Implementation Plan for Interoperability and Interaction with the Broader IoW Network

Prepared and submitted by the Internet of Water with funding from the Cynthia & George Mitchell Foundation

Internet of Water

CONTENTS

INTRODUCTION

The developing Texas Water Data Hub represents a first step in realizing the vision for "an ideal water data system for Texas" that allows those tackling water problems timely access to relevant data about water in Texas. The Internet of Water (IoW) shares a similar vision but is applied across the United States. As such, the collaborative project between the IoW and the TWDB lays the groundwork for the Texas Water Data Hub to interconnect with the broader internet of water, including the water data systems of federal agencies and neighboring states.

Currently, the Texas Water Data Hub provides an online location where metadata regarding many water datasets can be searched, directing users to where data of interest can be accessed. As part of this project, the IoW has developed the following implementation plan for the continued development of Texas Water Data Hub capabilities that will serve to make its data findable, accessible, and interoperable with data from other hubs within the internet of water. While this document was developed for the Texas Water Data Hub it is broadly applicable and can be used to support the development of other hubs in the future.

IMPLEMENTATION PLAN

The broader internet of water is based on the idea of hubs being able to serve out metadata and data about a particular type of water data in the same format, so that a user can get data from different hubs about the same topic seamlessly. However, different data types may require different formats and require different kinds of metadata to be usable, so this activity needs to be done on a data type basis. For example, groundwater monitoring wells are associated with a point location on earth and a time series of sensor data, whereas data about a groundwater management district might be associated with a polygon and less frequent reports. In general, the implementation plan should proceed as follows:

1. Identify key data types for standardization and integration with IoW
2. Choose a type for integration
3. Develop type-wide location-based metadata
4. Publish location-based landing content
5. Serve data for type using APIs, standardizing if feasible
6. Link metadata to data
7. Repeat 2-6 for all types identified in step 1.

Each step is elaborated in detail below.

STEP 1: IDENTIFY KEY DATA TYPES FOR STANDARDIZATION AND INTEGRATION WITH IOW

There are many datasets from many organizations that may come to be represented in the Texas Water Data Hub. Many of these datasets are on similar topics. For example, several organizations may operate their own Spring monitoring programs. In this step, datasets can be grouped into similar topics. Based on previous work that prioritized use cases and associated data, particular topics can be prioritized for standardization based on Texas stakeholder needs. These topics should be targeted for Internet of Water integration, because the marginal effort required is lower if these datasets are being standardized for Texas purposes anyway, and the value added by leveraging external data made available by the Internet of Water can more quickly be brought to bear on Texas water issues. Examples of these cross-organization data types include but are not limited to:

- Spring monitoring sensors
- Steam gages

- Monitoring wells
- Surface water quality samples
- Precipitation stations
- Management district aggregate water use or availability data

STEP 2: CHOOSE A TYPE FOR INTEGRATION

Criteria should be established among the enumerated data types developed in Step 1 to choose which among them should be integrated. Such criteria may include the importance of related use cases and the current ease with which existing constituent datasets of each type can be integrated currently. Determinants of ease include the availability of metadata and data in machine-readable formats and the quality of site-level metadata from each source. Proceeding with dataset types that are presently easier to integrate has the advantage of being easier to standardize in the further steps, developing confidence among stakeholders that such efforts are both doable and worthwhile. To begin with, one dataset type should be chosen. In Texas, the first suitable data type may be spring monitoring sensors, where progress on cataloging available spring monitoring networks has been most advanced.

STEP 3: DEVELOP TYPE-WIDE LOCATION-BASED METADATA

Most water data types are associated with particular places, such as point locations like stream gages, or management areas like groundwater districts. Data to be integrated within a hub and with the greater IoW should develop and standardize location-based metadata across sources. For example, if the Texas Water Data Hub is to standardize spring monitoring data, it should develop a standardized catalog of spring monitoring locations. This would involve collecting a common set of metadata from the contributing spring monitoring networks, such as those operated by USGS, TWDB, and TCEQ. Such metadata elements may include, but not be limited to:

- Texas Water Data Hub-wide location identifier
- Contributing/Monitoring organization
- Organization-specific Location identifier
- Location name
- Location (latitude and longitude with datum)
- Hydrologic address in the National Hydrography Dataset (if applicable)
- Aquifer identifier (if applicable)

This exercise accomplishes two important outcomes for future data management:

1. A full picture of the spatial extent of the data type within Texas, which can be useful for understanding what water management issues can be plausibly tackled with current data collection activities, and where data collection may need to be augmented.
2. A framework to identify "duplicate" locations that are monitored by multiple organizations.

For each data type that type-wide metadata is developed for, the resulting metadata can be published as a dataset in its own right in the Water Data Hub for use by interested stakeholders in the short term.

STEP 4: PUBLISH LOCATION-BASED LANDING CONTENT

The data type-wide metadata developed in step 3 can be integrated into the Internet of Water. The following substeps are necessary:

1. Publish the metadata such that every location has a unique web page with a unique URL. For example, https://texaswaterdatahub.org/locations/springs/<identifier>
2. Embed a JSON-LD version of the metadata elements within each page. This is similar to the schema.org markup used for search engine optimization for personal and commercial websites. The Internet of Water will be developing curating specific guidance for this at https://docs.geoconnex.us, but Internet of Water staff can also be contacted directly for assistance and advice.
3. Register persistent identifiers with a predictable scheme that redirects to the URLs published in step 4.1, with https://geoconnex.us. For example, https://geoconnex.us/texaswaterhub/springs/<identifier>
4. (Optional) Add metadata to the USGS Network Linked Data Index, a system allowing locations to be indexed to the National Hydrography Dataset to be searched up and downstream alongside any other contributed datasets (e.g. USGS, EPA, CUAHSI, and neighboring state monitoring locations). Internet of Water staff are available to assist in this process.

The simplest way to accomplish steps 4.1-4.3 would be to use pygeoapi, an open-source python server that the USGS and Internet of Water have contributed features to specifically to facilitate these steps. However, they can be accomplished with custom software using html templating frameworks such as Mustache or Jinja.

STEP 5: SERVE DATA FOR DATA TYPE WITH APIS, STANDARDIZING WHEN FEASIBLE

Ideally, any data within the data type for a given theme would be available through an API. Organizations may already make data available via API, such as USGS through NWIS or the TWDB through TexMesoNet. Organizations that do not have APIs may be engaged with to allow copies of their data to be hosted in Texas Water Data Hub, where an API version of the data can be served via the CKAN or ESRI platforms. In some cases, it may be worth exploring the deployment of a purpose-built standard API and database for observations data such as the OGC SensorThings API, which has recently been deployed by the USGS as a core service of the next generation of NWIS services. This choice represents a higher degree of difficulty to implement but would provide the added value of Texas Data being available from the exact same API as USGS and other participating state data organization efforts such as the New Mexico Water Data Initiative, easing the development of web applications and analysis workflows using data from across such organizations. In any case, ideally what should be available is for the actual water data from each location in the type-wide metadata to be available at a specific URL, so that it can be accessed and downloaded directly if a data user has found and accessed the specific location's metadata.

STEP 6: LINK METADATA TO DATA

Texas Water Data Hub Integration with the wider Internet of Water is complete for a given data type when URLs at which data for a particular location is findable alongside metadata about that location, along with spatially or hydrologically related locations from other hubs. This can be accomplished if links to the data URLs are embedded in the JSON-LD of the location metadata webpages developed in Step 4. The Internet of Water is still developing the best practices for how these links should be formatted at https://docs.geoconnex.us but staff can be engaged with directly for advice in the short term. Once this is accomplished, it becomes possible to build applications that allow users to query metadata and data for a particular type across all participating providers, including the Texas Water Data Hub.

STEP 7: REPEAT STEPS 2-6 FOR ALL DATA TYPES IDENTIFIED IN STEP 1

Each data type represents a distinct effort, since every data type is different, with different metadata and data standardization requirements depending on the associated

use cases. Thus, steps 2-6 must be repeated for each data type. In addition, they should be repeated periodically even for completed data types to maintain the resulting data system over time as data collections throughout the state evolve, and as data and API standards in the water sector evolve as well. Integration with the Internet of Water is always a work in progress and should be made an integral part of data stewardship after the initialization of the Texas Water Data Hub, to ensure that interoperability with the wider network is maintained over time.