

DRAFT

USCRN METADATA MANAGEMENT-

SURVEY to OPERATIONS

Marjorie McGuirk/Ed May

June. 2003

Section 1.0 Introduction

High quality data records are central to accomplishing the climate change detection and climate monitoring goals of the USCRN, and detailed metadata are crucial to establishing and interpreting a high quality data record. Each component of the observing system and its operating procedures must be fully documented. This is particularly important when changes to the station occur or are contemplated. Historically, metadata has included such variables as latitude, longitude, and elevation of the station, type of instruments, their exposure and height above ground. USCRN metadata also includes digital images, instrument specifications, calibration and maintenance records, sampling and validation procedures, and algorithms used to process and quality control the climate data.

1.1 Purpose of Document

The management of the metadata associated with the USCRN data is described in this document. Metadata is understood to mean information about data. It includes everything a researcher would need to know in order to process the USCRN data. Metadata include information on a station-by-station basis, such as location, identification and naming conventions, and detailed equipment information. Metadata also include information about the network in its entirety, such as the processes the data undergoes, the Quality Control programs, and the definitions of the data parameters, such as wind speed units and method of measurement. Just as metadata are a companion to data, this document is a companion to *USCRN Data Management – Ingest to Access*, a document that describes the management of the USCRN data itself.

Metadata are dynamic. Instruments, ground cover, station location, personnel, observation practices, processing algorithms, data formats - in short, virtually all aspects of metadata - change over time, and the temporal component is critical to interpreting the data that the metadata describes. To effectively model this changing environment, the metadata management system tracks change history for all data items over time, with the previous versions or values and their effective change dates available for retrieval.

Some metadata are under Configuration Management. That means that when the metadata parameter, such as the version number of the Data logger, changes, that change would have been preceded by a formal change request and approval from the Configuration Change Board (CCB), in compliance with the *Configuration Management Plan*. This document identifies which metadata parameters are Configuration Items under Change Management.

1.2 Metadata Categories

Metadata are roughly collected into categories as those specific to an individual station, and those for the network as a whole. **Table 1** lists Station-specific Metadata and **Table 2** list Network Metadata. These are the metadata that need to be collected, recorded and stored in a depository, and made available to the user community as part of commissioning (See *USCRN Commissioning Plan*).

1.2.1 Station Metadata

Metadata that are intrinsic to a particular station are maintained in the Station History Databases of NCDC. Currently, the repository for USCRN Station History is an Oracle-based database named CRNSITES and an imagery storage and access system called WSSRD. CRNSITES is somewhat modeled on NCDC's official Station history Database SHIPS (Lazar 1999), which is in process of migration to a new system called MI3 (for Metadata Integration and Improvement Initiative) (Arnfeld 2001). MI3 is intended to become operational for CRN metadata late 2003. NCDC's WSSRD tool, for Web Search Store Retrieve Display, permits digital imaging of photography, maps, forms, etc. USCRN metadata stored in CRNSITES and CD's will eventually be uploaded to the MI3 and WSSRD systems respectively.

The QA/QC manual procedures include monitoring data for potential problems with the stations instruments. Potential problems, such as a significant occurrence of data outage, are entered into the Anomaly Tracking Systems. When actions are taken to resolve the problem at a site, when a component is replaced for example, a record of the event is entered into the CRNSITES database.

1.2.2 Network Metadata

Metadata central to the USCRN network as a whole, not intrinsic to a particular station 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 as well as technical manuals for the suite of instruments, software and dataset documentation. Furthermore, software used to process the data are Configuration Items and are under 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.

Section 2.0 Management of Station Metadata

The CRN station metadata database uses the Oracle relational database engine, and resides at the NCDC. Oracle is a de facto industry standard for high performance relational database systems, and provides a variety of means to enforce business rules at the database level, thereby ensuring logical data integrity independent of application-level constraints and checks. The logical data structure is independent of a specific implementation platform. Eventually, the CRNSITES metadata will migrate to the NCDC MI3 Station History Database.

Read only access to USCRN Station Metadata is available on-line from the CRNSITES database at the Website <http://www.ncdc.noaa.gov/oa/climate/uscrn/>. Restricted read access limits some security-related metadata, such as directions to the station. CRNSITES utilizes a World Wide Web HTML forms-based user interface. A web-based interface simplifies visualization and selection of stations' metadata. The underlying database structure is independent of the interface, however, so that the user interface can evolve to take advantage of new technologies and techniques as they develop with little or no impact on the database.

CRNSITES has restricted write access. Updates and maintenance are performed by Network Monitoring Team (both at NCDC and ATDD) and also through an interface that permits field personnel to submit updates. A brief description of the procedures for CRNSITES is in **Appendix 1**.

Metadata stored in WSSRD database, the Station's imagery Metadata, can be made available upon request. Direct access to the WSSRD's USCRN Station Information cabinet may be granted to qualified individuals by the WSSRD system administrator. Requests for information may also be made through NCDC's Customer Service Division; these requests are filled and information is copied onto CD's and sent directly to the requestor on .

Station-specific Metadata are collected throughout the stages of selecting, installing and operating CRN station. A definitive list of Station Metadata can be read in **Table 1**, which lists the metadata in the order in which they are collected. The section below describes the cumulative growth of station metadata, which metadata parameters are added in each step, from Site Survey to Site Operations. At each step more metadata are collected, but the Metadata may not be added to the repository until later in the process. Furthermore, metadata may be modified at later steps along the way.

2.1 Site Survey

Sites are identified as candidates by Regional climate Centers (RCCs). Using a site survey checklist and supplemental survey scoring sheet, the RCCs survey the candidate sites. Information gathered in the survey, photographs, maps, host contact

information is collected. Forms and notes are scanned, combined with photos and powerpoint slides, then written onto a Site Survey Information Compact Disk. They are not entered into an official repository but are maintained at NCDRC in case the original surveys need later review for putting in a paired site, or for other reasons that may arise.

2.2 Site Approval

More metadata is collected if a site is approved. A Site Review Panel analyses the site survey metadata, collects survey scores and if the site is approved, a decision paper is written and signed and eventually a Site License Agreement is signed. This information is scanned and combined with metadata collected during the Survey, and stored on the Site Survey Information Compact Disk, for entry into WSSRD database.

2.3 Pre-Installation

Before a site is installed, NCDRC and ATDD collect and load metadata into CRNSITES, while ATDD readies for deployment. Most importantly, NCDRC ensure the GOESID is loaded into CRNSITES.

2.4 Installation

During installation, ATDD collects the necessary metadata items, as described in *The Complete Guide to Installing a USCRN Station*. Specific parameters include the serial numbers of the instruments, the calibration coefficients, driving instructions, host contact information. The installation team takes photos of the site. Some digital photos are taken before and after installation. These, along with other metadata listed, are scanned and submitted to NCDRC for inclusion in the WSSRD database. During installation, the Initial configuration of the site is recorded in CRNSITES. While a particular model of instrument has a known set of design characteristics, each individual instrument has unique characteristics. Pertinent metadata related to the instrumentation installed at each station is entered into CRNSITES.

2.5 Acceptance

After a site is installed, several checks are made to ensure that all was installed correctly. During Site Acceptance Testing, more metadata may be added to CRNSITES in preparation for Commissioning. In this time period, the Coop id number, WBAN identification number, information about other network membership, such as co-located with a Coop site, may be entered into CRNSITES. In this time period, NCDRC adds more metadata to CRNSITES, if not already entered.

2.6 Commissioning

Upon commissioning, CRNSITES database and the WSSRD databases would contain the parameters as shown in the Table 1. After completing the commissioning test and evaluation, when all is in order according to the Commissioning Plan, the commission

code in CRNSITES is changed from No to Yes and the date of commissioning is recorded.

2.7 Operations

After commissioning, during normal operations, regular maintenance will be performed on the instruments. During the life of the site, repairs will also be necessary. As instruments, exposure, and other metadata parameters change, new records will be added to the depository. Records of actual maintenance, calibration procedures performed are entered in the depository.

2.7.1 Host Maintenance

During normal operations, the host is expected to perform routine maintenance of the site monthly, such as cleaning the pyranometer, mowing the grass, and emptying the rain gauge. The host is expected to notify of maintenance actions through the Site Event Form. That form is scanned and entered into WSSRD. The Network Monitoring Team enters a record entered into the CRNSITES database describing the maintenance performed.

Often the rain gauge needs to be emptied more often than the monthly host's visit. A procedure exists wherein NCDC Network Monitors notes when the frequency of the rain gauge vibrating wire indicated that the bucket is near 50% capacity, and sends notification that the gauge needs to be emptied. This is described in the Manual Monitoring Handbook.

Another semi-routine function performed by the Site Contact is to use the Personal Digital Assistant (PDA) to download data from the Datalogger. . Following instructions supplied by ATDD the hosts opens the data logger door and downloads data. Later the data are uploaded to the Oracle Data Base. (The procedure is described in the Data Management Ingest to Access). Seeing the "data logger door open" flag, the NCDC Network Monitor would enter a new record into the CRNSITES database, as necessary if it were deemed that data quality could have been effected by the maintenance action.

2.7.2 Annual Maintenance

Presently, ATDD performs annual maintenance for calibrating instruments, swapping out components as necessary and so forth as described in *USCRN Maintenance Manual*.

An Annual Visit Form is filled out on site and sent to NCDC for inclusion in the WSSRD database. ATDD also enters a new record in CRNSITES to record modifications to metadata parameters, and the date of maintenance.

2.7.3 Non-Routine Maintenance

The Anomaly Tracking System (ATS) is used to record all performance-related potential problems with the network and stations. (See *ATS Users Manual*) ATS is a Web based

database available at site <http://nsi5.osd.noaa.gov/ats/CRN/>. The QA/QC manual procedures (described in *USCRN Data Management*), include monitoring the station's data for potential problems with the station's instruments, and for full network monitoring for generalized problems. Potential problems are entered into ATS. ATS provides the means of determining the performance measurements of specific hardware and firmware components of the Network and is a central component of Network Operations. (See The *Configuration Management Plan* and *Configuration Management Procedures*.) When actions are taken to resolve the problem at a site-specific level, a record of the event is entered into the Site Visit Form and CRNSITES database. When actions are taken at a Central Network level, such as modifying Quality Control Algorithm, an explanation is added to the Archive Dataset Documentation.

Automated quality control produces Quality Control Flags, which are recorded with the data and which are defined in the Archive Dataset Documentation. Manual Quality Control, besides resulting in Non-routine maintenance actions, also provides insight into the data. Reports can be generated from the ATS on Significant Data Quality notes, delivered to the Data Administrator and made available to the User Community. Interesting Science Source Notes resulting from Manual Quality Control are also stored and made available through the Data Administrator.

Repairs are performed by ATDD engineers or by the Site contact, at the request of ATDD. Components may be mailed to the host along with a set of instructions to replace the used part with the new part (see *USCRN Maintenance Manual*). When a host or engineer makes a repair, such as swapping out a datalogger, replacing a GEONOR wire, replacing the ball bearing on the wind gauge, ATDD enters a new record into CRNSITES describing the type of repairs done, updating all pertinent metadata parameter fields such as new calibration information, serial numbers, and a description of the action taken. A Site Visit Form is also filled out and sent to NCDC for inclusion in the WSSRD database.

Section 3.0 Management of Network MetaData

Metadata that are for the network in its entirety begin with the documentation of the instruments, including their known performances, references to research and experiments done on network's instruments, documentation on the software that is used to record, ingest, process, quality control and archive the USCRN data, and ends with the Data Set Documentation. See **Table 2** for a list of Network Metadata. The management of the USCRN data itself has been described in the document *USCRN Data Management – Ingest to Access*. USCRN data are archived under Data Set tag identification TD 3286.

Access to Instrument and Research metadata is made available through the CRN Website <http://www.ncdc.noaa.gov/oa/climate/uscrn/>. Access to Software and other Dataset documentation for USCRN is available from the Data Administrator. The TD 3286 manual is located URL: <http://www4.ncdc.noaa.gov/ol/documentlibrary/datasets.html>.)

3.1 Instrument Metadata

Each particular model of instrument has a known set of design characteristics, which are described in the Manufacturer's Manual, such as instrument specifications, instructions, reference values and diagrams for any maintenance or calibration procedures performed. Summaries from the Manufacturer's Manuals are available on the USCRN Webpage. Users may also request extensive documentation directly from the Manufacturer. Instrument Metadata made available from the manufacturer of each instrument or sensor include the sampling interval and the resolution and accuracy. See Table 2 for a complete listing.

3.2 Research Metadata

Results of Research performed on the various instruments and methods of observation and processing are made available through peer-reviewed literature, conference papers, presentations and technical notes. The USCRN Web page provides pointers to many of the known papers published, but these individual documents are not listed in Table 2.

3.3 Software Metadata

Metadata about the software that operate on the data, include ingest, processing, quality control and archiving of observations are important to the later use of the data. Often the only definitive source for the rules and transformations used is the application source code itself. Algorithms, procedures, specifications and the like may be documented via freeform text, documents, and diagrams or scanned images. These are Configuration Items and are maintained with the Archive Dataset Documentation. The Configuration Manager or the Data Administrator, as appropriate holds the source code documentation, track these documents, along with their version history, and makes them accessible to the user

3.3.1 Acquisition Software

After an instrument makes an observation, the recorded value may undergo processing prior to transmission. The CR23X micrologger manufactured by Campbell Scientific, Inc., is the data acquisition system used for the CRN station. It is a user programmable precision device that combines recording, processing and control capabilities in a single unit. The software for the datalogger is stored on CD and stored in the Archive Dataset Documentation. The CR23X is a Configuration Item. Transmission Software

The method of transmitting data can, and probably will change significantly with the lifespan of the network (See *USCRN Communications Study*). Though the transmission technique would be standard for the network as a whole, transmission instructions are specific to each individual station. Therefore, transmission metadata is recorded on a station by station basis.

3.3.2 Software for the processing of the data

Software for the ingest of the data is stored on CD and stored in the Archive Dataset Documentation under the Data Administrator

CRN_INGEST/SRC

CRN_IO

LOAD_LRGS

LOAD_SRRS

LOAD_NOAAPort

CRN_Ingest/EXE

Software for the quality control of the data is stored on CD and stored in the Archive Dataset Documentation under the Data Administrator.

CRN_QC

Software for the archiving of the data is stored on CD and stored in the Archive Dataset Documentation under the Data Administrator.

CRN2DB

CRNARCHIVE

CRN_INV

3.3.3 Other

Data inventories, produced by the processes that send the data to the archive, are valuable metadata. Data inventories for USCRN are provided on line from the USCRN Website under reports. Software for the display of the data are not included in this document. The display of the data does not in any way effect the data itself. Software for Data display software are not Configuration Items; and the documentation of that software is not metadata.

Software that produces Data Flag Summaries and Data Inventory Summaries, which are used by the Manual Quality Control process, do not effect the data itself, are not identified as Metadata, and are not Configuration Items.

3.4 Dataset Documentation

Dataset Documentation includes a manual that describes the data format. The TD 3286 manual is located URL <http://www4.ncdc.noaa.gov/ol/documentlibrary/datasets.html>.. Other documentation available from the Data Administrator), is primarily document a the variety of data products that may be derived from the basic observations, each with its specific details. Product details include:

- systems, procedures, algorithms and values used in production of the data product
- description
- period of record
- geographic coverage
- data elements
- product media and format(s)
- inventory and location of the product
- access systems.

Section 4.0 Other References to Metadata

Details regarding the administration of the USCRN are best described in documentation that is under Configuration Management.

- Site Information Handbook
- Demonstration Phase Evaluation Plan
- Site Acquisition Plan
- Configuration Management Plan
- Test and Evaluation Master Plan
- Functional Requirements
- Concept of Operations
- Program Development Plan

Acknowledgments

Much information for this document was based on earlier work by John Jensen, April 10, 2000 (especially <http://lwf.ncdc.noaa.gov/oa/climate/research/crn/crmdatametadadata.html>) and from the References listed below.

Arnfield, et al, U. S. Climate Reference Network, Part 4: Metadata. American Meteorological Society, (May 8 – 11 2000), 12th Conference on Applied Climatology, Asheville, NC,

Arnfield, Jeff, A Flexible System to Manage and Query NOAA Station History Information January 2001 IIPS AMS

Viront-Lazar, A., A Meteorological Station Information Data Base, Proceedings of the American Meteorological Society Second International Conference on Interactive Information and Processing Systems for Meteorology, Oceanography, and Hydrology, (January 14-17, 1986), Miami, FL

Viront-Lazar, A., The Definition of Station and Management of Station Metadata Information in Support of Climatological Data Bases , Proceedings of SDM-92 Planning Workshop on the Role of Metadata in Managing Large Environmental Science Datasets, November 3-5, 1992, Salt Lake City, UT

Viront-Lazar, A., Pete Seurer, Metadata for Climate Data, A Geographic Data Base Model for Station History, First IEEE Metadata Conference, (April 16-18, 1996) Silver Springs, MD

Viront-Lazar, A., K. Robbins, Advancements in the Integrated Management of Site Metadata for Multi-Agency Weather/Climate Data Networks. Third IEEE Computer Society Metadata Conference, Bethesda, MD April 6-7 1999

WMO Commission for Climatology Statement of Guidance on Metadata and Homogeneity, Draft, 2003

Appendix 1

Procedures for Utilizing CRNSITES

CRNSITES is a password protected Oracle database. Read-only Web access to the non-restricted data contained within CRNSITES is provided from the location <http://www.ncdc.noaa.gov/oa/climate/uscrm/>. Temporal alterations of any of the metadata parameters shown in CRNSITES is also recorded.

Some parameters are not available for read only access, such as the complete latitude and longitude of the site, driving directions to the site, and site host contact information, for privacy and security reasons.

Write access is password protected. Members of the Network Monitoring Team at NCDC and ATDD have permission to Edit Existing Records and to Add New Records.

Editing an existing record is reserved for fixing an entry that was made in error. This function is almost exclusively used when, following entries by the ATDD during the site installation, and before Site Acceptance, and error is noted in the initial parameters.

When entering a new record, the last record that exists for the site pre-loads all the parameter fields. Then the Monitor can change one or all of the parameters, filling in the date of modification, and the effective date of the change. An example of editing an existing record is shown in Figure 3.

New records are entered by Network Monitors upon getting notification from the site host of mowing the grass or emptying the gauge. ATDD Engineers enter a new record during annual maintenance trips. ATDD