RECLAMATION

Managing Water in the West

Sharing Time Series Data throughout Reclamation: water.usbr.gov

Research and Development Office
Science and Technology Program
Final Report ST-2015-969-1

U.S. Department of the Interior
Bureau of Reclamation
Research and Development Office

September 2015

Mission Statements

The U.S. Department of the Interior protects America's natural resources and heritage, honors our cultures and tribal communities, and supplies the energy to power our future.

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.

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188
T1. REPORT DATE September 2015	**T2. REPORT TYPE** Research		**T3. DATES COVERED** 2015
T4. TITLE AND SUBTITLE Sharing Time Series Data throughout Reclamation: water.usbr.gov			**5a. CONTRACT NUMBER** **5b. GRANT NUMBER** **5c. PROGRAM ELEMENT NUMBER** 969 (S&T)
6. AUTHOR(S) Karl Tarbet, ktarbet@usbr.gov James Nagode, jbnagode@usbr.gov Deena Larsen, delarsen@usbr.gov Kyle Cavalier, kcavalier@usbr.gov Steve Malers steve.malers@openwaterfoundation.org			**5d. PROJECT NUMBER** 969 **5e. TASK NUMBER** **5f. WORK UNIT NUMBER** PN 6214
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Karl Tarbet, Hydromet Manager Bureau of Reclamation, Pacific Northwest Region Steve Malers Open Water Foundation, 320 E. Vine Drive, Suite 315, Fort Collins, CO 80524			**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Research and Development Office U.S. Department of the Interior, Bureau of Reclamation, PO Box 25007, Denver CO 80225-0007			**10. SPONSOR/MONITOR'S ACRONYM(S)** R&D: Research and Development Office BOR/USBR: Bureau of Reclamation DOI: Department of the Interior **11. SPONSOR/MONITOR'S REPORT NUMBER(S)**
12. DISTRIBUTION / AVAILABILITY STATEMENT Final report can be downloaded from Reclamation's website: https://www.usbr.gov/research/			
13. SUPPLEMENTARY NOTES			
14. ABSTRACT ***(Maximum 200 words)*** A modern web service at water.usbr.gov provides time series data in computer-friendly formats that supports access by software programs including analysis tools, web browsers, and web applications. This service uses open-source software to pull data from a variety of different databases. It uses Pisces, a Reclamation-developed software, to provide an interface to graphically format and present time-series data.			
15. SUBJECT TERMS Data sharing, time-series data, databases, Pisces, TStool, Hydromet, HDB, AgriMet, weather, water, reservoir, streamflow			
16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT** U
a. REPORT U	**b. ABSTRACT** U	**c. THIS PAGE** U	**18. NUMBER OF PAGES** 42
19a. NAME OF RESPONSIBLE PERSON Karl Tarbet			
19b. TELEPHONE NUMBER 208-378-5272			

S Standard Form 298 (Rev. 8/98)
P Prescribed by ANSI Std. 239-18

PEER REVIEW DOCUMENTATION

Project and Document Information

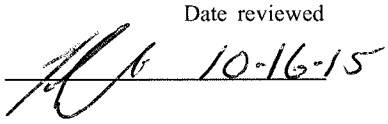
Project Name: Time Series Web Service: water.usbr.gov
WOID: Z0969
Document Final Report: water.usbr.gov
Document Authors: Karl Tarbet, Deena Larsen
Document date: September 2015
Peer Reviewer: Kyle Cavalier

Review Certification

Peer Reviewer: I have reviewed the assigned items/sections noted for the above document and believe them to be in accordance with the project requirements, standards of the profession, and Reclamation policy.

Reviewer
(Signature)

Date reviewed



Acknowledgements

Reclamation's Research program provided the ideal way to bring together Reclamation's regions to work on this project at the grass roots. Thanks to Andrew Gilmore's research Jam question in 2012; which was the critical spark for this project.

Acronyms and Abbreviations

CGI	Common Gateway Interface (method for web programs)
GP	Great Plains
HDB	Hydrologic Database used in several Reclamation locations.
LC	Lower Colorado
MP	Mid Pacific
Open Source	Software that can be freely used, changed, and shared by anyone.
Pisces	Reclamation developed time-series database and user interface
PN	Pacific Northwest
UC	Upper Colorado

Executive Summary

Problems and Needs

Managing data is central to the Bureau of Reclamation's (Reclamation) core missions of delivering water and power. Time-series data that track flows, operations, power deliveries, and more will drive Reclamation future planning and operations. With the Open Data initiative, timely delivery of machine readable data becomes an even more vital part of Reclamation's mission. Moreover, a growing number external partners and interested parties are requesting information about Reclamation's water operations, particularly measurements to support regional and local water planning and forecasting.

> The web service at water.usbr.gov provides a variety of Reclamation time series data in computer-friendly formats for software programs, including analysis tools, web browsers, and web applications.

Reclamation has several disparate water measurement systems that have publicly accessible data and yet are not easily discoverable. Providing access to Reclamation's data via web services will improve efficiency for external users, the public, and specifically university-type researchers. For example, users would be able to obtain reservoir data across Reclamation with a single method. A standard data-sharing framework is vital to share and use Reclamation's time-series datasets.

Solution

This Science and Technology Research project demonstrates that we can share time series water data Reclamation wide, even though very different databases are used within and across regions. Individual database owners keep control of their data but can publish data on a shared infrastructure. A website, water.usbr.gov, provides a central portal to list available datasets and publish data (Figure ES-1).

Results

This project created a Reclamation-wide team and a prototype to:

- Provide time series data in computer-friendly (machine-readable) formats to support software programs access, including analysis tools, web browsers, and web applications.
- Automatically pull data each day from internal databases in the Pacific Northwest (PN), Great Plains (GP), Upper Colorado (UC), and Lower Colorado (LC) regions.
- Make various sets of water operations data machine-readable and available in multiple formats.
- Connect to the Tessel web application (on the 'Monitoring Stations' layer).
- Develop a Common Gateway Interface (CGI) web data feed for Reclamation's Hydrologic Databases (HDB).

We also identified issues (e.g., security, synchronization, interfaces, data consistency, and Reclamation's Information Technology [IT] constraints) and suggested methods to resolve these issues.

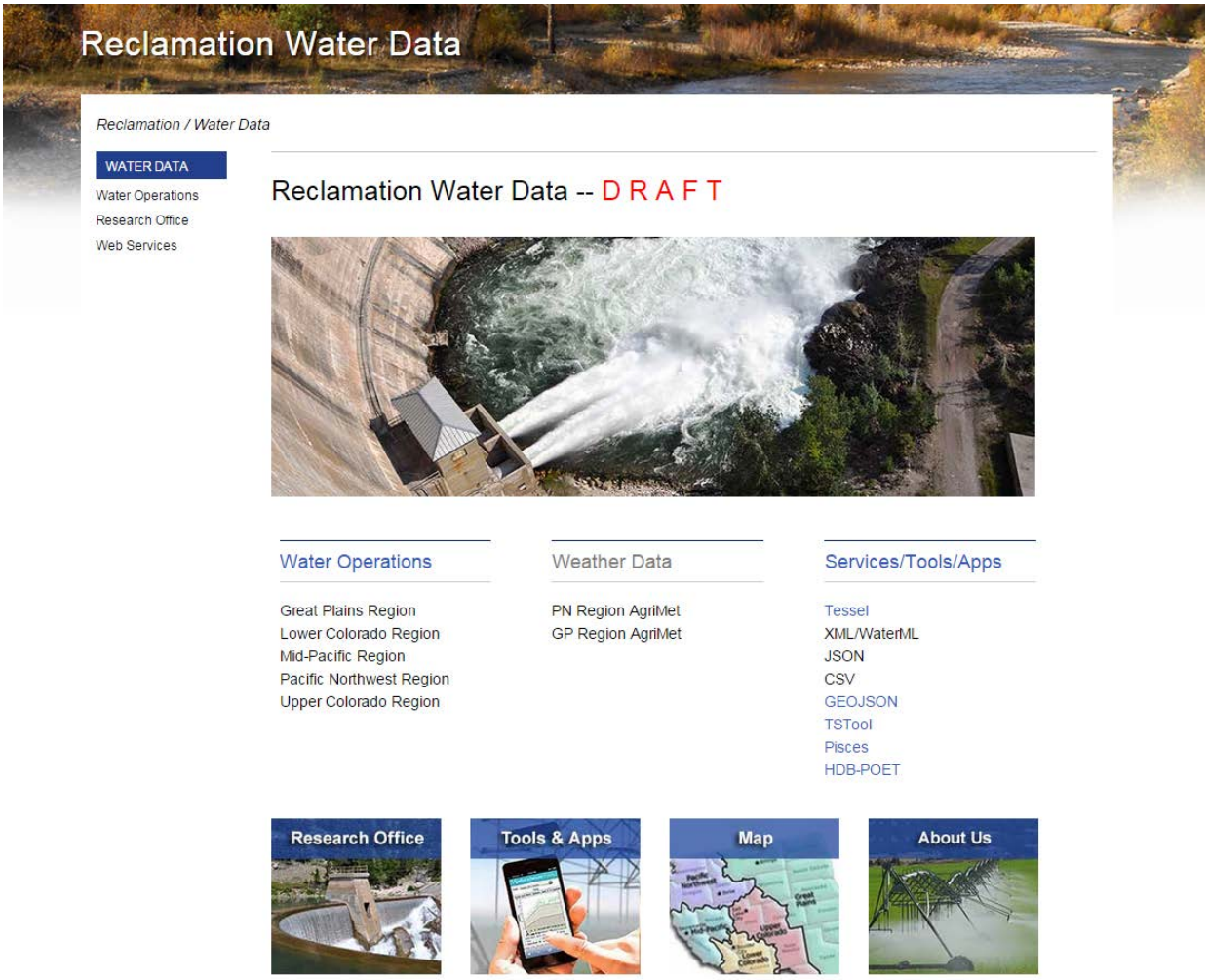


Figure ES-1. Screenshot of water.usbr.gov

This project developed a data-sharing community by bringing Denver/corporate IT expertise and assistance and the Information Resources Office (IRO) division management together to address Reclamation's dataset sharing needs. This collaboration will serve as a model for future Reclamation-wide collaboration between programmatic and IRO to support mission-related information technology development. Others, such as the Data Resource Manager, the Research Office, and water operations data producers are working together to develop a Reclamation Community of Interest for water data. This group is poised to work with the Data Resource Manager and the now-forming Open Data Team to help solve data management issues and challenges.

Recommended Next Steps

The initial project that led to water.usbr.gov helped Reclamation support the President's Open Data Initiative. Further development of this website will make this data-sharing platform continuous and sustainable. By providing an environment for datasets to become available in multiple machine-readable formats, program managers will be able to make better informed decisions, and directly share data, information and processes with other Reclamation programs. In addition, the web service could be used for other data domains (e.g., hydropower, lands, buildings and structures, invasive species, river restoration).

> Now Reclamation's program managers and their data subject matter experts have a set of tools for making data available to the public as appropriate.

The potential for this project to grow Reclamation's abilities to standardize and share our data will be realized in a short period of time. Over the long term, this could translate into better informed decision-making, reduction in data management costs, and improved program management effectiveness. To make this a Reclamation-wide application, we would need to:

- Create more custom data connections to Tessel[1] and other Geographic information systems (GIS) and spatial databases.
- Create some web apps such as a web enabled version of Pisces[2] for use by water resource professionals.
- Develop a consistent metadata schema (e.g., variable names, latitude/longitude format, elevation format). Currently some variables have different names in different databases.
- Determine advanced synchronization protocols (this may vary by database and region). For example the LC region has plans to automatically push real time changes to data to water.usbr.gov.
- Publicize the availability of the database throughout Reclamation and our partners.

[1] Tessel is a web mapping application that delivers Reclamation enterprise geospatial data as web services along with a suite of base maps and external web services that allow users to create context-rich dynamic maps of Reclamation features. See http://intra.usbr.gov/borgis/Tesselstart.html or contact the authors for access.

[2] Pisces is a desktop application that graphs and analyzes time series data. Pisces is designed to organize, graph, and analyze natural resource data that varies with time: gauge height, river flow, water temperature, etc. Pisces provides convenient access to time series model results and the ability to easily compare many different scenarios or alternatives. It has a scenario selector unique to Pisces that converts massive amounts of output data into a summary form for each time series by scenario/alternative. See http://www.usbr.gov/pn/hydromet/pisces/.

Next steps are to make this a Reclamation-wide application not only for water operation domains but other data domains such as hydropower, lands, buildings and structures, invasive species, and river restoration. The data-sharing communities of interest are working with the Data Resource Manager and the now-forming Open Data Team to help solve data management issues and challenges.

Contents

Tables

Figures

Problems and Needs

Time series datasets play a critical role in planning and operations. Without clear and simple access to historical records on flows and water and power operations, Reclamation cannot effectively plan to meet the challenges ahead. Reclamation has many dedicated employees, significant interagency agreements, and a dedicated water quality lab developing water data. However, Reclamation does not have a clean and consistent way to distribute the data. Reclamation needs to better catalog the public water data that are essential to delivering power and water.

Data stewardship varies across Reclamation with different databases and different water management needs at the field office, area office, and regional levels. Different regions have different methods to share water data via various websites and applications. For example, PN and GP Regions have Perl[3] scripts that access data from three different Hydromet servers. This technology is working well, but variations between the regions makes building Reclamation-wide tools more complicated than it should be. Moreover, Reclamation's web presence is managed by an information technology group, which is separated from water operations. Thus, there is no central way to pull data from the many disparate time-series datasets that Reclamation maintains.

Reclamation needs a web service for sharing data that:

- Provides time series data in computer-friendly (machine and human-readable readable) formats to support software programs access, including analysis tools, web browsers, and web applications.

- Uses the standard web communication channel to share information between computers.

- Provides programmable datafeeds (i.e., processes that can be used by web browsers or set up by a programmer to get data automatically).

- Uses Pisces, a Reclamation-developed software to process time-series data for Reclamation's data needs, including analyses, graphs, model input, and tracking alarms.

- Integrate with Tessel, Reclamation's overall geographic information system to provide location-specific time-series data.

Challenges

The biggest challenge in this project is that data stewardship varies across Reclamation. Some areas that vary are:

- Technologies and databases for data storage.

[3] A programming language.

- Different water management approaches at the field office, area office, and Regional levels.

- IT and IRO are separate organizations and thus priorities are isolated from programmatic efforts. Web services are managed separately from datasets and Open Data directives.

Technologies and databases vary throughout Reclamation

Most regions have different databases and associated software for managing water. Table 1 summarizes the databases explored in this research project:

Table 1. Databases in this Research Project

Regional Database	Database technology	Hardware/Operating systems
Great Plains – Hydromet	Custom Hydromet	Itanium / OpenVMS 8
Pacific Northwest – Hydromet	Custom Hydromet and postgresql	Alpha/ OpenVMS 7 / Linux
Upper Colorado HDB	Oracle	X86 / Linux
Lower Colorado HDB	Oracle	X86 / Linux
Mid Pacific HAR database	Oracle	X86 / Windows

Different water management approaches at organizational levels

Field office, area office, regional, and corporate have varying organizational structures to manage water-related databases. For example, the PN Region has dedicated staff managed by the water operations group that maintains the Hydromet hardware and software. The GP Region Hydromet software and hardware are managed by the IT group.

Web sites are managed separately from water operations

Reclamation's public web sites are managed by the corporate level information technology group in Denver. This is a different management structure than both the Regional IT and the Regional water operations. From a regional perspective, the corporate IT is a required 'vendor'. However, that vendor does not work for the Regions, and has differing priorities from water operations. This encourages regional independence as regional IT staff are managed under water operations and thus can better reflect regional operation priorities in using limited IT resources. From the corporate/Denver IT perspective, the wide variety of programs in the regions is difficult to standardize.

Solution and Results

This project created a Reclamation-wide team and a working web based service to share a variety of Reclamation's time-series datasets:

The web service at water.usbr.gov provides a variety of Reclamation time series data in computer-friendly (machine and human-readable) formats for software programs, including analysis tools, web browsers, and web applications.

Made various sets of time-series water operations data machine-readable and available in multiple formats.

- Created with Pisces software (a Reclamation-developed software) that automatically pulls data each day from various regional databases in PN, GP, UC, and LC regions and stores data in a central database (water.usbr.gov[4]) in Denver.
- Identified database owners or other points of contact in each region.
- Created a web connection to the Tessel web application (on the 'Monitoring Stations' layer).
- Developed a CGI web data feed for Reclamation's HDB databases.
- Populated the data and metadata for a representative sample of time-series data within Reclamation.
- Identified and addressed issues (e.g., security, synchronization, interfaces, data consistency, Reclamation's IT constraints).

Developed a data-sharing community. This project raises the potential for working together to solve common issues and challenges across Reclamation.

- Corporate IT Services: This project used Denver/corporate IT expertise and assistance, and it also generated interest at the IRO division management level to help mission areas with their IT needs. This collaboration will serve as a model for future Reclamation-wide collaboration between programmatic and IRO to support mission-related information technology development.
- Reclamation-wide data community: The project got the attention of Reclamation's Data Resource Manager, the Research Office, and other data managers in the water operations realm. The Data Resource Manager has since advocated a Reclamation Community of Interest for water data. This group is

[4] water.usbr.gov is the name of the database and the website. This is currently housed on test.water.usbr.gov until web teams finalize the website.

poised to work with the Data Resource Manager and the now-forming Open Data Team on helping to solve data management issues and challenges.

Addressing organizational differences.
The web service solution was influenced by the variability in Reclamation water databases and organizational structure. Two changes in the final design compared to the original project proposal are:

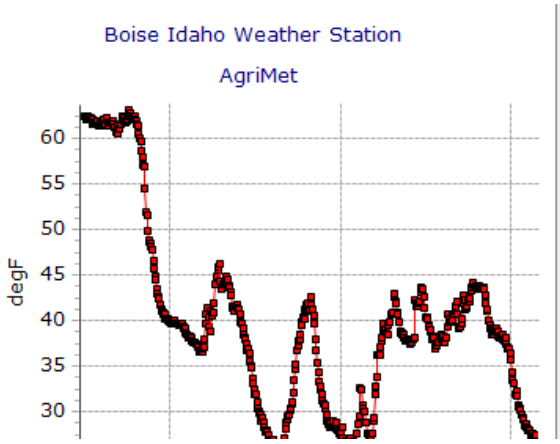
- A master database was created instead of dynamically linking to internal databases. This satisfied IT security concerns and makes the system easier to manage from the enterprise IT perspective. A central database also fits better with the corporate vision of architecture.

- The database technology was switched from PostgreSQL to MySQL to minimize the different types of technology used in Denver while keeping the open philosophy of this project.

Description of water.usbr.gov

Serving Time Series Datasets

This project is focused on daily time series data. Each time series dataset represents different measurements (or calculations) such as river flow, and air temperature varied over time. Individual datasets can be represented as a series of dates and values as shown in Figure 1.



	A	B	C
1	DATE TIME	BOII OB	Flag1
2	12/12/2014	62.58	
3	12/12/2014 0:15	62.52	
4	12/12/2014 0:30	62.09	
5	12/12/2014 0:45	62.46	
6	12/12/2014 1:00	62.34	
7	12/12/2014 1:15	62.21	
8	12/12/2014 1:30	62.09	
9	12/12/2014 1:45	62.34	
10	12/12/2014 2:00	61.61	

Figure 1. Example screenshot of a time-series dataset.

Using Pisces for the Graphic Interface

Pisces is a desktop application that graphs and analyzes time series data and this interface is used to explore the time series data. Pisces is designed to organize, graph, and analyze natural resource data that varies with time: gauge height, river flow, water temperature, etc.

Pisces provides convenient access to time series model results and the ability to easily compare many different scenarios or alternatives. It has a scenario selector unique to

Pisces that converts massive amounts of output data into a summary form for each time series by scenario/alternative. Figure 2 shows a screenshot where Lake Powell forebay data from the UC region is plotted in a graph generated from Pisces.

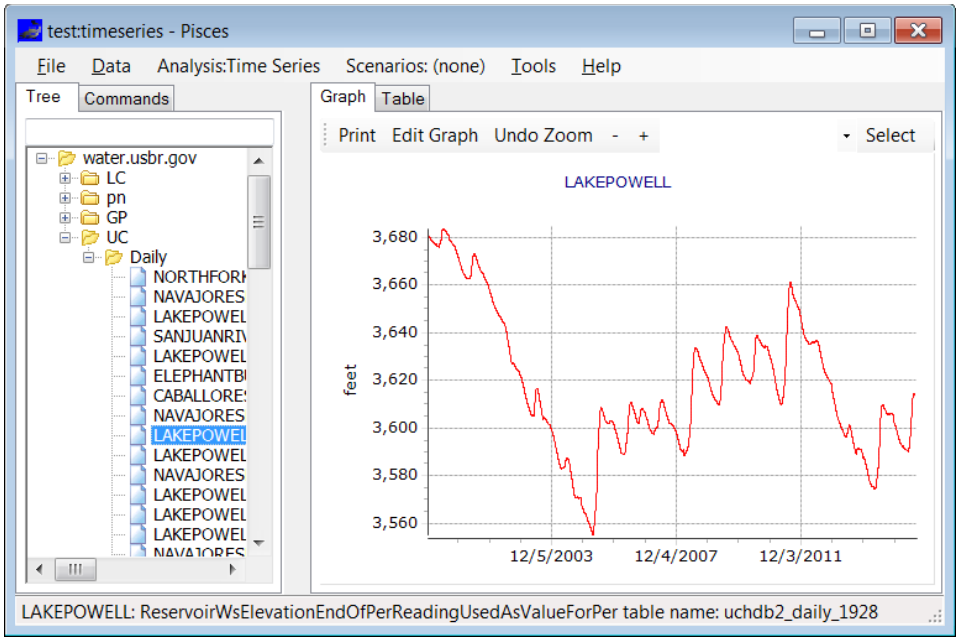


Figure 2. Screenshot of Pisces user interface.

Managing Multiple Datasets in the Time Series Web Service

Many different sets of time series data are managed at different locations throughout Reclamation, and examples are listed in Table 2. Water database owners manage the list of time-series datasets that are queried. Database owners can determine what data to share and can add, delete, and rename what is published at any time. The databases discussed in this document are not all inclusive and the design allows additional sources of data to be accommodated.

Table 2. Example Databases within Reclamation

Datasource	Data Contact
GP Region Hydromet/AgriMet database	Darren Knuteson/Tim Grove
LC Region HDB database	Kyle Cavalier
PN Region Hydromet/AgriMet database	Jama Hamel and Karl Tarbet
UC Region HDB database	Paul Davidson
MP HAR database (***not included in this research project***)	Teresa Schredl, Elizabeth Kiteck
Others, including spreadsheets??	Other Reclamation Datasets are welcome!

Combining Datasets into a Central Database

The Pisces software pulls or pushes data daily to the water.usbr.gov central database with HydrometServer.exe, the command line version of Pisces that runs on Linux or Windows computers. When Pisces is run from Denver, data are pulled from the HDB and Hydromet Servers from the regional servers daily via a CGI program. Pulling data into the central database from the regional databases is preferred for:

- **Security.** Pulling is secure because data is retrieved using read-only web queries.

- **Maintenance.** Pulling allows maintenance to be done in Denver without coordinating with each Region's database managers.

However, the database platform also allows regions to push data to synchronize databases in real time, such as when data are modified in a regional database. Regional database owners can push data to the central database in Denver whenever their data are updated or they can use automated processes. To avoid having multiple programs to read data, regional servers use a CGI program to share data from the regional servers in response to web requests. A CGI program is installed on each regional database. The CGI in LC Region has the special ability to also share data directly from the LC Region and within the region—rather than going through Denver. To minimize work and facilitate coordination (e.g., on the Colorado River), the LC database is linked with the UC database with Oracle (the pre-existing hydrologic database [HDB] database technology using pre-existing links). The PN and GP Regions both have separate CGI programs that work with their respective databases.

A CGI is a very simple web service to provide secure read-only access to data. CGI programs were selected for the following reasons:

1) CGIs require minimal technology because data travels as plain text over the network. Web queries are the standard simple way to share data.

2) CGIs allow data sharing even when Reclamation Regions use different types of databases and operating systems. Building a CGI program takes a day or two, but changing the database or operating system can take many years.

3) A CGI is not limited to use by web browsers. Many programs and spreadsheets (e.g., Excel) may be used to read data from a CGI.

4) Databases communicate with custom protocols that are vendor specific and require specific technology to setup communications.

5) Each regional database manager has control of what data is allowed to be read by customizing the CGI program.

6) A CGI program is the most basic web service that can be created with any programming language.

Figure 3 illustrates how this CGI works with the HDB database in the LC and UC regions. The components are:

- Database (lchdb)
- Web server (running on intra.lc.usbr.gov) .
- hdb_cgi.exe program
- Web browser
- Excel

The database runs on one computer and the web server runs on another computer. Lchdb is the Oracle LC Region water database. The webservers connect web requests from web browsers, spreadsheets, and custom programs to the hdb_cgi.exe program. For example, the Pisces software is a custom program that reads from hdb_cgi.exe program that the LC region database uses and saves data in the water.usbr.gov database.

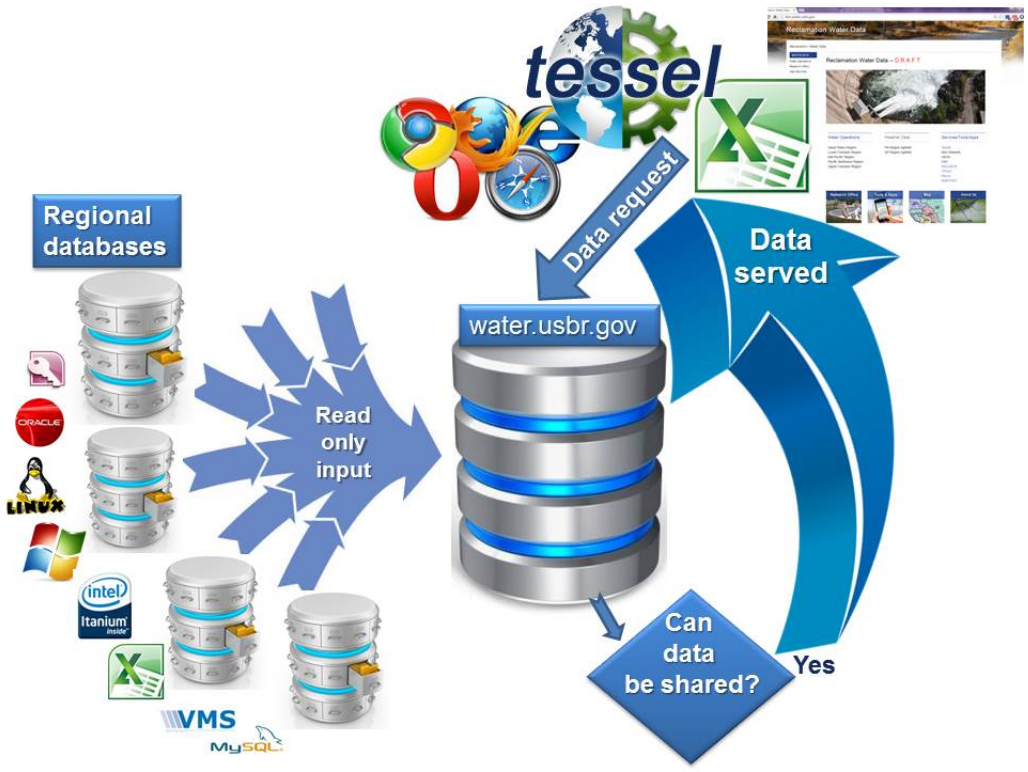


Figure 3. Data sharing process with water.usbr.gov

The hdb_cgi.exe program allows computers to talk to each other and share data in a sequence of steps:

1) Users go to a URL in their browser that need specific time-series data from a participating database (e.g.,http://intra.lc.usbr.gov/hdb_cgi.exe?interval=daily&sdi=123&t1=1-2-1981&t2=12-31-2004).

2) The user's computer calls the computer (in this case intra.lc.usbr.gov) with the request for data.

3) The computer running webserver software that handles this request, starts the program hdb_cgi.exe (intra.lc.usbr.gov in this example).

4) hdb_cgi.exe checks the config file to see if this data can be shared, then queries the regional database for data using a limited account (in this example, the HDB Oracle database).

5) The results are sent back to the user's computer and browser.

Water.usbr.gov or spreadsheets (such as Excel) use the same method to gather data.

Using Open Source Software and Standards

Water.usbr.gov is an Open Source project allowing anyone to contribute to or explore the project without needing expensive software and hardware. This project uses these existing open source software and open data standards:

- Linux
- Apache web server
- MySQL http://www.mysql.com/
- HydroJSON https://github.com/gunnarleffler/hydroJSON
- Pisces https://github.com/usbr/Pisces
- WaterML (example http://waterservices.usgs.gov/ and http://www.waterml2.org/)
- TSTool http://openwaterfoundation.org/owf-for-users/tstool

Time Series tool (TSTool) (http://openwaterfoundation.org/software-tools/tstool) is used for the under the hood data connections to the Pisces database. The web service is built upon this tool. Programming is done in the Java programming language.

Pisces, the source code and database schema for water.usbr.gov, is programmed in the C# language. See http://github.com/usbr/Pisces.

How to Share and Manage Your Datasets

Any data put in the Pisces database is available in the web service. So by managing your data in the Pisces database, you control what data and meta-data are published. Once the meta-data is in place, the associated time series data is automatically updated on a daily basis. Pisces is a Reclamation developed interface and time series database tool. The source code is freely available at http://github.com/usbr/Pisces

Connect to the Central Database with Pisces Software

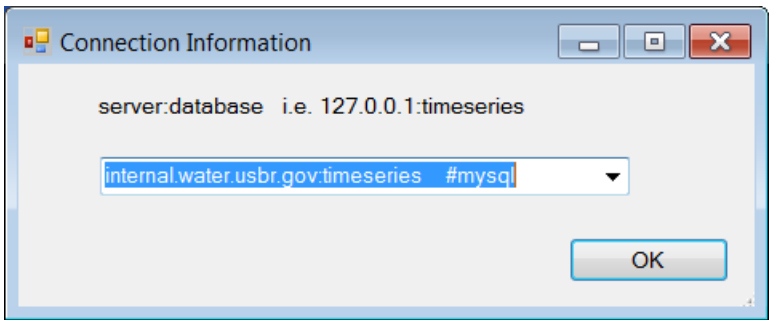
Use Pisces to import and manage datasets.

1) Download and install Pisces from http://www.usbr.gov/pn/hydromet/pisces/

2) Request a database login from ktarbet@usbr.gov. Karl will send you a database key file that works with your Windows login to provide access to the database. Copy the key file to the same directory where you installed Pisces.

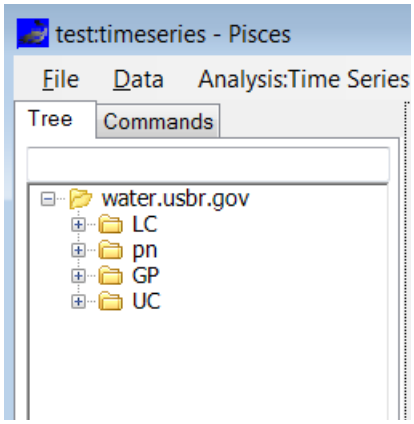
3) Start Pisces

4) Select the Menu: File->Connect to Server. Select internal.water.usbr.gov:timeseries



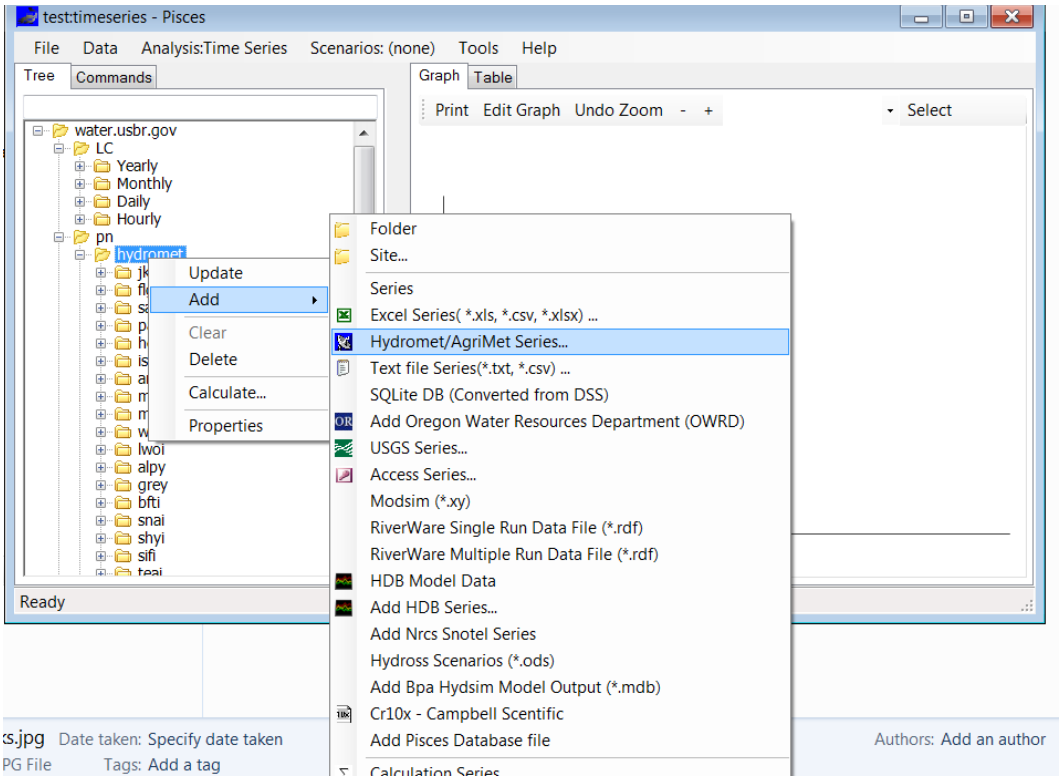
Manage Data for Your Database

Once connected with Pisces, you will see folders for each Reclamation internal water databases.



The folders organize and manage data to be published. While the folder structure is not related to how the web service works, it provides a structure for human users to find your datasets more easily. Please keep your dataset in your respective region and coordinate with other users in your region to keep an orderly folder system.

To import only one or two series from regional databases, use the Pisces interface by right-clicking on a tree folder and selecting the Add menu. To import more than two series, use the custom programs written for this project to import the data automatically. Pisces has a feature that allows expert users to edit the meta-data tables directly.



Define Your Metadata

Right-click on the series to enter the detailed meta-data in the Series Properties window. For example, Figure 4 shows the properties dialog for general information for Lake Powell in Figure 5.

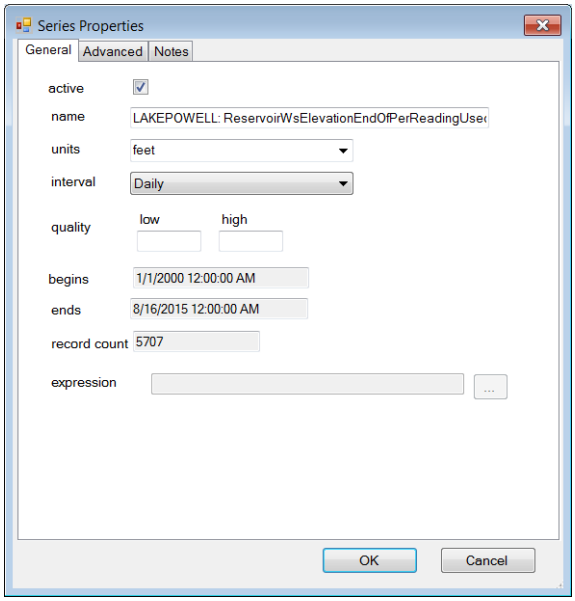


Figure 4. Properties dialog example (Lake Powell)

Each series has a set of critical metadata required by Pisces and the web service (Figure 5shows the example for Lake Powell metadata). Enter these under the advance tab in the Series Properties window. The meta data is contained in three database tables as detailed in the next section:

- Sitecatalog: Location data and contacts
- Seriescatalog: List of datasets with details about the datasets
- Seriesproperties: Ways to tag datasets (e.g., time, period of record)

Pisces has other optional tables to enter metadata, including site properties, Scenario, PiscesInfo, quality_limit, alarm, and Rating Table. Consult with Karl to determine the metadata needed for your dataset.

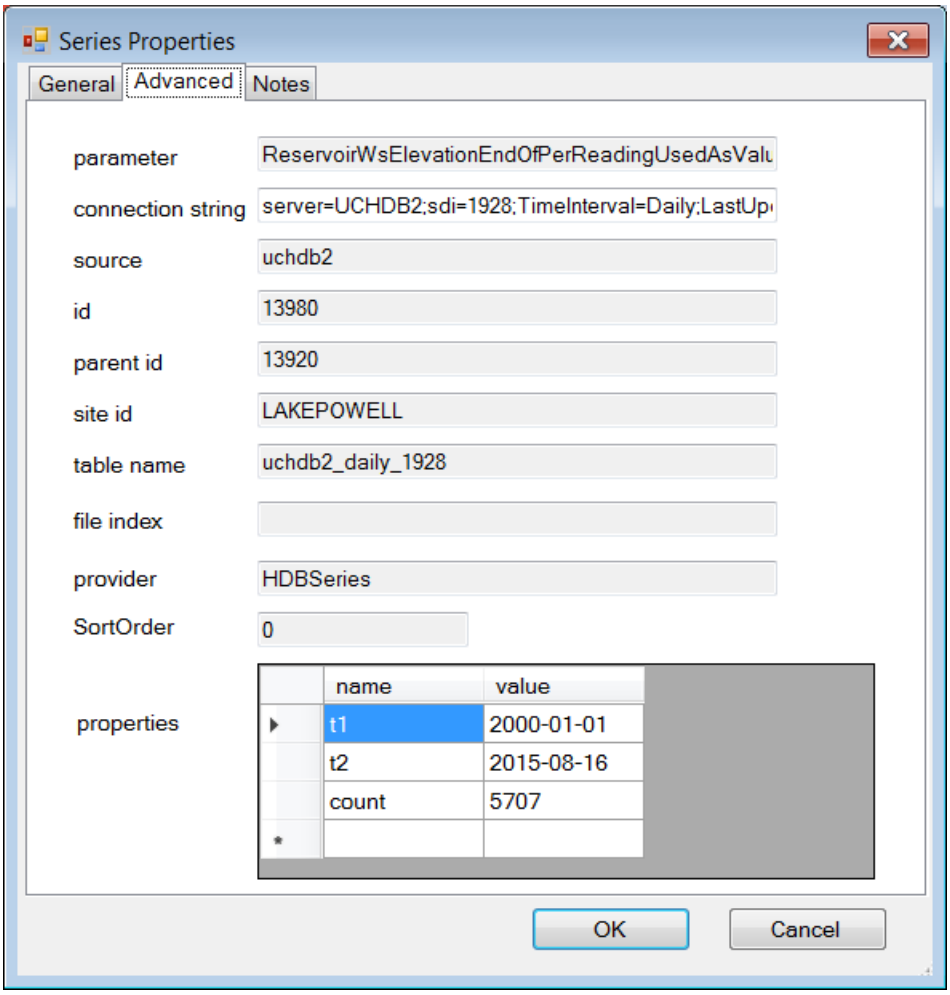


Figure 5. Example screenshot of metadata fields.

Water.usbr.gov Database Tables

Pisces metadata are contained in three database tables.

SiteCatalog Table

The sitecatalog table is used to describe a point where timeseries data is collected. This is used for GIS and other location-specific applications. Responsibility and agency_region are used to determine contacts for any questions or problems.

Table 3. SiteCatalog table

column name	description
siteid	identifier for the site example: LakeMead
description	description for this site location
latitude	latitude of site
longitude	longitude of site
elevation	elevation of the site (in units of vertical datum description)
timezone	full name example: US/Pacific
install	date site was installed
horizontal_datum	datum description for lat/long. Example: (WGS84)
vertical_datum	description of vertical datum for the site. example(NGVD29)
vertical accuracy	accuracy of elevation
elevation_method	method used to determine elevation
tz_offset	optional hours -08:00
active_flag	site is currently being used T/F default T if blank
responsibility	maintenance responsibility

column name	description
agency_region	grouping by organization regions

SeriesCatalog Table

The SeriesCatalog table (Table 4) lists all time-series datasets in the central database along with properties such as units, parameter, name, and the name of a separate table that stores the time Series data.

Each time-series dataset is stored in a separate table. A table named ‘SeriesCatalog’ is a listing of these tables and each column of the SeriesCatalog maps to properties of Series such as database table names, units of measure, and parameter name for each table (Figure 6 shows an example of this mapping).

Table 4. SeriesCatalog Table

column name	description
id	Primary key (integer)
parentid *	SiteDataTypeID of containing folder
isfolder *	When true this row represents a folder not a series
sortorder *	Sort order within a folder for user interface
iconname *	Use to render an icon based on the source of data
name	Display Name and name for equations referencing this Series/row
siteid	Reference to site/location information
units	Units of measurement such as: feet,cfs, or acre-feet
timeinterval	One of : (Instant, Daily, Monthly)
parameter	Description for data such as: daily average flow
tablename	Unique database table name for this Series/row

column name	description
provider	Name of a class derived from Reclamation.TimeSeries.Series (or Series)
connectionstring	Provider specific connection information such as a path to an excel file, sheet name, or specific parameter code
expression	Equation expression for computed series
notes	User notes
enabled	Used to active or deactive calculations and presentation of data

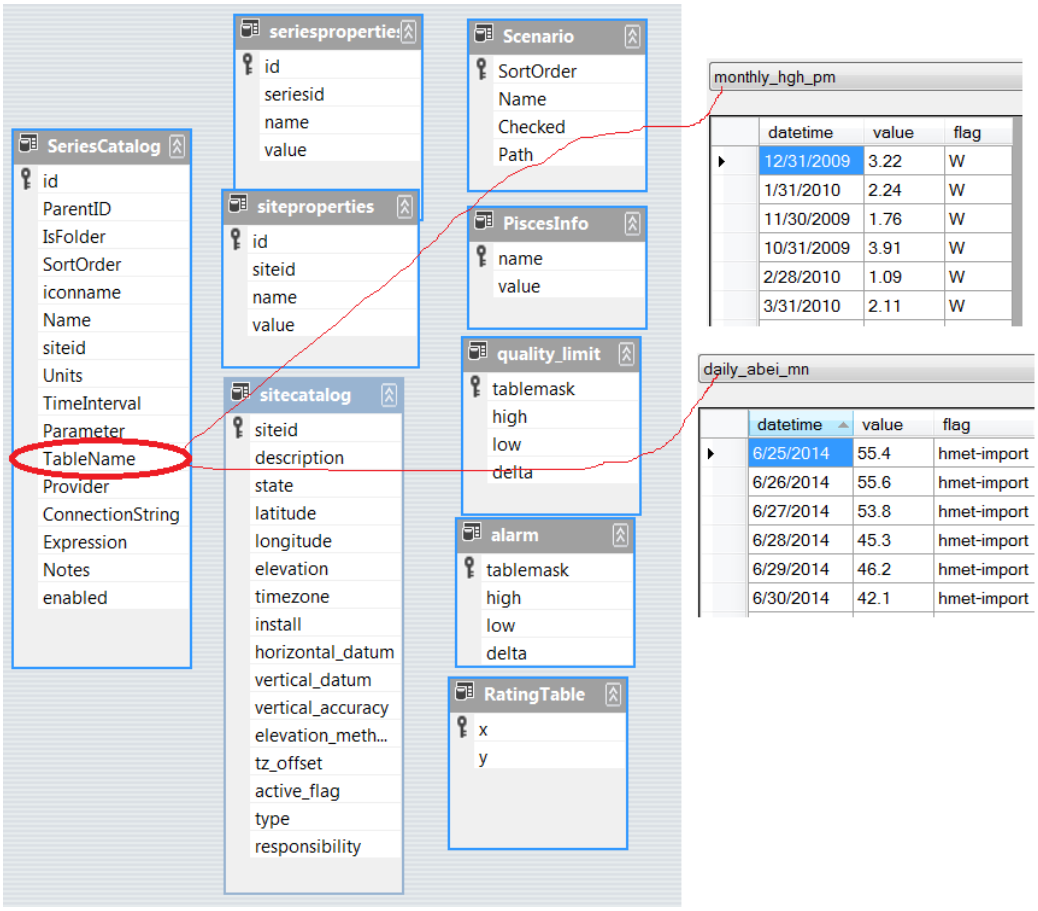


Figure 6. Example of mapping from the SeriesCatalog table to datasets (note: the third column—Scenario, etc. are optional metadata tables).

SeriesProperties Table

The seriesproperties table is used to store time-series dataset specific information. This is supplemental or additional data to the seriescatalog. For example, the period of record information is stored as properties with the names t1, t2, count where:

- t1 is the date of the first record
- t2 is the data of the last record
- and count is the total number of records in a series

The corresponding values for these would be in the value column. In this example, with daily records starting January 1, 2001 and ending December 31, 2001, the values would be:

- t1 = 1/1/2001
- t2 = 12/31/2001
- count = 365

Table 5. SeriesProperties Table

column name	description
id	Primary key (integer)
seriesid	link to seriescatalog id
name	property name
value	property value

Reliability and Security Considerations

Denver IT operations management asked that water.usbr.gov be designed for reliability, security, and with potential to go to the cloud in the future. As shown in Figure 7, the architecture allows for local regional databases to provide data to the Denver production center. This architecture adds further protection to the production water databases in the Regions because they are three steps away from the open internet. The Denver server pulls data rather than pushes data from the regions. This allows for maintenance and the ability to switch between different failover servers. Table 6 lists risks and mitigation.

Table 6. Risks and Mitigation

Risk	Mitigation
Denial of service attacks	This type of attack on the external server will have no impact on Reclamations internal water databases.
SQL injection	If the external database is hacked with SQL Injection, it will be rebuilt using the internal version. SQL Injection occurs when a malicious query tricks a computer program to do unwanted database commands.
Modification of data	Modification of external data will not affect internal data, because the data flows one way from Regions to the external server.
Network intrusion	Network intrusion is more difficult because both the web server and database are isolated from Reclamations internal network.
Disclosure of private data	Two layers work to prevent disclosure of data. 1) Database in Denver only has a catalog of public data, and copy of public data. 2) Regional water database CGI design prevents sharing of private data.

Data are housed on an internal production site and must cross a “demilitarized zone” (DMZ) to be served to the public on (www.usbr.gov) (Figure 8). The external database will not be accessible directly by Reclamation water database owners. As data are automatically replicated from internal to external production, the external database is more secure because access is extremely limited. If the external database is compromised (i.e., hacked) it is simply deleted and re copied from the internal regional databases. Moreover, scalability can be accommodated with multiple databases in the Denver production center and a failsafe backup database in Sacramento.

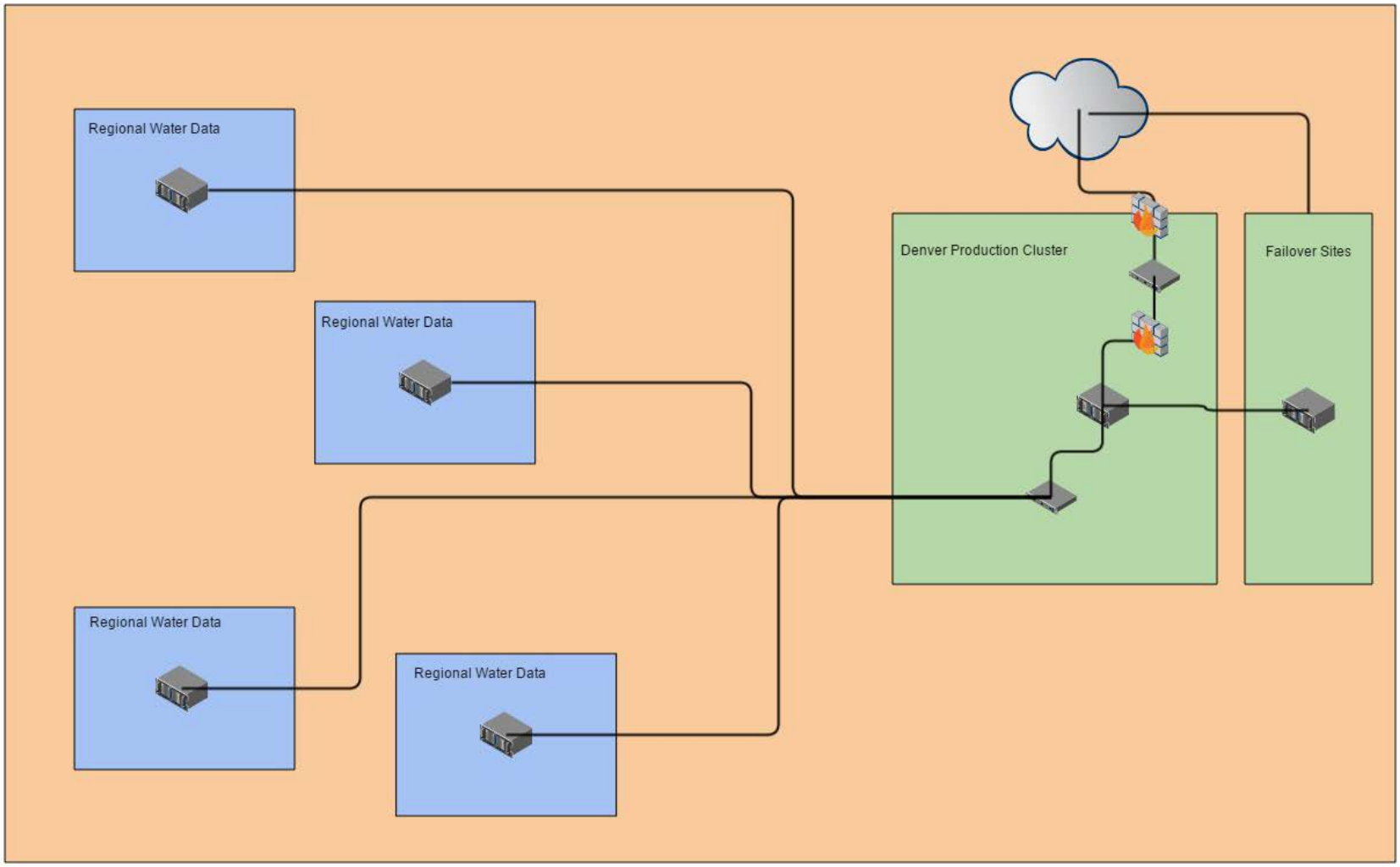


Figure 7. Network design

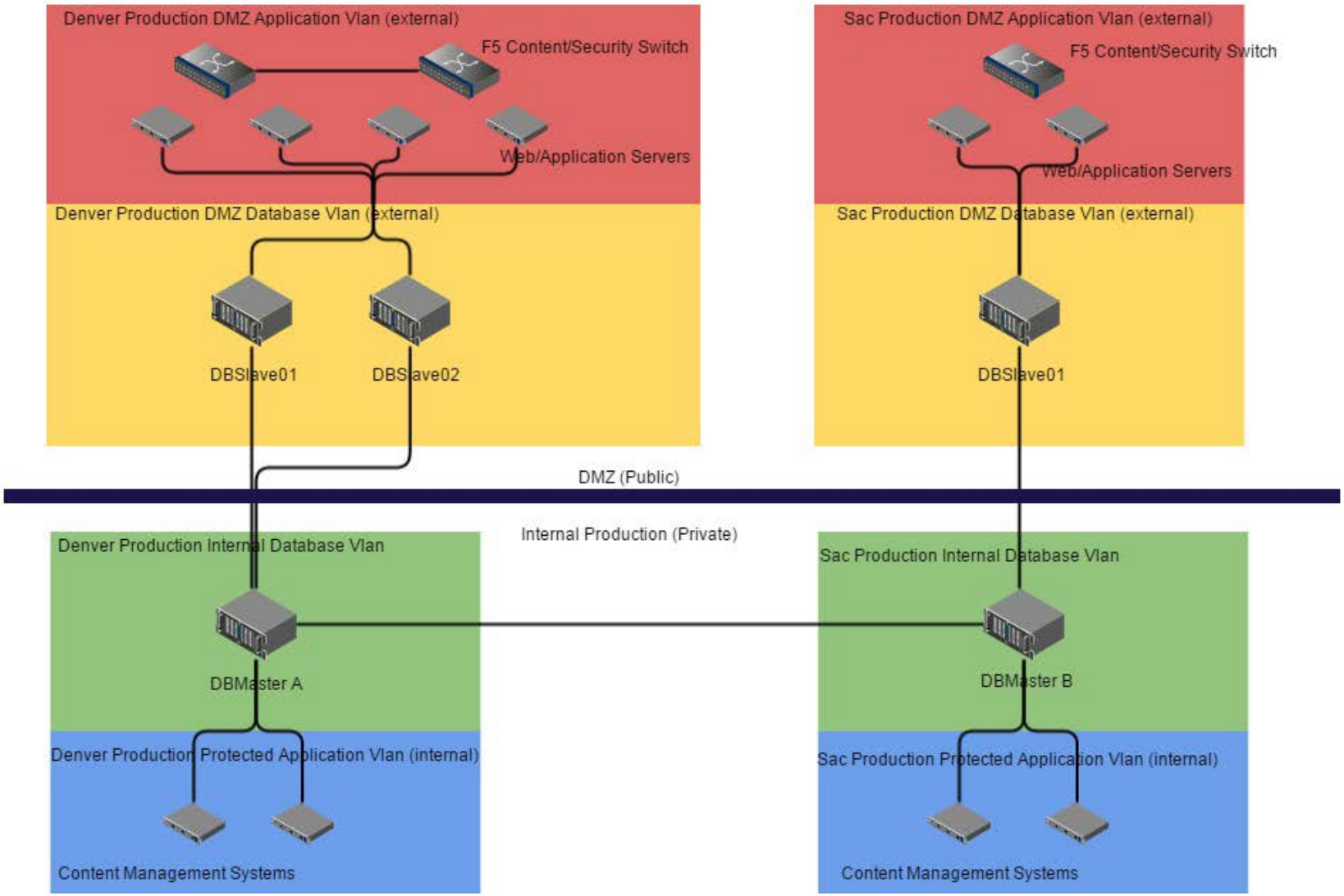


Figure 8. Failover design

References

Fielding, R.T., 2000. Architectural Styles and the Design of Network-based Software Architectures. PhD Dissertation in Information and Computer Science at the University of California, Irvine. http://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm

Reclamation, 2012. Pisces User Manual. http://www.usbr.gov/pn/hydromet/pisces/Pisces_User_Manual.pdf

Associated Websites

- Data base interface: water.usbr.gov
- Pisces Website: http://www.usbr.gov/pn/hydromet/pisces/
- GitHub usbr/Pisces: https://github.com/usbr/Pisces

Metadata for Supporting Materials

- Share Drive folder name and path where the programs are stored **Github.com/usbr/Pisces**
- Point of Contact name, email and phone: **Karl Tarbet**
- Keywords: **datasharing, metadata, databases, open software, Pisces, time-series data, water resources data**
- Approximate total size of all files: **50 MB**

Appendix 1. Web Service Project

Representational State Transfer (REST) REST is a hybrid style derived from several of the network-based architectural styles combined with additional constraints that define a uniform connector interface. As Fielding (2000) explains “REST provides a set of architectural constraints that, when applied as a whole, emphasizes scalability of component interactions, generality of interfaces, independent deployment of components, and intermediary components to reduce interaction latency, enforce security, and encapsulate legacy systems.”

Design Ideas: The following design is a supplement intended to express the requirements for RESTful URLs.

☐ Main Landing Page
/beta/daily (provides information about querying the web services)

☐ Inventory Landing Page
/beta/daily/inventory (provides information about querying daily time series)

☐ Inventory Queries
/beta/daily/inventory/uc -- list of UC series data
/beta/daily/inventory?basin=uc -- list of UC series data (alternate design)
/beta/daily/inventory/gp – list of gp region data
/beta/daily/inventory?region=gp -- list of gp region data (alternate design)
/beta/daily/inventory/all (provides a full list of all series)
/beta/daily/inventory (provides a full list of all series, alternate design)

☐ Time Series Data Queries
/beta/daily/series?sdi=sdi14&t1=2012-10-01&t2=2013-09-30[&format=csv|waterml2|hydrojson] – query data from October 1, 2012 through September 30, 2013 in format specified.
/beta/daily/series?sdi=sdi1,sdi24&back=30[?format=csv|waterml2|hydrojson] --- query the last 30 days for the time series data specified by sdi1 and sdi24 (multiple locations)
/beta/daily/series/sdi1 -- get all data for the series indicated by sdi1
.