Ending Month	Image Type	Image Path
1880	04	0000	00	0	\\FORTS\\AE04526\\00012.JPG

Please do not hit 'enter' when typing in the boxes below. Thanks.  
 Confused about what to key, or how to key? [Click here!](#)  
 Errors on the page? [Questions?](#) [E-mail Mike R.](#) for help.

---

State - Current: Massachusetts

Station Name: SPRINGFIELD As appears on form H. Armorey Springfield Alias on form:

Station Name and State(Territory) on Form: Hampden

Beginning correction - Year:  Month:  Assume? ☐ Check if YES ☐

Ending correction - Year:  Month:

Image Type Correction:

Temporary Form Code: 19 04 2007

Image Type: check all that apply

☒ Daily\_Data\_Page ☒ Monthly\_Summary\_Page ☒ Remarks ☐ Hourly ☐ Other ☐ Tracings ☐ Cover\_Page

Collecting Organization Code: SS

Form Number:

Form Media: Printed

Title:

Done Internet

**Figure 4. CDMP meteorologists use web-based tools to enter metadata and identify weather elements for keying.**

Information about the observers themselves can reveal important insight about their observations. For many stations, the same volunteer observer was noted to be on the job for decades. In addition, observer continuity is valuable for resolving different stations within the same town. Name, title, and any other observer information are also entered with the metadata.

Any pertinent observer comments are also noted. These may include notes on instrument exposure or reliability, station moves, explanations of missing data, or comments about specific weather events, all of which can provide valuable insight into data reliability.

### 3.3 Final PLCS review

After elements and metadata are keyed for each image, another CDMR meteorologist makes the final selection of images for keying. Duplicate or illegible images are excluded, as well as any that simply do not contain keyable daily data.

In addition, the PICS reviewer passes along special instructions to the daily data keyers. These may include specific instructions on how to key troublesome entries on a particular form, or general keying instructions that apply to many or all forms from a particular station.

In the Forts era, it was common for larger towns to have multiple observers collecting data simultaneously. An important part of the final PICS review is resolving these individual stations from the collection of observational forms in a particular city. A unique temporary ID number is assigned to each individual station. Later, a permanent Cooperative Station ID will be assigned by NCDC's Archive Branch to replace the temporary ID.

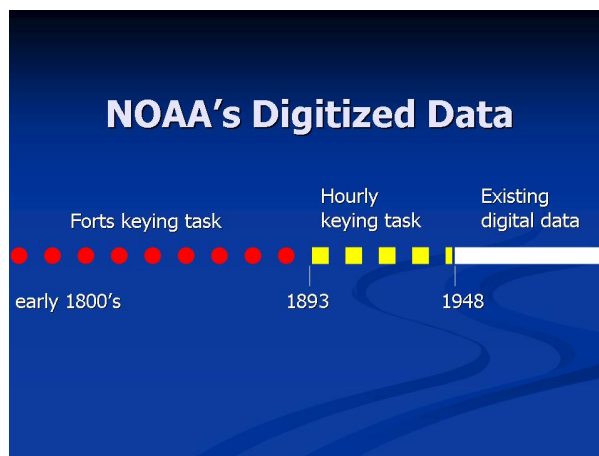
#### 4. FORTS PROJECT STATUS AND OUTLOOK

Nearly all of the 160 first-priority Forts stations' images have been prepared for keying through the ongoing PICS process. Daily data keying also continues (Figure 5), and it is anticipated that just over 100 stations' data will be keyed by the end of the current keying contract period in the spring of 2006.



**Figure 5. Representative Forts stations keyed as of late 2005.**

Based on these projections, over three million daily data records will have been keyed by spring, 2006 for the Forts era stations (Figure 7). QC and conversion to NCDC-standard formats is ongoing. When these records are added to NCDC's digital database, the database will be extended back close to 100 years for some stations. For example, daily data from New York City have been keyed all the way back to 1789! Many other locations have data that extend into the early 1800's, some for almost continuous periods of record over that time (Figure 6).



**Figure 6. For some locations, the Forts project will nearly double the period of record currently in NCDC's digital database holdings.**



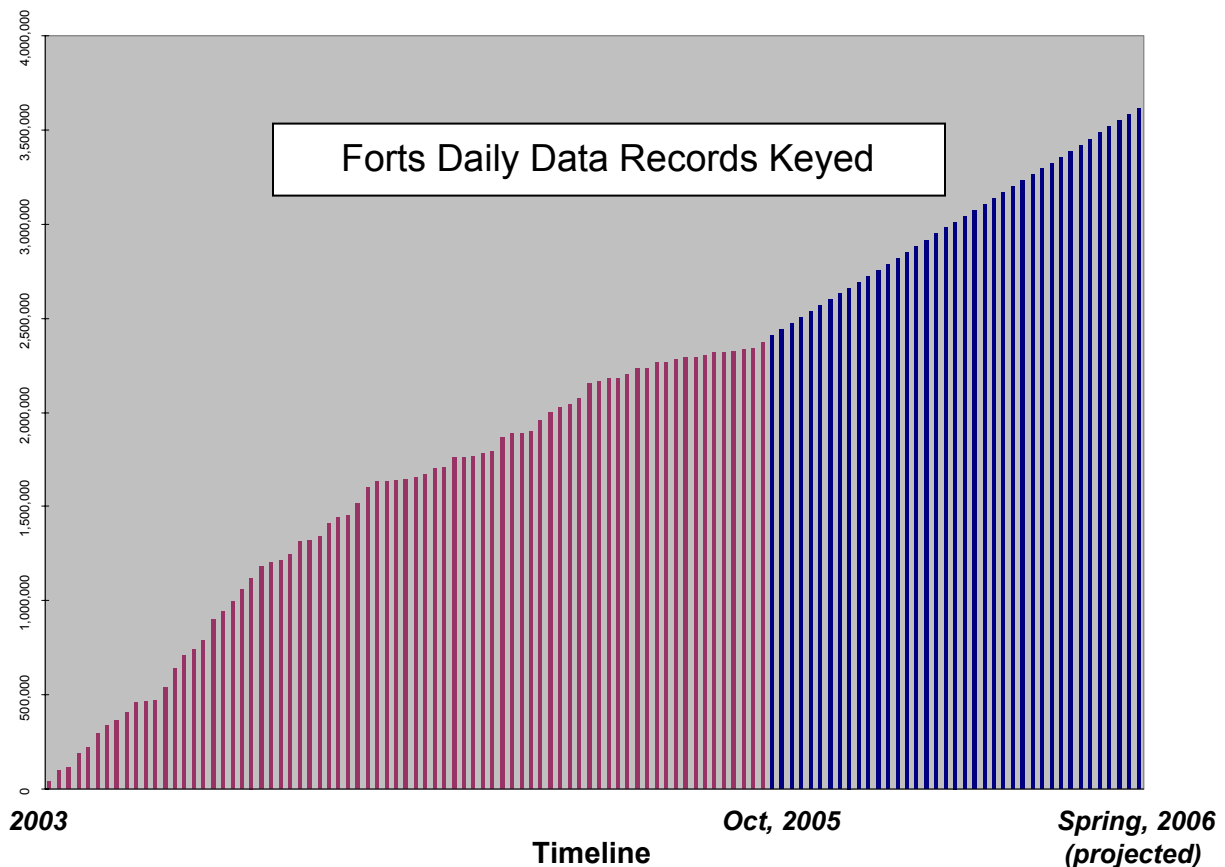
Once Forts' first phase is complete, data from some 160 locations will have been keyed. Another several hundred stations' data still remain for possible future keying, pending continuation of project funding.

## 5. CONCLUSION

The "Forts era" weather records of the 1800's are an invaluable national scientific resource. Through the CDMP Forts project, these records are being made available to the research community for the first time. Numerous scientific endeavors, including those that

may help shed light on the ongoing global climate change question, will benefit from this robust data set.

The unique nature of these data requires a similarly unique keying process. The pre-keying review of the data forms by CDMP meteorologists provides a singular opportunity to not only provide form-by-form guidance for keying contractors, but also to capture valuable metadata about the weather observers and their observations. These concurrent efforts will each add significant value to the final data product.



**Figure 7. The cumulative number of records keyed in the Forts project, showing the records added by each station keyed. By the end of the current keying contract period this spring, well over 3 million records will be keyed and ready to QC and add to NCDC's digital holdings.**