Technical Memorandum

Water Temperature Modeling Platform: Data Management Plan (DRAFT)

Central Valley Project Water Temperature Modeling Platform

California-Great Basin Region

U.S. Department of the Interior

May 2023

Mission Statements

The U.S. Department of the Interior protects and manages the Nation's natural resources and cultural heritage; provides scientific and other information about those resources; and honors its trust responsibilities or special commitments to American Indians, Alaska Natives, and affiliated Island Communities.

The mission of the Bureau of Reclamation is to manage, develop, and protect water and related resources in an environmentally and economically sound manner in the interest of the American public.

Water Temperature Modeling Platform: Data Management Plan (INTERIM DRAFT)

Central Valley Project Water Temperature Modeling Platform

California-Great Basin Region

prepared by

United States Department of the Interior Bureau of Reclamation

California-Great Basin

With Technical Support by:

Eyasco, Inc.

Cover Photo: Keswick Dam on the Sacramento River by John Hannon

Contents

Page

Chapter 1 Introduction **1-1**

Chapter 2 Data Types **2-1**

Chapter 3 Data Management Tasks **3-1**

Chapter 4 Hardware and Software **4-1**

Chapter 5 Summary **5-1**

Chapter 6 References **6-1**

Figures

Figure 3-1. DMS data flow showing the six steps from import to model delivery and reporting (numbered activities correspond to listed descriptions)... 3-2

Figure 3-2. DMS example of data gap graphical reporting tool to identify data gaps in a record (top) and for final data set review (bottom) for Nimbus Dam (USGS 11446500): 2000-2021. Red periods in upper graph represent data gaps and green in lower graph represents gap filled data.. 3-5

Figure 3-3. Flow chart of DMS showing Post-Processing step and the role of modeler filling/correcting data to address gaps or erroneous data and meta-data assignments.......... 3-6

Figure 4-1. DMS Software Components and the respective presentation layer, application layer and data later.. 4-1

Tables

No table of figures entries found.

Abbreviations and Acronyms

API	Application Programming Interface
CDEC	California Data Exchange Center
CIMIS	California Irrigation Management Information System
CVP	Central Valley Project
DMS	Data Management System
GB	Gigabyte
GUI	Graphical User Interface
I/O	Input/Output
MS	Microsoft
NAS	Network Attached Storage
NOAA	National Oceanic and Atmospheric Administration
OS	Operating System
PDT	Pacific Daylight Time
PST	Pacific Standard Time
QA	Quality Assurance
QC	Quality Control
RDBMS	Relational Database Management System
Reclamation	U.S. Department of the Interior, Bureau of Reclamation
TB	Terabyte
TCD	Temperature Control Device
USGS	United States Geological Survey
WTMP	Water Temperature Modeling Platform

Contents

This page intentionally left blank.

Chapter 1 Introduction

Flow and water temperature simulation models are useful and necessary tools to support resource managers in their understanding of temperature dynamics in US Bureau of Reclamation (Reclamation) CVP reservoirs and downstream river reaches. Such tools support evaluation of how operational decisions and various influencing factors can affect water temperature in reservoirs and rivers, and the resulting potential impacts to fishery species that are sensitive to water temperature. The improvement of models, modeling approach, and associated tools to support operational decision making is a necessary adaptation strategy that takes advantage of ongoing technological advancement, and additional information and data. Reclamation's objective for the development of the Water Temperature Modeling Platform (WTMP) is the effective and efficient management of resources for downstream regulatory and environmental requirements within the context of an uncertain environment. A primary development goal of the WTMP is to provide realistic predictions of reservoir and downstream river water temperatures with sufficient confidence to carry out the necessary planning for seasonal and real-time applications while also describing situational risk and uncertainty.

Data management is an important element in developing the Reclamation WTMP. The implementation of a successful Data Management System (DMS) reduces the time and labor required to assemble the high-quality datasets that are used for model input. The DMS also provides an inventory system for adding new data as it becomes available. Specifically, data management activities include identifying necessary data and data sources; establishing procedures for data acquisition; instituting measures for data quality analysis (and potentially for data quality control); creating a system for data storage; developing consistent metadata definitions and formats; and other related tasks. In addition to these important elements of data management, the DMS will be used to develop and export model-ready data to improve the efficiency of the modeling process, reduce the potential for error, and track the data/metadata sources for model simulations. Further, because reporting of model results will necessarily require the use of observed data, the DMS will also play a role in the automated production of selected modeling reports.

The DMS stores time series model data in a relational database and avoids spreadsheets for data collection and management. However, data can be queried from the DMS and processed in a spreadsheet if desired, but primary modeling data sources will be collected and managed in the relational database. This approach facilitates consistent application of data management rules, establishes hierarchical relationships that make the inventory and reporting process more efficient, and minimizes transcription error. As new data sources become available, they can be integrated quickly and efficiently either as new boundary conditions or as replacements for existing data.

This document describes data types, data management tasks, data management flow, automated data collection, and database organization.

This page intentionally left blank.

Chapter 2 Data Types

Three general types of data are used in the WTMP: time series data, physical data, and operational data. These data types are defined below, followed by a discussion of how these data are addressed in the DMS.

- *Time series data* represents a sequence of data points in chronological order. These data are usually successive measurements made by the same source, at set time intervals, over a period and are used to track changes in conditions over time. Time series data are used for two principal purposes in the WTMP: boundary conditions that represent information flow into the models, and field observations used for calibration and validation. Some time series data occur at variable periods of time, such as reservoir thermal profiles; spot measurements of temperature, flow or stage; and historical reservoir operations (see operational data, below). Typically, time series data types include reservoir stage, inflows and inflow temperature, reservoir outflow and outflow temperature time series; reservoir vertical profiles (at discrete times); river inflows (headwater and tributary) and inflow temperature time series; and meteorological data. Specific data descriptions and sources for each system are described in Reclamation (2023a)
- *Physical data* define physical aspects of the system that are not time dependent for the purposes of this temperature modeling project. Physical data include reservoir and river geometry (e.g., latitude/longitude, morphology, location of tributaries/withdrawals), reservoir intake descriptions (e.g., elevation, diameter, capacity, stage-discharge relationships), conveyance capacities, and similar information. Occasionally, physical data change when new infrastructure is added (e.g., spillway capacity), reservoir sedimentation occurs, river channel form changes notably (e.g., sediment transport and associated geomorphic processes).
- *Operational data* include reservoir operating rules, TCD management protocols, minimum instream flows, and similar information. Operations data that change at regular intervals or that exert control over other time series data can also be included in the DMS. Historic TCD operations are an example of operational data that change over the spring through fall and at uneven intervals, yet have an important relationship to flow and temperature time series data.

The DMS largely focuses on time series data because physical data do not vary or vary slowly over time. Operational data are also typically static or specified for longer periods of time and without fixed frequency. There will be exceptions and/or potential overlap to these three basic categories that will be documented accordingly.

The initial project focus will be to develop the historical data necessary to implement, calibrate and validate the models included in the WTMP. Development of the historic data includes the aforementioned steps of collection, quality assurance, data gap filling, preparing the information for model input, and documentation. The DMS will include 2000-2021 flow/hydrology, temperature, and meteorological data available to support the WTMP for the Sacramento River system (Shasta

Lake, Keswick Reservoir, the Sacramento River from Keswick Dam to Red Bluff, Whiskeytown Lake and Clear Creek below Whiskeytown Lake), the Trinity River system (Trinity lake, Lewiston Lake, and the Trinity River from Lewiston Dam to the North Fork Trinity River), the American River (Folsom Lake, Lake Natoma, and the American River below Nimbus Dam), the Stanislaus River system (New Melones Lake, Tulloch Lake, and the Stanislaus River downstream of Tulloch Lake (including Goodwin Dam).

Development of the DMS was closely coordinated with the WTMP modeling framework implementation and development. Combining the DMS and framework development activities provided an efficient mechanism for model implementation and calibration/validation (testing). Validation was the preferred nomenclature to describe the testing performed. Additional discussion on the sensitivity of this term and how validation is defined is found in *Technical Memorandum: Water Temperature Modeling Platform: Model Development, Calibration, Validation, and Sensitivity Analysis* (Reclamation 2023b). Although this data management plan addresses both Phase I and Phase II elements of the WTMP project, Phase II data management elements associated with managing forecasting information were further developed during Phase II activities.

Chapter 3 Data Management Tasks

The DMS proposed for the WTMP will automatically import and store data to as great a degree as possible, providing an efficient means of assembling, qualifying, and processing data. The tasks required to implement this type of system include a wide array of activities to ensure efficient collection, validation, storage, retrieval, and reporting of relevant data, including:

- Develop rules for organizing time series data for storage and retrieval from a relational database. This includes applying uniform naming rules and metadata tags that can be applied across all current or potential future data sources.
- Create scripts for automating collection from available sources into database tables for QA/QC processing, storage, retrieval, and reporting.
- Develop import processes and manual entry methods to supplement automated data collection for time series data that are unavailable on line, or operational and other periodic data (e.g., temperature profile data or operational data from USBR data store) used for boundary conditions, initial conditions, calibration, forecasting or reporting.
- Collect and store metadata for time series data that tracks the data source, data quality and data revisions that are used to modify source data to make it "model ready" data.
- Provide visualization tools for post-processing source data (QA/QC, gap-filling, etc.).
- Provide a means for on-demand delivery of model ready time series data to the framework.
- Document data types and date ranges for time series included in each specific model run.
- Use specific identifiers for time series to track the relationship between model input series and model output for rapid comparison and report preparation.
- Preserve record of raw data information for transparency and full process replication.

The DMS constructed as a result of implementing these tasks will relegate the common activities of collecting and organizing data to continuously running background services and make reviewing data points and trends easier and more efficient, whether for the purpose of developing a better understanding of real time physical processes or for preparing datasets for modeling. Retention of raw source data provides opportunities to reproduce the entire process for confirmation purposes. Using automated methods of data collection will also make real time data continually available for comparison with model forecast data (Phase II).

Flow of data through the DMS begins with (1) identifying available data sources and progresses through (2) importing data collection, (3) reviewing data, (4) post-processing, (5) data service, and (6) potentially importing selected model results. The data flow process from collection through delivery to the WTMP is shown in Figure 3-1 and described below.

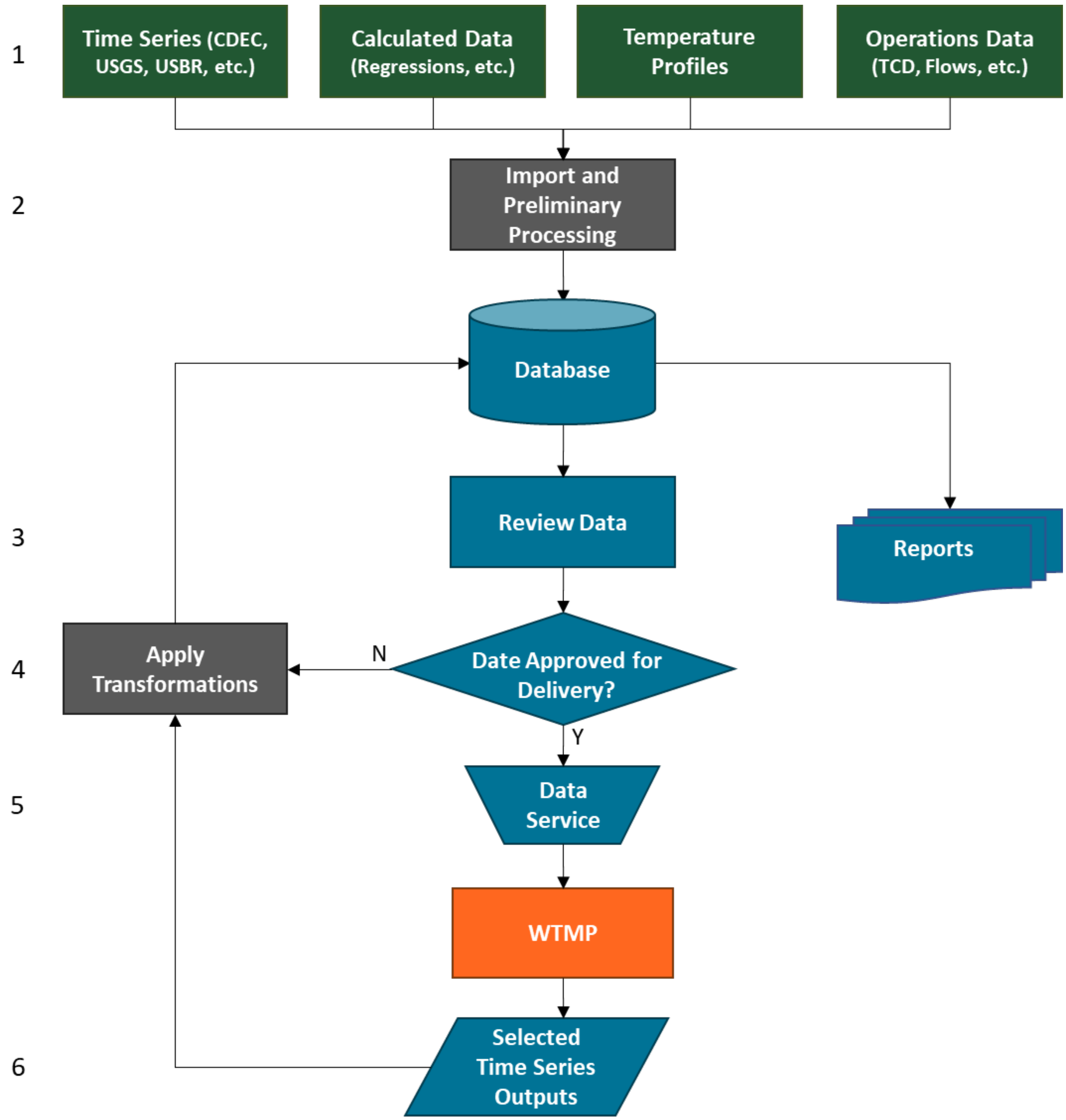

Figure 3-1. DMS data flow showing the six steps from import to model delivery and reporting (numbered activities correspond to listed descriptions).

1. Data Sources: Currently defined data sources include:

- Data collected directly from online sources (e.g., CDEC, USGS, CIMIS, NOAA).
- Data imported from manual measurements (e.g., temperature profile data).
- Data imported from Reclamation (e.g., reservoir operations, TCD schedules–).

- Calculated data, which are values derived from an algorithm to represent a measurement where no direct measurements were available (e.g., applying a rating curve to stage data to calculate flow or storage).

2. Preliminary Data Processing: Import data and associated information into the database. Apply basic data processing during the import process including QA/QC filters based on statistical analysis of historical data, scale factors and offsets, simple data gap filling, and flagging data for review. Examples include excluding flow measurements below zero or water temperatures below freezing, converting temperature units, or calculating flow from stage when rating curves are available. Basic data descriptions include station name/number, source, parameter, unit, latitude/longitude, time step, and start/end date.

3. Review Data: Data review using web-based graphical tools (time series graphs, vertical temperature profiles) and interactive reports that can support data review, interpretation, and presentation. This step requires manual review of flagged data to determine which datasets require additional attention and which data are "model ready" for delivery to the WTMP. For those data sets that are not model ready, some level of post processing is required.

Web-based graphical tools are used for visualizing and comparing data series quickly and efficiently over any date range to prepare data series or collections of series to the WTMP. Reports are useful for applying calculations over a large collection of data series and assembling the results in a paginated format for distribution and review. They can also be used to speed up the preparation of graphics and tables for inclusion in formal reports.

4. Post Processing: Those data flagged in the previous step may include data that fail the QA/QC filters, data gaps that need to be filled, or other data flagged or qualified. Apply transformations (modification) to data that is not qualified for use in models using browser-based graphical tools and other techniques. Transformations can include:

- Standardizing time (all model simulations in the WTMP will be based on Pacific Standard Time (PST) and many data sets are a combination of PST and Pacific Daylight Time (PDT).
- Identifying data gaps or erroneous data.
- Data filling. This includes filling identified data gaps and replacing erroneous data points in the raw data. This step can include different methods such as simple linear interpolation for short periods (e.g., a few hours), or for longer periods using nearby stations or algorithms (e.g., statistical relation or other calculation approach). Erroneous data can occur in the raw data as a code (e.g., -999), sensor deployment problem (e.g., water temperatures sensors exposed to the atmosphere), sensor malfunction, or other conditions that can produce a non-representative state of the system. An example of gap filling is shown in Figure 3-2.
- Normalizing time steps. Data are reported on different time steps depending on the source. Further, these data may not be on a time step consistent with the models operating within the WTMP. For example, a common model simulation time step is one hour, but available data may be on longer (daily) or shorter (15 minute) frequencies. Also, collection frequency for longer time series sometime occurs at different frequencies over the span of the record. As these data sets are prepared for delivery to the WTMP, normalizing time steps to a particular frequency may be required.

Certain data may require considerable transformation if there are significant periods of missing observations (multiple seasons or years). In some cases, there may be no data available for a model boundary condition. When significant data gaps exist in the record, the data set is sent to a modeler or technician to develop the necessary data to fill these missing records for modeling.

Qualified data sets are "approved" for model input by saving the data transformation rules and steps applied. This high level of transparency allows model analysts to track the data used in the model from its initial entry point into the DMS through any applied transformations and to the "model ready" data set.

5. Data Service: Delivery of Data to the Modeling Framework: The DMS provides a means of on-demand data delivery via web services. This is the process of transferring the final modeling input data to the WTMP in a form that the framework models can use to develop model specific input files (See Technical Memorandum: Water Temperature Modeling Platform: Model Framework Selection and Design)

6. Import Model Results: The process of potentially storing selected model results in the DMS, design decisions regarding forecasting data development from real-time sources, and reporting elements are activities that are under continuing development as the WTMP is constructed. The activity of storing model results at selected locations where field data are available will support model calibration and validation reporting. These activities will be considered in Phase II of the project.

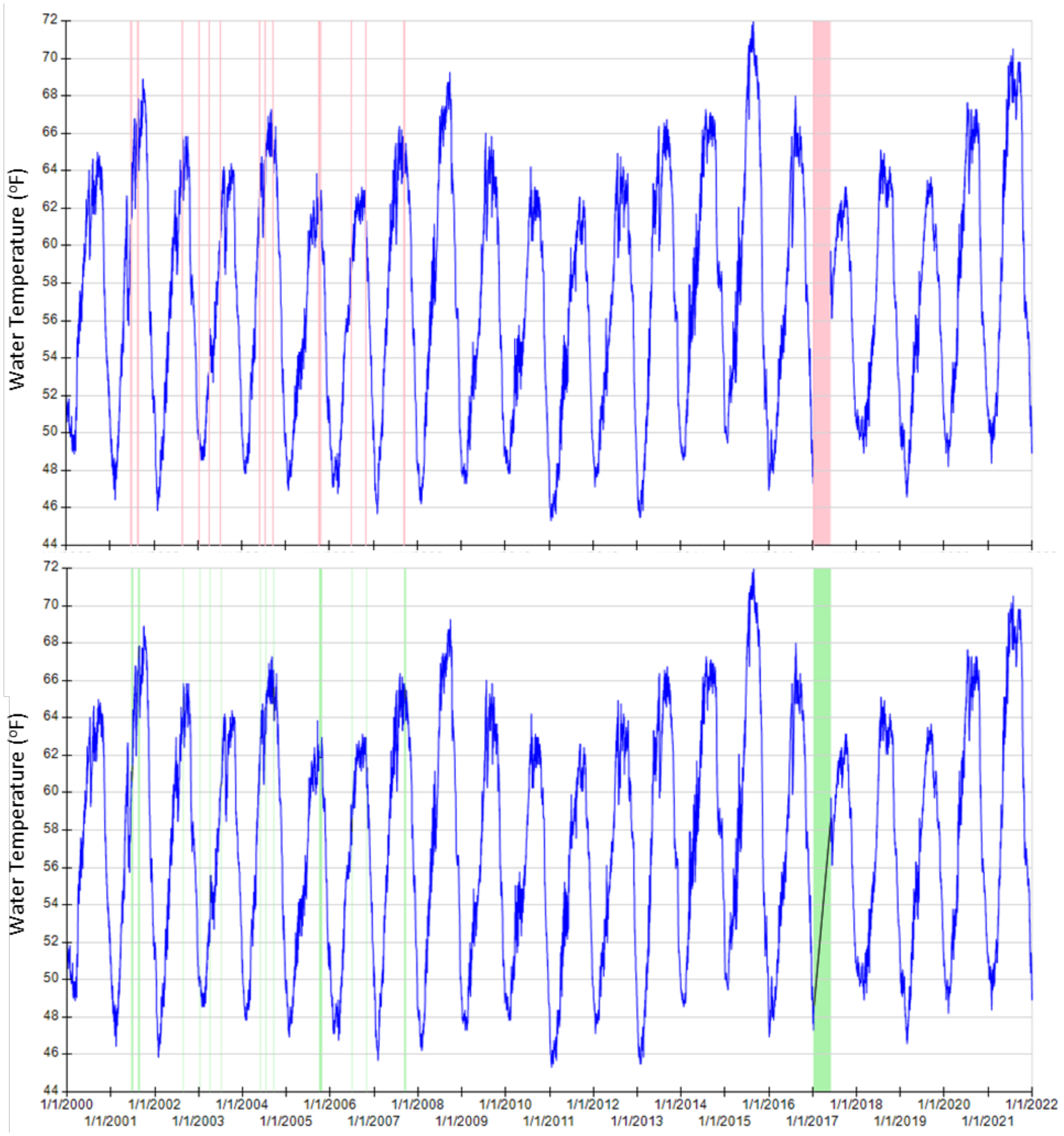

Figure 3-2. DMS example of data gap graphical reporting tool to identify data gaps in a record (top) and for final data set review (bottom) for Nimbus Dam (USGS 11446500): 2000-2021. Red periods in upper graph represent data gaps and green in lower graph represents gap filled data.

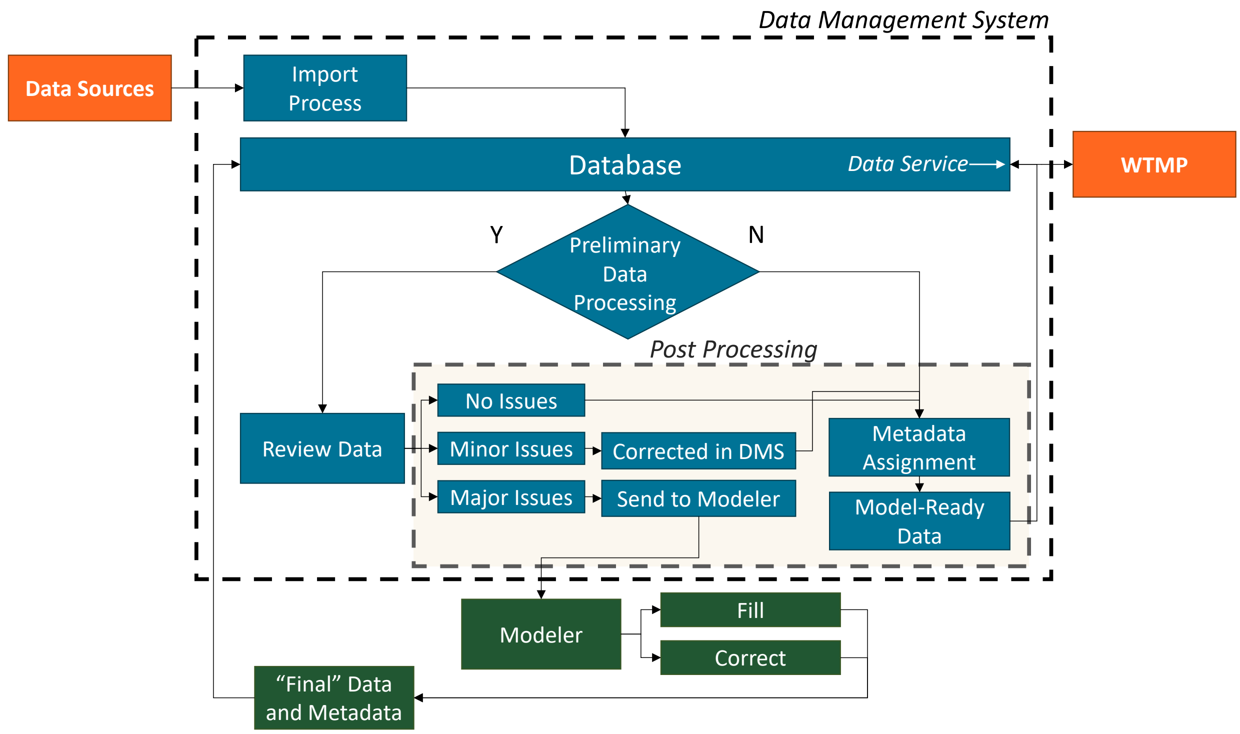

Figure 3-3. Flow chart of DMS showing Post-Processing step and the role of modeler filling/correcting data to address gaps or erroneous data and meta-data assignments.

Chapter 4 Hardware and Software

The DMS utilizes a suite of tools to collect and manage water resources and environmental data. The DMS consists of the enterprise-level software components, meaning individual components are deployed within an organization and communicate at high speed with each other over one or more networks. The DMS architecture organizes applications into three logical and physical computing tiers:

- data layer, where the data associated with the application is stored and managed.
- application layer, where data is processed
- presentation layer, or user interface.

Each of these levels, shown in Figure 4-1 with their sub-components, are described briefly below. The database is designed to have the capacity to accommodate 10 additional years of field observations (data).

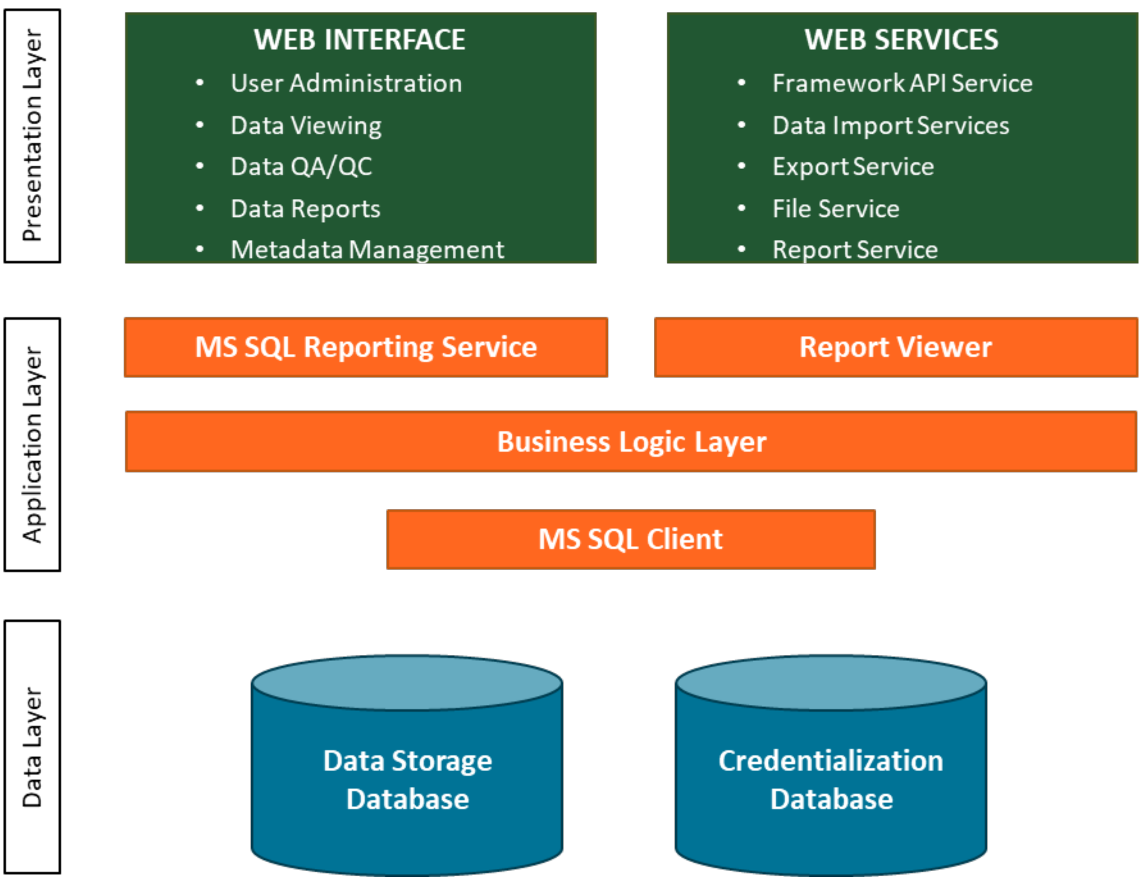

Figure 4-1. DMS Software Components and the respective presentation layer, application layer and data later.

DATA Layer includes the two Microsoft (MS) SQL Server databases used for data storage and credentialization. The data storage database includes all project data and associated metadata to support the WTMP. The credentialization database defines and manages role-based access credentials to specific datasets and applications.

APPLICATION Layer defines the computer code that acts on datasets in the database(s) in a rules-based manner to transfer data from database tables to a specific application. For example, the statements or procedures use information in the database to deliver data to a specific tool the user is credentialed to use (i.e., secure transfer).

PRESENTATION Layer is where most users, as well as automated processes, interact with the data. Most users accessing the data management system used for the WTMP will interact with one of the following presentation specific groups:

- Web Interface client is a user interface implemented in the form of a web page that can be navigated using a web browser. The Web-Interface provides a non-proprietary means of interacting with the data in a secure, rules-based manner, and can be accessed over a network. Elements of the Web Interface include user administration, data viewing, data QA/QC, data reporting, and metadata management.
- Web Services run on a server and wait for requests at a particular port over a network to serve data or documents. These services can facilitate exchange between two devices on a private network or exchange data between devices over the internet (if the web service is exposed to the internet). The DMS uses web services that run asynchronously in the background to import and process data as described above in Preliminary Data Processing. A web service is also used to deliver data to the WTMP. Elements of the Web Services include a modeling framework application programming interface (API) service, services for importing and exporting data, and a reporting service for logging transactions and emailing automated notifications about system status.

The above software components make up the data resource management system. These components are developed in the Windows .NET environment and designed to run on Windows servers. The hardware requirements vary depending on the data throughput rate (i.e., the amount of data moved successfully from one place to another in a given time period). For high throughput rates, high bandwidth/processing is required. The database, web services, web server and enterprise client may be installed on discrete computers to take advantage of the combined computational capacity. For low data throughput rates these software components can be installed on a single server with sufficient CPU, and memory and storage resources.

For future data collection, the DMS will use a combination of automated data collection for available online sources to import information for use in the WTMP. These sources include internal Reclamation database data, other agency data, manual import for pre-qualified data (such as historic model-ready data) or data hand-collected at irregular intervals, and perhaps forms-based entry for some operational data. Considering both hourly and daily data series, the total number of points envisioned to be processed is on the order of 5,000 points per day. This is a relatively low data throughput rate and as such all the software components – database, web services, web server and enterprise client – can be installed on a single server with sufficient CPU, memory, and storage resources. System backup will consider Reclamation's resources and needs, but one common option is a daily database backup schedule combined with a weekly backup to a local network-attached storage (NAS) drive and a cloud server. (A NAS is dedicated file storage that enables multiple users and heterogeneous client devices to retrieve data from centralized disk capacity.)

Minimum Windows server requirements for a single host server are listed below.

- Windows Server (latest enterprise version)
    - Core Processors: 8
    - RAM: 80 GB
    - Storage: 5 TB
    - RDBMS: MS SQL Server (latest enterprise version)
    - Web Server: IIS7

This page intentionally left blank.

Chapter 5 Summary

The primary benefits of using the DMS are to:

- Accept data sets via an automated process or manually imported into the system and manage these data behind the scenes. This includes historical data sets to support the WTMP modeling period (2000-2021) as well as future data to support modeling efforts as new information/data become available.
- Provide complete data sets (raw data to model ready data) to the WTMP to develop model input files (boundary conditions and initial conditions), as well as historic observations for model calibration and validation, and full transparency for replication. An example of model ready data is a complete boundary condition time series at the desired time step for the models in question (e.g., hourly, daily).
- Provide data related information (e.g., metadata) for modeling data sets, making available critical information regarding data source, transforms (e.g., data gap filling), and other information for the model analyst and managers to consider when applying models and using model results in decision making.
- Storing selected final model output.

The DMS will include all flow/hydrology, temperature, and meteorological data available to support the WTMP for the Sacramento River system (Shasta Lake, Keswick Reservoir, the Sacramento River from Keswick Dam to Red Bluff, Whiskeytown Lake and Clear Creek below Whiskeytown Lake), the Trinity River system (Trinity lake, Lewiston Lake, and the Trinity River from Lewiston Dam to the North Fork Trinity River), the American River (Folsom Lake, Lake Natoma, and the American River below Nimbus Dam), the Stanislaus River system (New Melones Lake, Tulloch Lake, and the Stanislaus River downstream of Tulloch Lake (including Goodwin Dam)).

The DMS includes this broad range of data sources in a consistent, comprehensive, and accessible package, including all metadata on any gap filling (missing or erroneous data). This approach provides a powerful and valuable tool for the analyst to fully explore historical data sets with graphical and tabular tools, as well as support modeling studies, to assist Reclamation staff in operating Central Valley Project facilities. The DMS will reduce the time and effort required for qualifying and assembling data sets, reduce errors in processing data manually, and efficiently transfer model ready data to the WTMP. Further, the DMS will ease the comparison of measured data with model simulations (either directly or indirectly[1]).

---

[1] Directly: if the final DMS and WTMP design includes the option for simulation results to be imported to the DMS, comparison with historical data or among model results can be made. Indirectly: if model results will be reported directly form the WTMP, the DMS can be used to populate those reporting efforts by integrating historical observations form the database.

Chapter 5 Summary

The entire system can be configured, delivered, and plugged in either within an existing business network or as a firewalled network separate from existing business networks. The process of storing selected model results in the DMS, design decisions regarding forecasting data development from real-time sources, and reporting elements are activities that are still under development and will occur in Phase II of the project.

Chapter 6 References

United States Bureau of Reclamation (Reclamation). 2023. Technical Memorandum: Water Temperature Modeling Platform: Data Development (DRAFT). In progress.

United States Bureau of Reclamation (Reclamation). 2023. Technical Memorandum: Water Temperature Modeling Platform: Model Development, Calibration, Validation, and Sensitivity Analysis (DRAFT). In progress.