IN23B-0602

HIS Modernization

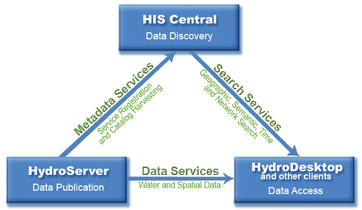
Integrating the Hydrologic Information System (HIS) with Dendra to Provide Enhanced User Experience and Real-time Management Capabilities

Martin Seul[1], Collin Bode[2], Scott Smith[2]
1.Consortium for the Advancement of Hydrologic Science, Inc. 2. University of California, Berkeley

www.cuahsi.org

HIS History

The Hydrologic Information System (HIS) system is a 20 year old internet-based system for sharing hydrologic data. Its purpose is the publication, discovery and access of data. It started as a federated system of databases connected via web services to client applications. It was developed by the Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI)



Why Modernize

Hydrologic research relys on time-series data. CUAHSI is a major provider of cyberinfrastructure for the hydrologic community. HIS was primarily for published datasets. The community asked for a cost-effective way to manage streaming sensor data. With streaming comes a need to manage the sensor deployment.

HIS was built on Microsoft products. High licensing fees and monolithic proprietary components make it difficult to maintain. The Federated design, i.e. many servers with a central catalog, proved impractical. Controlled vocabularies diverged. Always-on internet made centralization an obvious choice. All HIS servers are now centralized on Azure, but the architecture is still federated. The system design cannot scale up easily.

Key Requirements

- Real-time streaming sensor data
- Management of sensor deployments
- Leverage existing open source technologies
- Migrate existing data into a new system
- No downtime during modernization
- Fully cloud based
- Reduced costs
- Backwards compatibility

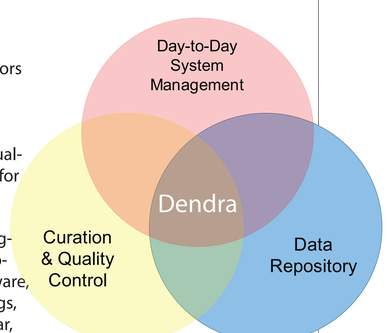
Chosen Approach

Partnered with Dendra Science to replace the existing HIS system, while maintaining existing APIs, data, search, and metadata services. Contracted with Eastern Research Group, Inc. (ERG) to help with developing the new system.

- Develop metadata conversion from the Observations Data Model to Dendra schema.
- Data migration from the existing system into Dendra.
- Deploy Dendra services to CUAHSI's the commercial cloud (GCP).
- Development of a WaterOneFlow (WOF) compatible service in Dendra.
- Development of an improved UI for a common administration interface.

Introducing Dendra

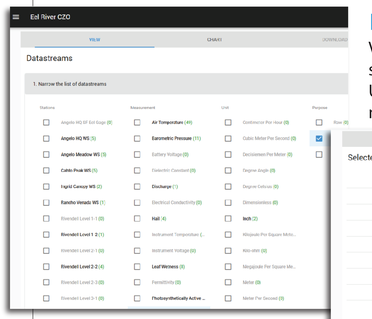
Dendra Science, an EarthCube funded project (NSF EAR-2126386), is a cloud-based service for managing sensors and time-series data. It can handle any kind of data paired with a timestamp. Dendra's strength is pulling data in real-time from sensors, coupled with managing equipment status, issue reporting, quality control, while providing visualization & distribution. This combination of services allows for efficient management of large sensor networks.



Dendra has integrations for many major vendors of dataloggers and most kinds of communications. This includes support, for example, of Campbell Scientific's LoggerNet software, Onset Computing's HOBOlink system, and Internet of Things, The Things Network (TTN). Communications can be cellular, WiFi, LoraWAN, commercial satellite, GOES, Iridium, SWARM, or simply a manual upload of a CSV file.

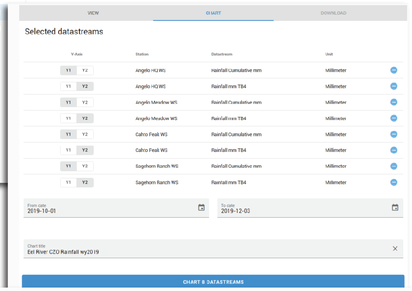
Dendra Key Features

Modern Search, Graph & Download



Faceted Search for Data Discovery

Visual filter that helps inform the user as they select vocabulary terms. Stations, Measurements, Units grey out if there are no Datastreams that match. Counts of Datastreams listed for matches.



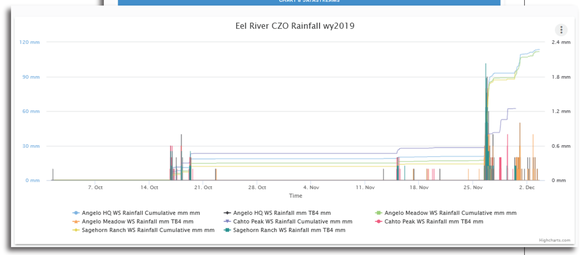
Download

Dendra's export system can download huge datasets without running out of memory.

Download functionality is accessible via API as well.

Graph

Plotting Datastreams can use both Y-axes. Multiple plots can be placed.



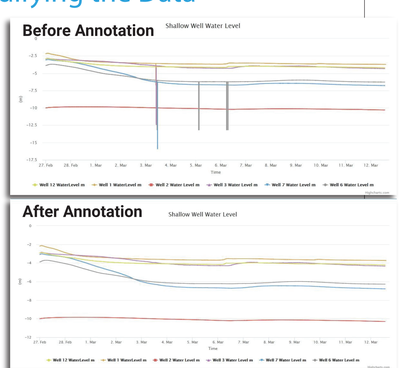
Real-time Alerts & Equipment Status

A critical need for large-scale monitoring is real-time status reporting. Dendra has a Status page showing last timestamp each station reported in, and a sparkline of battery voltage. Email alerts are sent if a logger has not reported in for 24 hours.

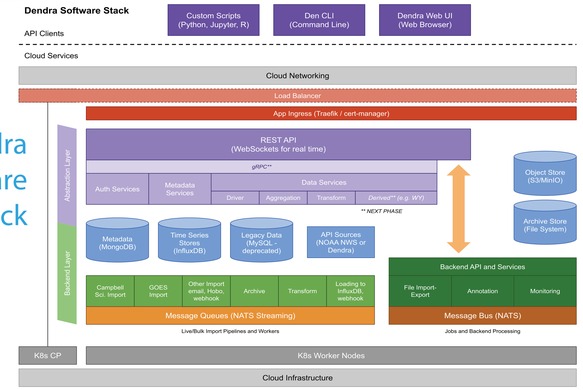
Data Cleaning without Modifying the Data

Dendra has the ability to dynamically version datasets. This is done through Annotations.

- Annotations are used to record events that happen to equipment, the environment, or data.
- Annotations are only applied when data is retrieved.
- Source data is unchanged. This allows the user to "roll back" changes to an earlier time, if needed.



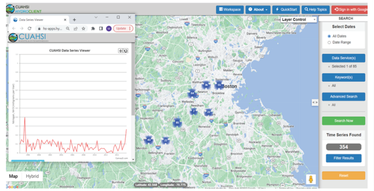
Dendra Software Stack

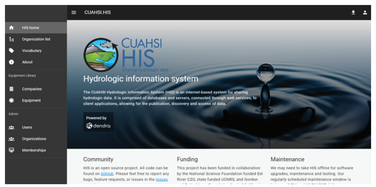


Outcomes

Data Migration Results

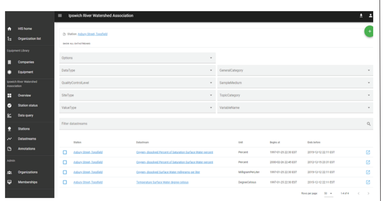
- 84 SQL Server instances migrated into Dendra (InfluxDB).
- Developed metadata mappings from ODM 1.1.1 to Dendra schema.
- Added ODM vocabulary to faceted search.
- WaterOneFlow wrapper for Dendra deployed.
- HIS-Dendra Staging environment mimics HIS for most services.
- HIS Central harvested metadata from Staging.
- Hydroclient interoperates with Staging for search and download.





CUAHSI Customizations

- Dendra is hosted by ACCESS on Indiana U. Jetstream2 supercomputing facility. Deployed containers for HIS-Dendra to Google Cloud Platform (GCP).
- Container orchestration designed for reproducibility and future expansion.
- CUAHSI branding on Dendra UI.
- Added new management features to Dendra UI for Users, Memberships, & Organizations.



Next Steps

- Develop backup and restore process and policies.
- Implement existing Dendra cloud CSI LoggerNet VM for streaming data.
- Documentation: system and user.
- Develop outreach and engagement plan.
- Train CUAHSI staff to support the system.

Acknowledgements

- National Science Foundation Award (EAR-1849458)
- BHP Foundation
- EarthCube (NSF EAR-2126386)
- CA Dept. of Fish & Wildlife (P2091015)
- Eel River Critical Zone Observatory (ERCZO; NSF-EAR1331940)
- University of California Natural Reserve System (weather station budget)
- Gordon and Betty Moore Foundation's California Heartbeat Initiative (7063)

Contact: Martin Seul, mseul@cuahsi.org