



US Climate Reference Network

Data Management Plan



**Compiled by
US Climate Reference Network (USCRN) Team at
NOAA's National Climatic Data Center
September 2012**

Compiled by:

U.S. Climate Reference Network Team at:

National Climatic Data Center (NCDC)
151 Patton Avenue
Asheville, NC 28801-5001

Howard J. Diamond, USCRN Program Manager
e-mail: howard.diamond@noaa.gov

Jay Lawrimore, USCRN Data Projects Manager
e-mail: jay.lawrimore@noaa.gov

Michael A. Palecki, USCRN Science Project Manager
e-mail: michael.palecki@noaa.gov

USCRN Web Site:

<http://www.ncdc.noaa.gov/crn>

U.S. Climate Reference Network (USCRN) Data Management Plan¹

September 2012

1. General Description of Data to be Managed

- 1.1. Name of the Dataset or data collection project [→ gmd:title].

U.S. Climate Reference Network (USCRN)

- 1.2. Keywords that could be used to characterize the data, and vocabulary from which those keywords were obtained (e.g., GCMD, CF Conventions, etc.) [→ gmd:MD_Keywords]

Air Temperature

Precipitation

Solar Radiation

Surface IR Temperature

Soil Moisture

Soil Temperature

- 1.3. Summary description of the data to be generated [→ gmd:abstract].

DSI-6117: Raw USCRN data that is transmitted continuously via the Geostationary Orbiting Environmental Satellite (GOES) datastream, and sometimes via the NOAAPort system, to the National Climatic Data Center (NCDC).

DSI-6126: Raw USCRN data collected from station dataloggers in the field during site maintenance visits. These data are put on an anonymous FTP site from where they are pulled by NCDC. These raw data are also in a different format than data transmitted through the GOES datastream.

DSI-3286: Derived geophysical parameters with other quality indicators processed from raw data by the USCRN Team. The raw data input may include PDA Files and/or Raw Data from GOES and NOAAPort. These data records are versioned based on the processing methods and algorithms used for the derivations, and data are updated when more accurate and complete raw data become available from stations' datalogger storage PDA backup files.

- 1.4. Anticipated temporal coverage of the data [→ gmd:EX_TemporalExtent].

The first USCRN stations were installed in late 2000 and the network was initially commissioned in 2004. The full conterminous U.S. network was completely commissioned in 2008. Stations are now being installed in Alaska, and as of this point, a total of 5 stations have been commissioned. Data are collected and summarized

¹ [Extracted from Appendix A of the NOAA EDMC Data Management Planning Procedural Directive (2011): https://geo-ide.noaa.gov/wiki/index.php?title=Data_Management_Planning_PD]

on timescales of 5-minutes, hourly, daily, and monthly². The network is intended to provide national climate scale data for the next 50-100 years.

- 1.5. Anticipated geographic coverage of the data [→ gmd:EX_Extent]

Conterminous U.S., Alaska, and Hawaii.

- 1.6. What data types will you be creating or capturing? (e.g., digital numeric data, photographs, video, acoustic records, database tables, spreadsheets, paper records, physical samples, etc.)

Digital numeric data

Database tables

Paper and imaged metadata records

Other metadata including photographs and maintenance reports

- 1.7. How will you capture or create the data? (e.g., satellite, airplane, unmanned aerial system, radar, weather station, moored buoy, research vessel, autonomous underwater vehicle, animal tagging, manual surveys, enforcement activities, etc.)

Data are collected from automated *in-situ* climate stations via satellite communication [GOES Data Collection System (DCS)] and direct downloads from station datalogger during annual maintenance visits. A minimal requirement level of 98% exists with respect to the receipt of data; the typical receipt rate is closer to 99.6%. Data not able to be collected due to satellite outages or other communications problems are stored on-site and are retrieved manually at a later time and then added to the database; this brings the data receipt rate up to a level typically closer to 99.8%.

- 1.8. Where will this plan be stored electronically besides in the NOAA DMP Repository?

a) **USCRN online document library at <http://www.ncdc.noaa.gov/crn/docs.html>.**

b) **NCDC's Archive document library.**

- 1.9. What volume of data is anticipated to be collected in the Project Time Frame?

This is a 50+ year project. Data collected at a rate of 26GB per year (approximately 2.2GB per month).

- 1.10. Will the data contain Personally Identifiable Information or any information whose distribution may be restricted by law or national security?

No.

2. **Points of Contact** (Give name, title, location, e-mail address, phone number and mailing address, as appropriate.) [→ gmd:CI_ResponsibleParty]

² It is planned that a total of 29 stations in Alaska will be commissioned by 2019

- 2.1. Who can, or could, represent this data collection project on NOAA's Data Management Integration Team (DMIT)? Current members of DMIT are listed at https://geo-ide.noaa.gov/wiki/index.php?title=DMIT_Membership.

Jay Lawrimore, NESDIS/NCDC (jay.lawrimore@noaa.gov; 828-271-4750)

- 2.2. Who is the overall point of contact for the data collection?

Jay Lawrimore, NESDIS/NCDC

- 2.3. Who is responsible for verifying the quality of the data?

Michael Palecki, NESDIS/NCDC (michael.palecki@noaa.gov; 828-271-4340)

- 2.4. Who is responsible for answering questions about the data collection?

Michael Palecki, NESDIS/NCDC

- 2.5. Who is responsible for data documentation and metadata activities?

Jay Lawrimore, NESDIS/NCDC

- 2.6. Who is responsible for the data storage and data disaster recovery activities?

Nancy Ritchey, NESDIS/NCDC (nancy.ritchey@noaa.gov; 828-271-4445)

- 2.7. Who is responsible for ensuring adherence to this data management plan, including ensuring that appropriate resources are available to implement the data management plan?

Howard Diamond, NESDIS/NCDC (howard.diamond@noaa.gov; 301-427-2475)

3. Data Stewardship

- 3.1. What quality control procedures will be employed?

A layered approach to quality control is employed. At the most basic level each station is monitored on a continual basis through the USCRN Station Monitoring and Reporting Tool (SMART) automated station monitoring and reporting system and through one or more meteorologists who manually review station operations on a daily basis. Data quality also is ensured through triple configuration of instrumentation. The redundant sensors provide cross-checks which make it possible to quickly identify instrumentation malfunction. A final layer of quality control is applied through the use of automated quality assurance algorithms applied at the point of ingest and insertion into NCDC datasets through checks that include basic integrity, outlier, and temporal and spatial consistency.

- 3.2. What is the overall lifecycle of the data from collection or acquisition to making it available to customer?

USCRN data are made available in near real-time via hourly satellite transmissions. For support of NWS operations, the data are collected and disseminated via the Meteorological Assimilation Data Ingest System (MADIS) at <http://madis.noaa.gov/> and the Hydrometeorological Automated Data System

(HADS) at <http://www.nws.noaa.gov/oh/hads/>. Climate quality observations are simultaneously collected at NCDC and made available via the NCDC website within 1 hour of data transmission. NCDC also transmits quality controlled hourly and daily summary data via the World Meteorological Organization's Global Telecommunication System (GTS) as an added benefit to the U.S. and international user community.

4. Data Documentation

- 4.1. Which metadata repository will be used to document this data collection?

NCDC designed a metadata repository tailored to the needs of USCRN and its stakeholders. The Integrated Station Information System (ISIS) provides archive and access for all metadata associated with the USCRN network including station id, geolocation information, agency partners, dates of installation, commissioning, and maintenance. All event details associated with scheduled and unscheduled maintenance visits and any factor that would affect the performance of a station is included. Included are details regarding instrumentation including manufacturer, model and serial numbers, and calibration coefficients for each sensor on the USCRN platform.

- 4.2. In addition to discovery-level metadata, what additional metadata or other documentation is necessary to fully describe the data and ensure its long-term usefulness? How will that metadata be collected and updated?" Is there a requirement to document this data collection in other metadata repositories?

During site survey, installation, and each subsequent annual maintenance visit, dozens of aspects of the USCRN instrumentation, site, and surrounding area are photo documented. Photographs and other documentation such as engineer site visit forms and site survey forms are stored in NCDC's Image and Publications (IPS) system. This system supports digital imaging of photography, maps, forms, etc. The photographs are reviewed by project scientists to identify changes to the station environment and other factors that may alter the true climate signal.

USCRN standard operating procedures include monitoring data for potential problems with station instrumentation and equipment. Tracking and resolution of such problems is managed in NCDC's Anomaly Tracking System, an internal web-based system that enables engineers and quality control specialists the ability to log system anomalies and track their progress toward resolution.

In addition, metadata central to the USCRN network as a whole, are maintained as part of the Archive Dataset Documentation. The storage and access of these metadata vary by the type. All are under the authority of the NCDC Data Administrator. These metadata include information about the data ingest, data processing, and data storage, commissioning information, site survey processes, and technical manuals for the suite of instruments, software and dataset documentation. Furthermore, software algorithms used to process the data are identified as Configuration Items, and are under formal NCDC Configuration Management.

When actions are taken that change the configuration of the network as a whole, such as modifying the Quality Control algorithms, an explanation is added to the Archive Dataset Documentation.

- 4.3. What standards will be used to represent data and metadata elements in this data collection? *Note:* The [EDMC Data Documentation Procedural Directive](#) calls for the use of ISO 19115 and related standards for data documentation.

The collection metadata is compliant with FGDC RSE and ISO 19115-2 standards. The use of ISO 19115-2 is in line with the NOAA Documentation Procedural Directive (<https://www.nosc.noaa.gov/EDMC/PD.all.php>). Documentation on the USCRN data stream elements is referenced within the collection metadata.

5. Data Sharing

- 5.1. Will the data be made available to the public? If so, what is the expected date of first availability? Is this a one-time data collection, or an ongoing series of measurements? Will there be a Principal Investigator hold or other delay between data collection and publication, and if so for how long? [Note: the [Data Sharing for NOAA Grants Procedural Directive](#) provides useful guidance for sharing data in a timely manner.]

USCRN data are transmitted via the GOES satellite to NCDC each hour and processed and made available to the public on the USCRN website within the hour. In addition preliminary temperature and precipitation observations are included in the satellite transmission for use by MADIS and HADS for distribution in support of forecasting and other near real-time requirements. There is no hold on the data provided at NCDC; however, the quality of the data are higher than the feed to MADIS and HADS, because data collected at NCDC are subjected to the full suite of automated processing and quality control algorithms associated with calculating official USCRN measurements from the triplicate sensor configuration. Also USCRN data are stored on the on-site datalogger and retrieved during annual maintenance visits. These data are used to confirm observations collected via satellite transmission.

- 5.2. If the data are not to be made available to the public, explain why and under what authority distribution may be restricted. [NOAA Administrative Order 212-15](#), "Management of Environmental Data and Information" (2010) states that Environmental data will be visible, accessible and independently understandable to users, except where limited by law, regulation, policy (such as those applicable to personally identifiable information or protected critical infrastructure information or proprietary trade information) or by security requirements.

All USCRN data are provided to the public as soon as they are available (usually within the hour they are collected).

- 5.3. Will users be subject to any access conditions or restrictions, such as submission of non-disclosure statements, special authorization, or acceptance of a licensing agreement?

No. Users who redistribute or use USCRN data to create products and conduct analyses are asked to cite NCDC and the USCRN program as the source of the data.

- 5.4. What data access protocols will be used to enable data sharing? The use of open-standard, interoperable, non-proprietary web services is recommended (for example, OPeNDAP, or Open Geospatial Consortium (OGC) web services).

A variety of mechanisms are used for data sharing in order to reach the widest group of users having differing preferences, needs, and level of technical sophistication. These include web services via NCDC's Climate Data Online web access system, NOAA's climate portal, the NIDIS Drought Portal, Data.gov, transmission over the GTS, tabular ASCII files provided via ftp, and web-generated html tabular data.

- 5.5. In what catalogs will these services or data be made registered to enable discovery by users and other Catalogs?
- **U.S. Federal Government's Data.gov portal**
 - **WMO Information System**
 - **Global Earth Observation Data portal**
 - **Global Observing Systems Information System [<http://gosis.org>]**

6. Initial Data Storage and Protection

- 6.1. Where and how will the data be stored initially (i.e., prior to being sent to a long-term archive facility)?

All data are stored in the USCRN database and archived at NCDC on a daily basis.

- 6.2. How will the data be protected from accidental or malicious modification or deletion? Discuss data back-up, disaster recovery/contingency planning, and off-site storage relevant to the data collection.

USCRN data are stored in the NCDC archive and in off-site storage facilities in keeping with NCDC standard operating procedures.

- 6.3. If there will be limitations to data access, how will these data be protected from unauthorized access? How will access permissions be managed? What process is to be followed in the event of unauthorized access?

Only USCRN developers and Data Access and Archive professionals have access to NCDC data storage and archive. Authorized access is granted by USCRN program management and administered by NCDC's Information Technology staff. The security of USCRN systems and data are assured by IT security personnel who monitor systems on a 24-hour basis. In the event of unauthorized access, IT security personnel will disable access, identify the source of the security breach, and develop and implement additional processes to fully secure the system. USCRN staff will evaluate the database to determine if any data have been comprised. If a data com-

promise occurred the database will be replaced by the most recent authenticated archived version.

7. Long-Term Archiving and Preservation³

7.1. In what NOAA Data Center (e.g., NODC, NCDC, or NGDC) will the data be archived and preserved? Have you begun discussions with that Data Center regarding your intended submission?

USCRN data are archived at NCDC. An internal submission agreement with the Archive system has been completed.

7.2. If you have not identified a NOAA Data Center, what is your long-term strategy for maintaining, curating, and archiving the data?

N/A

7.3. How will the costs of long-term data archiving be provided and maintained?

Long-term data management and archive costs are paid for by the USCRN program and NCDC operating funds.

7.4. What transformations or procedures will be necessary to prepare data for preservation or sharing? (e.g., quality control, format conversion, anonymization of personally-identifiable information, etc.). What related information will be submitted to the archive to enable future use and understanding of the data [e.g., metadata, references, reports, research papers, algorithms, audio or video codes, special character sets or fonts, etc.]?

All USCRN data quality controlled and converted to formats that are suitable for sharing when they arrive at NCDC. Personally-identifiable information including site host contact information and driving directions to station sites are limited to a select group of authorized NOAA personnel. No other information needs to be removed from the USCRN data or metadata. All data, metadata, FTP products, and related documentation including algorithm flowcharts and software used for binary-to-ASCII record decoding are part of the official NCDC archive.

Identify the Record Schedule applicable to these data and provide the retention time for these data.

Per records schedules in development at NCDC, raw observational inputs and station metadata will be permanently retained. Derived products will be reviewed after twenty years to assess their ongoing usefulness and value. Processing methods and algorithms for these products will be permanently retained.

³ Note: NOAA's Procedure for Scientific Records Appraisal and Archive Approval describes how to identify, appraise and decide what scientific records are to be preserved in a NOAA archive.

END OF REPORT

