



National Park Service

Inventory and Monitoring Division

The I&M Enterprise Climate Database Project Description

Revision Date: 17 July 2012

Author: Brent Frakes, Peter Budde

Table of Contents

1.	Background.....	3
2.	Workflow	4
3.	Costs.....	6
4.	Risks	6
5.	Benefits.....	7

1. Background

Climate is integral to understanding variations and changes within a park's natural resources. In fact, every unit of the national park system regardless of its designation is influenced by climate; it is one universal variable affecting our entire NPS portfolio. Unfortunately, to date, managing the climate information has been problematic for a number of reasons. First, the volume of data precludes effective management using text files, spreadsheets, or MS Access databases. While enterprise database management tools, including SQL-Server, can manage large volumes of data, they also require devoted personnel with a background in enterprise-level database administration. Second, because the sources of climate data vary, their formats and means of data delivery are inconsistent, placing a large burden on the end user to transform that information for any type of integrated analysis. Third, acquisition of the data from the source agencies isn't always a straightforward process since there is often a required level of understanding of what values to ignore and how to trap data errors. Finally, there has been the ongoing hope and expectation that "other" entities would identify the functional needs and develop a sustainable solution; this continual organizational uncertainty has only been a hindrance. It is both naive and unreasonable to expect that any federal entity charged with the task of providing/managing climate data will meet every need of every client.

Therefore, with this project, we hope to alleviate these issues in a cost-effective, agile, and scalable manner. As a proposed solution, IMD has developed a basic climate data repository with the purpose of meeting the most fundamental needs across all I&M networks:

- Data acquisition from national level data sources
- Data management
- Easy access to the data

The goals of this project include the following:

- Streamline and centralize the acquisition of national-level climate data (i.e., NOAA, SNOTEL, Snow Course, RAWS, and USGS Gaging Stations)
- Provide a stable and supported repository for climate data requested by NPS users to inform resource decisions (considering the full range of applicability to natural, cultural, as well as physical resources)
- Provide a flexible and open means to access, manage, and use these data in a single and consistent format
- Provide tools to analyze gridded climate data

- Mentor the data management community in the use of an enterprise database to help them better understand where the functionality can be leveraged effectively at a regional, network, or park scale.
- Prevent duplication of effort across multiple parks, networks, and regions
- Help clarify the business requirements for managing and analyzing climate data to address resource management issues within the NPS
- Be positioned to provide the data in a consistent and well-documented format if and when another more appropriate entity has capacity to fully support the NPS's business requirements for the management of climate data.

The scope of the project includes the following:

- Automated quarterly acquisition and basic quality control of national level data, including information available from GHCN, SNOTEL, Snow Course, RAWWS, and USGS Gaging Stations.
- Management of the national-level data as well as other data submitted by the respective Networks
- Management of daily data (currently, we will not support the management of multiple observations for a single day).
- Management of all types of climate information, including: precipitation, temperature, wind, river flow, snow water equivalent, etc.
- Development of simple analytical tools that would likely meet the needs of most annual reports
- Internal (i.e., all NPS) read-only accessibility to the climate data via direct, ODBC, and REST connectivity
- Technical support and training to the I&M network data managers
- Database administration and regular data backup using SQL Server 2008 enterprise edition
- Quarterly synchronization of data with the authoritative data sources, including COOP, Snotel and Snow Course.

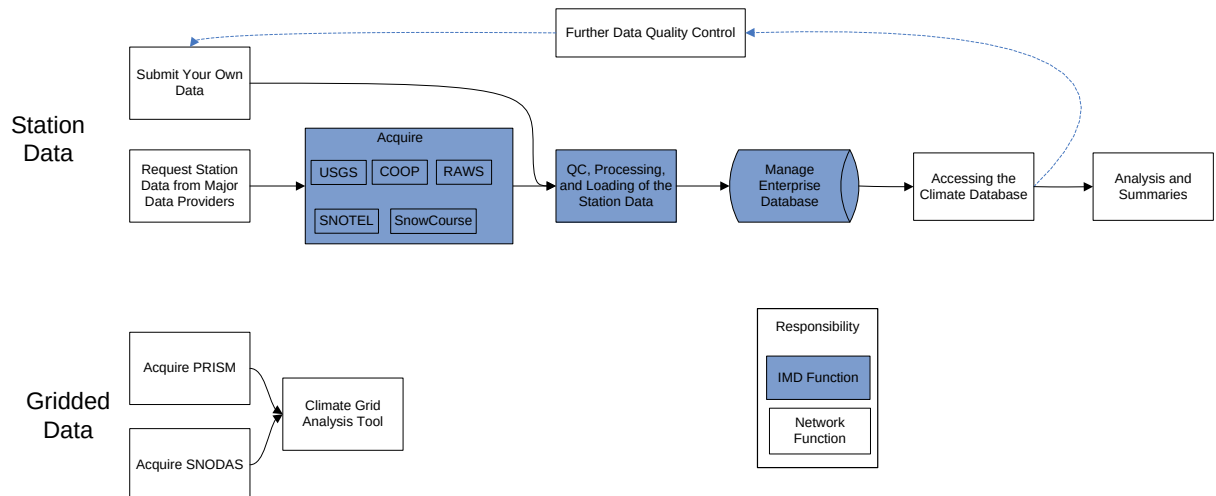
The following are out of scope of this project:

- Version control of datasets
- Management of sub-daily observations
- Development of advanced QC procedures and other analytical tools
- Oversight of when it is best to appropriately use and interpret climate data

2. Workflow

The following figure illustrates the general workflow that users are likely to follow. This workflow represents three critical and fundamental steps for supporting the NPS's climate reporting and analysis needs. Those are (1) acquisition of data from partner agencies, (2) storage and management of all data in a common format using a robust

database engine, and (3) easy and flexible access to the data for annual reporting and other in-depth analyses. This workflow is described in the top half of the following figure where the station-data workflow is described.



The second half of the workflow figure describes the general steps for working with gridded data.

All steps are documented with the Standard Operating Procedures (SOPs) and can be found on the NRInfo Portal.

Station Data

- [Processing Climate Data from the Authoritative Data Sources Using Automated Techniques](#)
- [Connecting to the I&M Enterprise Climate Database for the Purpose of Data Retrieval, Summary and/or Analysis](#)
- [Climate Database Schema, Table and Field Definitions.](#)
- [Uploading Station Data to the I&M Climate Database](#)
- [Requesting COOP, SNOTEL, Stream Gage, and Snow Course Data for Acquisition, Compilation, Basic Quality Control, and Distribution by the NPS Natural Resource Program Center](#)
- [Submitting Data for Upload to the I&M Enterprise Climate Database.](#)

Gridded Data

- [Acquiring National Operational Hydrologic Remote Sensing Center \(NOHRSC\) Snow Data Assimilation System \(SNODAS\) Data](#)
- [Acquiring PRISM Climatic Data.](#)
- [Climate Grid Analysis Toolset - Tools for Assessing Regional Climatological Trends - Standard Operating Procedure.](#)

3. Costs

Initial investments include the following:

- Interviews with network staff
- Development of the database schema, views, and functions
- Preparation of SOPs

In total, all tasks required approximately two full weeks of staff time.

Ongoing annual costs will be associated with the following:

- Quarterly loading of data – 1 week tech time
- Update of SOPs – 1 week staff time
- Training and Support – 1 week tech time
- IT costs associated with data storage

4. Risks

Below are a number of identified risks associated with the project along with how we intend to minimize the risk:

Networks Will Not Utilize the Functionality of the Climate Database – If the I&M Networks decide that this endeavor does not meet their needs, we will work closely to determine why and evaluate the cost of meeting their needs. At worst, a couple of weeks of development and effort will be lost if this project does not satisfy anticipated needs. Additionally, networks would be able to leverage the SOPs, scripts, and database schema.

Another Agency Will Develop a Solution That Fully Meets All Climate Data Needs for the I&M Networks – If and when this happens, we will have all of our data in a consistent format which can easily be provided to the new supporting agency. At worst, a couple of weeks of development and effort will be lost.

Networks Will Fully Utilize the Functionality and Require More Assistance – If this were to occur, the I&M program would likely need to expend more money on technician time required to acquire, process and load the data into the database.

IT Costs for Storage Space are Beyond What I&M Can Support Pro Bono – Networks will be assessed for the amount of disk storage to manage the data.

There are changes to the source data (e.g., COOP, Snotel) – We fully recognize that the authoritative data source is the source agency. We will synronize annually with the respective authoritiave data sources.

Proprietary Information Is Compromised – The climate database will only be accessible internally to the NPS.

5. Benefits

In summary, the project is expected to yield the following benefits:

- Overall cost savings to the I&M program since data loading and management tasks are consolidated
- Continuity of information since oversight of the enterprise database is handled by a dedicated database administrator
- Greater access to the data since all data is in a consistent format and can be accessed through a variety of channels