2021-04-03

Introducing the Water Data Explorer Web Application and Python Library: Uniform Means for Data Discovery and Access from CUAHSI and the WMO WHOS Systems

Elkin Giovanni Romero Bustamante
Brigham Young University

Introducing the Water Data Explorer Web Application and Python Library:

Uniform Means for Data Discovery and Access from CUAHSI

and the WMO WHOS Systems

Elkin Giovanni Romero Bustamante

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

James E. Nelson, Chair
Daniel P. Ames
Gustavious P. Williams

Department of Civil and Environmental Engineering

Brigham Young University

ABSTRACT

Introducing the Water Data Explorer Web Application and Python Library:
Uniform Means for Data Discovery and Access from CUAHSI
and the WMO WHOS Systems

Elkin Giovanni Romero Bustamante
Department of Civil and Environmental Engineering, BYU
Master of Science

There has been a growing recognition in recent years of the need for a standardized means for sharing water data on the web. One response to this need was the development of the Consortium of Universities for the Advancement of Hydrologic Science (CUAHSI) Hydrologic Information System (HIS) and its accompanying WaterOneFlow and WaterML protocols. To date, the primary means for accessing data shared using these protocols has been limited to the Microsoft Windows HydroDesktop software, the WaterML R package, and the web based CUAHSI HydroClient which serves as an access point to the CUAHSI HIS database. We recognized the need for a new web-based tool for accessing data from any system that supports WaterOneFlow web services and WaterML and that could be regionally customizable, giving access to the most locally relevant portions of the HIS database, and providing a means for international government agencies, research teams, and others to make use of the accompanying protocols on a locally managed web application. To fill this need, we developed the open source, lightweight, installable web application, Water Data Explorer (WDE) which supports any WaterOneFlow service and can be customized for different regions containing WaterOneFlow web services. The WDE supports data discovery, data visualization, and data download for the selected WaterOneFlow services. The WDE's structure consist of WaterOneFlow catalogs, servers, and individual measurement stations. The WDE provides a different User Interface for administrators and regular users. A server administrator can specify which datasets an individual instance of the WDE supports so that end users of the application can access data from the specified datasets. We modularized the core WaterOneFlow access code into a new open-source Python package called "Pywaterml" which provides the methods used by WDE to discover, visualize, and download data. This thesis presents the design and development of the WDE and the associated Pywaterml package, which was done in partnership with end-users from the WMO and was done in an iterative design-build process. We present two case studies which involve data discovery and visualization from the CUAHSI HIS and WMO Hydrological Observing System (WHOS). Both case studies demonstrate the regional customization of the WDE which allows creation of different custom versions of the same application to meet specific end-user needs. The WDE data discovery in both case studies focuses on discovering the different sites contained in a WaterOneFlow web service, and ontology-based data discovery for the different concept variables in each web service. The data visualization we present, focuses on the time series observation for the different sites in each system. Finally, we tested data downloading in data discovery and visualization by downloading the information of each site to the WDE database and allowing the user to download the time series data.

Keywords: observation networks, WMO, WHOS, CUAHSI, Tethys, HydroShare, HydroServer

ACKNOWLEDGEMENTS

I would like to thank all the wonderful people that raised me when I was a child. My aunts, parents, brother and cousins that were always being patient with me throughout the years. Their mistakes and triumphs influenced the different paths that I have lived, and I am grateful for it. The product of this research will not be possible without the teachings and time that Dr. Nelson, Ames, and Williams provided me. I would like to thank Dr. Nelson for being a mentor and giving me the opportunity to be a master's student. Dr. Ames for giving me innovative ideas and helping me to polish the final product of my research, and Dr. Williams for being always available when I needed help and being willing to be part of my committee. I also feel that my research work will not be possible without the influence of the people in the Hydro informatics laboratory at BYU. Finally, I would like to thank the WMO for the funding and support in this research, especially Enrico Boldrini and Igor Chernov for their time and feedback.

TABLE OF CONTENTS

LIST OF FIGURES

1 THESIS STATEMENT

Data sharing of observation time series and data interoperability between data providers and users involves the development and use of cyberinfrastructure including data servers, data catalogs, and client tools for data access, visualization, and download. Such systems typically follow a Service Oriented Architecture (SOA) based on web services and standardized communication protocols that connect each of these components. The Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) Hydrologic Information System (HIS) (Tarboton et al., 2009), and the World Meteorological Organization (WMO) Hydrological Observing System (WHOS) (Boldrini et al., 2020) are examples of cyber infrastructure developed for hydrologic data using the SOA system approach. The specific architecture of each of these systems differs in terms of databases, software, and hardware, however, they both aim to provide data storing, sharing, and interoperability of discrete water resources observations while allowing data dissemination and sharing among users through WaterOneFlow services. The ability to have data dissemination and sharing capabilities available through web services opens the opportunity for data discovery, download and analysis from multiple datasets. Even though the different SOA architectures for hydrologic data have developed specific client component tools for the data discovery, download and analysis of the different datasets, there is not a client component that can be customized for specific regions and datasets from multiple SOA web services. The Water Data Explorer (WDE) objective is to serve as client component that can be customized for a particular region of interest while allowing data discovery, download and analysis of WaterOneFlow web services from different SOA systems. WDE will provide end users with the ability to customize their service for region of interest to provide data discovery, download and analysis

2 INTRODUCTION

2.1 Background

Monitoring water quantity and quality with observation networks is fundamental to study, manage, and treat water. Observation networks are used to monitor the distribution and nature of water, and they measure and record different variables related to water quantity such as precipitation, streamflow, water depth, and water quality parameters such as temperature, turbidity, and the concentration of phosphorous, nitrogen, and other chemical components. However, there is concern regarding the adequacy, consistency, data archiving and accessibility, and long-term maintenance of high-quality water data observations networks. According to the World Meteorological Organization (WMO), some of these challenges are declining network observations, matching data collected to their intended use, data warehousing and dissemination, and accounting for the uncertainties in networks for forecasting ("Challenges to Hydrological Observations," 2015).

Data archiving and accessibility is a challenge because it is often difficult to locate, obtain, and compare data between different regions, especially when different parties and agencies collect data and manage in different formats. Dissemination of water data is also difficult due to the lack of integration and interoperability across the various data archives within geosciences to find and access those archives. Data archiving and dissemination can be overcome if the architecture of the water observation data storage and dissemination system is developed with the aim of expanding operational data management across multiple users and institutions.

Data archiving among organizations generally involves relational databases that can be updated and accessed by different users. Similarly, dissemination of hydrological data from different countries at a transboundary level involves standardizing various aspects of the hydrological observation data and organizing the transfer of technologies to provide the hydrological data and information required for sustainable development of nationally and internationally shared water resources (UN-WATER, 2008). As a result, the technology to provide the hydrological data sharing and warehousing between national organizations requires cyberinfrastructure that allows data interoperability between the different data providers and systems from the different organizations.

A number of different cyberinfrastructures, using the principles of service-oriented architecture (SOA), have emerged for sharing and storing spatially discrete observation data, including the (CUAHSI) Hydrologic Information System (HIS) (Tarboton et al., 2009), the World Meteorological Organization (WMO) Hydrological Observing System (WHOS) (Boldrini et al., 2019), the Critical Zone Observatory Integrated Data Management System (CZOData) (Zaslavsky et al., 2011), the Integrated Earth Data Applications (IEDA) and EarthChem system (Lehnert et al., 2004, 2011, 2009), and the Integrated Ocean Observing System (IOOS) (IOSS, 2010; Lubchenco, 2010). The core focus of these service-oriented architecture systems is on publishing or sharing data on the internet via web services such as the WaterOneFlow protocol and markup languages allowing data interoperability and availability among users. SOA systems also accomplish discovery and visualization through a client component tool, which is a platform that communicates with the different data services and the registry of the SOA system. For example, the CUAHSI HIS system has different client components such as HydroDesktop (Ames et al., 2012). Similarly, the WHOS system can use HydroDesktop or the web client from WHOS (http://www.wmo.int/pages/prog/hwrp/chy/whos/).

In the following sections of this introduction, we will address two different SOA system architectures: the CUAHSI HIS and WMO WHOS, and the information model and the web services used by both SOA systems. We will also address research objective of the WDE as a client component to provide data discovery, downloading, and visualization of time series observation and its role in both the CUAHSI HIS and WHOS systems in the following sections.

2.2 Information Model

SOA systems such CUAHSI HIS and WHOS WMO relies on common information model for organizing, storing, and publishing observational timeseries collected at point locations in order to identify the informational elements that are common across all data sources. The information model also contributes to the representation of the semantics of the data from each source. A representation of this common model for time series collected at point locations is depicted at Figure 2-1. Part (a) from Figure 2-1 shows the high-level view of the information model, an organization operates a network of monitoring sites. At each monitoring site a number of variables are measured resulting in a time series of data values for each variable at each site. Each data series is made up of individual, time-indexed values as depicted by part (b) of Figure 2-1.

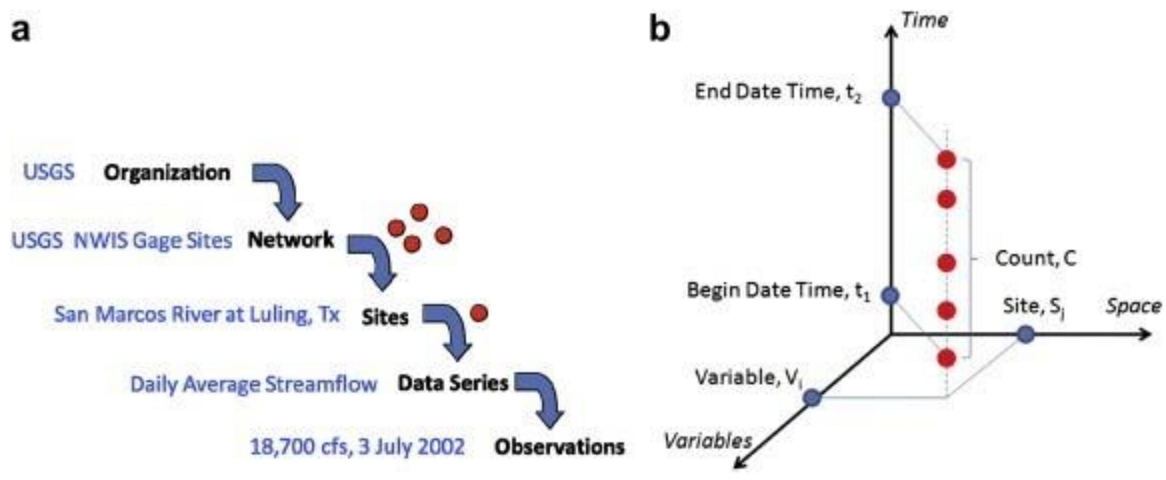

Figure 2-1: Simplified Information Model for Point Hydrologic Observations: (a) High-Level View of The Information Model, (b) Data Series Dimension (Ames et al., 2012).

The implementation of the information model can vary from one SOA system to another. For example, the CUAHSI HIS systems implement the information model using the Observations Data Model (ODM) (Horsburgh et al., 2008a) allowing the observations data to be stored with their associated metadata. Storing the metadata associated to the observation ensures that observations can be published and transmitted with their associated metadata through WaterOneFlow web services in WaterML format (Ames et al., 2012). As a result, the implementation of an information model allows WaterOneFlow web services to enable data discovery and access across the many different sources of observations data within the CUAHSI HIS.

2.3 WaterOneFlow Web Services

Web services are computer applications that interact with and exchange information with other applications over the internet. CUAHSI developed the WaterOneFlow web services as a standard mechanism for the transfer of hydrologic data between hydrologic data servers and users' computers. WaterOneFlow web services streamline the often-time-consuming tasks of extracting data from a data source, transforming it into a usable format and loading it into an analysis environment. WaterOneFlow web services allows data access through methods rather than requiring the user to navigate to the data provider's web page, query data, and save the data locally. The methods provided by the WaterOneFlow services are the following:

- GetSiteInfo — Given a site number, this method returns the site's metadata as a string
- GetSiteInfoObject — Given a site number, this method returns the site's metadata of an object.
- GetSitesXml — Given an array of site numbers, this method returns the site metadata for each as a string.
- GetSites — Given an array of site numbers, this method returns the site metadata for each as an object.
- GetVariableInfo — Given a variable code, this method returns the variable's name as a string.
- GetVariableInfoObject — Given a variable code, this method returns the variable's siteName as an object.
- GetValues — Given a site number, a variable, a start date, and an end date, this method returns a time series as a string.
- GetValuesObject — Given a site number, a variable, a start date, and an end date, this method returns a time series as an object.
- GetWaterOneFlowServiceInfo — Get all registered data services from a registered HIS catalog.

- GetControlledVocabulary — Get the terms and definitions of controlled vocabulary (CV), which are dynamically updated from a registered HIS catalog.

2.4 CUAHSI HIS System

As depicted in Figure 2-2, the CUAHSI HIS system is composed of three components which allows for storage and data publication using client tools for data search and discovery, download, and visualization (Ames et al., 2012a; Crawley et al., 2017a; Tarboton et al., 2009). The CUAHSI HydroServer is one of the components that accomplishes the role of a server tool for data publication. The second component of the system is the HIS Central catalog tool, it uses the Observation Data Model (ODM) model to structure a central metadata catalog database that supports data discovery services (Whitenack et al., 2010). HIS Central indexes both ODM and WaterOneFlow services to provide access to data served from an ODM database (Tarboton et al., 2014; Tarboton et al., 2009). The last component is the client component that includes both desktop and web interfaces which are able to retrieve local copies of observational data from HydroServers, taking advantage of the relational database schema from HIS Central.

2.4.1 HydroServer

CUAHSI developed HydroServer (Conner et al., 2013; Tarboton et al., 2010), which supports the sharing of distributed water data through OGC Web services and tools for data discovery, access, and publication (Horsburgh et al., 2010). HydroServer solves the problem of having different parties and agencies collect data in different formats that makes data archiving a challenge due to syntactic heterogeneity. HydroServer uses an ODM database, a relational model for storing, managing, and manipulating point observations data.

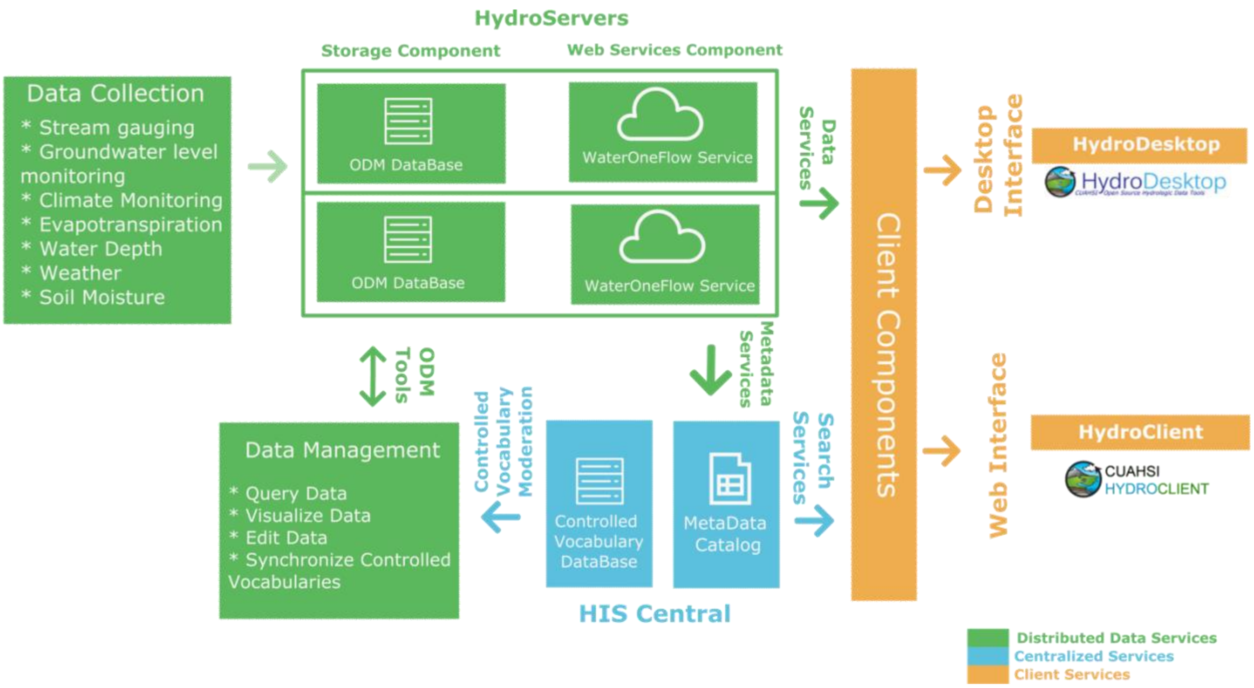

Figure 2-2: CUAHSI HIS Architecture

The ODM also provides physical implementations within various file systems and databases for data storage, using extensible markup language (XML) schemas and file formats for data transfer to overcome heterogeneity. The ODM database provides a common and encompassing database so that all of the observations, regardless of source, collection method, or original file type and format, can be stored along with their metadata (Horsburgh et al., 2008b). As a result, once different water data formats are loaded into an ODM database, they are syntactically similar and become available to analytical tools that exploit the WaterML format.

Once syntactic heterogeneity has been overcome, data from ODM databases can be accessed through web services such as WaterOneFlow. WaterOneFlow web services have been built to extract data from an ODM database based on a user defined query and transmit it over the Internet, preserving the syntactic homogeneity achieved by loading data into the ODM database (Horsburgh et al., 2008b). The ODM database and the WaterOneFlow web services serve as the basis for publishing point observations on a HydroServer. In other words, the different point observations can be loaded into one or more ODM databases, each of which is then connected to a WaterOneFlow web service, which transmits the data over the Internet in WaterML format.

2.4.2 HIS Central

HIS Central is part of the CUAHSI HIS system, and it enables data discovery of water data services from HydroServers. HIS Central is the web application which provides an interface for adding and managing registered water data services and the HIS central metadata catalog (Whitenack et al., 2010). The central metadata catalog maintains observation series information, including 1) site information, 2) variable information, 3) period of record, and 4) project metadata for all registered data sources of hydrologic observations.

HIS central users can register HydroServer WaterOneFlow data services to the central registry, making it possible for other users to discover and access the data and integrate it with information from other similarly registered data services. In addition, the metadata catalog also maintains a hydrologic ontology in order to solve the semantic heterogeneity challenge by using standardized concepts across multiple water data services using different semantics for their variables. Finally, once data have been published and registered in HIS Central, the metadata catalogs can be discovered by different client components.

2.4.3 CUAHSI HIS Client Components

Different HIS client components can be used with the CUAHSI HIS System and allow users to search available WaterOneFlow services. These include CUAHSI HydroDesktop (Ames et al., 2012) which supports the discovery and retrieval of data hosted on any of the distributed hydrologic data servers registered with the CUAHSI HIS system. In addition, HydroDesktop provides an interface with multiple functionalities: discovery of hydrologic time series data; map-based visualization of monitoring locations and other geographic information systems (GIS). The HydroDesktop application also supports data download, organization; visualization, editing, maintenance of hydrologic time series, and linkage with integrated modeling systems such as OpenMI (Gregersen et al., 2007).

CUAHSI has also developed the HIS HydroClient (https://data.cuahsi.org/), which is a web application alternative to HydroDesktop. HydroClient provides a basic functionality to discover time series by specifying the geographic extent, variables of interest, and specific published services from the CUAHSI catalog. HydroClient searches the catalog and returns a list of time series data that are plotted on the map. The points that are close geographically are grouped into a cluster, then markers can be explored by clicking on them. The different time series are contained in the group marker and are returned in a table format to explore and download the data as depicted in Figure 2-3. Time series data can be downloaded in a CSV format. Users are also encouraged to add time series to their workspace that creates files in WaterML format that can be used as input for other apps.

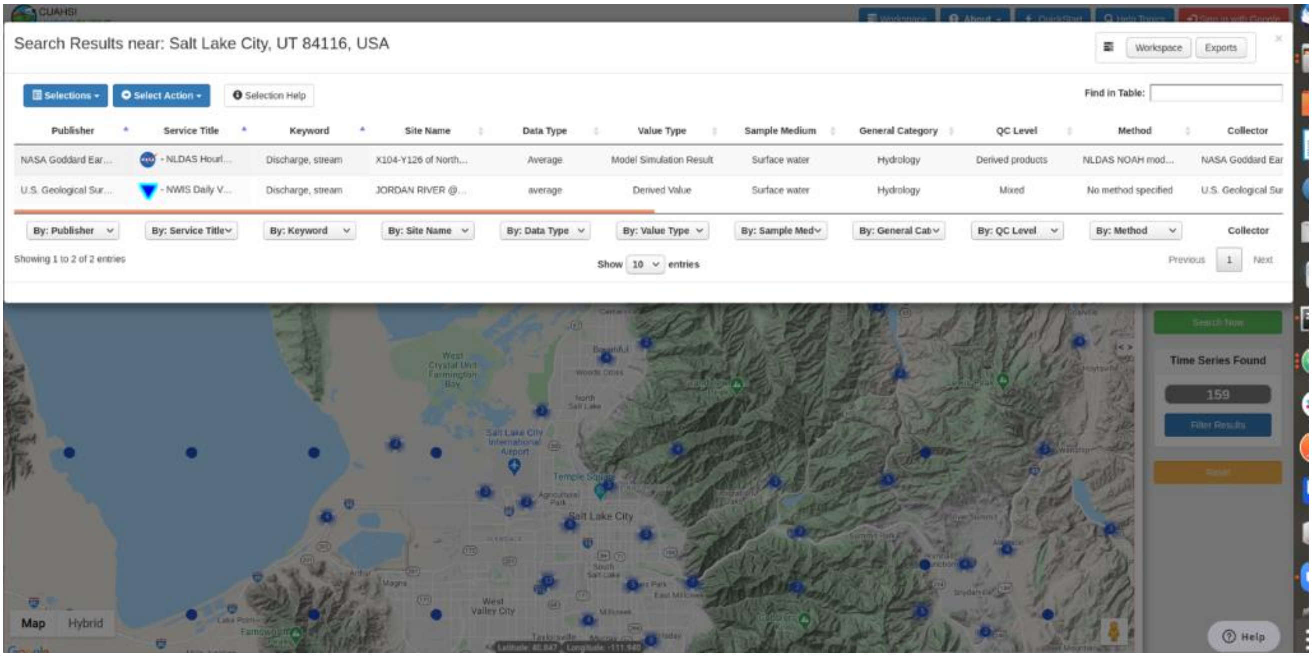

Figure 2-3: CUAHSI HydroClient Web Interface

2.5 WMO Hydrological Observing System (WHOS)

In 2014, the WMO initiated the development of the WMO Hydrological Observing System (WHOS), with the aim of providing a services-oriented framework linking hydrological data and users through an information system enabling data registration, data discovery, and data access (WMO and WIGOS, 2019). WHOS provides the possibility to exchange data among heterogeneous data providers (e.g., National Hydrological Services) through the already existing publication systems used by them. WHOS is designed to overcome the various challenges associated with the discovery, access, interoperability and reusability of data from heterogeneous web services. Data discovery and access can be performed uniformly through the WHOS data exchange allowing various applications and modelling systems for different regions to be built more easily.

WHOS also supports the publication of customized subsets of data shared by various data providers through the concept of “views”. E.g., a “basin view” can contain all the data sets that are collected in a specific basin by neighboring countries Figure 2-4. Different standardized data formats and service protocols are supported by WHOS to publish shared data of the different “views” by means of service interfaces required by different clients. Therefore, clients components can access the WHOS “views” or customized datasets for data discovery, download and visualization through WaterOneFlow web services.

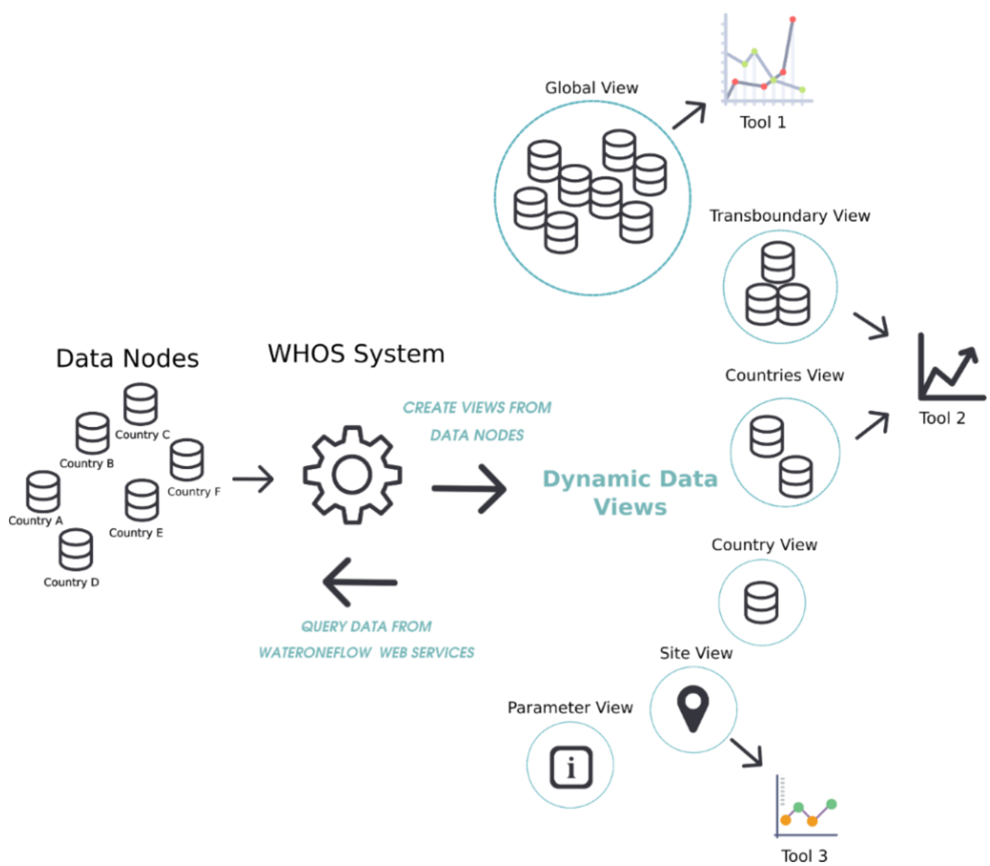

Figure 2-4: WHOS Dataset Interoperability and Customized Data Views

WHOS is built around seven fundamental components including: (1) data, (2) format, (3) service, (4) mediator, (5) broker, (6) ontology and (7) client (Pecora and Lins, 2020). Each logical WHOS component serves a specific purpose inside the WHOS architecture. A key component of the WHOS architecture is the broker which consists of three parts: discovery broker, access broker, and semantic broker. The discovery broker allows for data discovery in heterogeneous data providers by means of heterogeneous search clients. The semantic broker allows uses to discover concepts and terms from heterogenous ontologies, and the access broker allows uses to access datasets from heterogenous data providers by means of heterogenous import tools. As a result, each of the broker parts provides specific functions needed for distributed and harmonized data discovery and access by different client applications.

2.6 Research Objectives

This paper describes the design and development of WDE, a web-based client tool, that enables the discovery and retrieval of data hosted on any of the distributed hydrologic data servers from the CUAHSI HIS and WHOS systems. WDE can be thought of as tool that allows users to have a workspace in which they can create and customize catalogs containing datasets from different SOA systems. The WDE provides data discovery and access functionalities that allow data importation through WaterOneFlow web services. Ontology-based search of imported data is also possible through the WDE using WaterOneFlow web services. Finally, WDE provides data visualization and downloading functionalities when working with catalogs customized by users.

We developed the WDE application using the Tethys Platform framework (Nelson et al., 2019; Swain, 2015; Swain et al., 2016). Tethys Platform provided the framework and tools to build the WDE which serves as a cataloging tool for data discovery, download, and visualization of data from WaterOneFlow services. WDE can be customized for a specific region, and create different catalogs for that specific region, providing the capability for WDE to be deployed by any organization implementing a data interoperability through WaterOneFlow web services. As a result, the regional customization and cataloging ability permit the WDE application to not only serve as a client component for a specific SOA, but to serve as a regional tool to manage WaterOneFlow web services from different SOA systems. This means that the WDE can be customized for a particular region and serve as a client component to manage discovered data from one or more SOA systems. As part of WDE development we created the pywaterml python package which provides the different data discovery operations of the WDE through WaterOneFlow web services. The pywaterml package can also be used in the development of other client components involving WaterOneFlow web services, or as a standalone python package in python notebooks for further development.

We present two experimental cases studies for the WDE application to test the different functionalities of the application in the CUAHSI HIS and WHOS systems. We test the CUAHSI HIS Central Catalog WaterOneFlow registered services within the different 98 regions available. Similarly, we test the different WaterOneFlow web services provided from the WHOS for the La Plata Basin and Artic regions.

3 SOFTWARE DESIGN

3.1 Components

The WDE is composed of two main components the Tethys Platform framework, and the pywaterml python package. The Tethys Platform framework provides the WDE with the regional customization ability and the software development kit (SDK) for the app development. Tethys Platform served to implement all the WDE functionalities though the use of third-party and built-in libraries and the User Interface (UI) for the different users. On the other hand, the pywaterml provides the WDE with the ability to use WaterOneFlow methods responses in different formats. The WDE relies on the information model described in section 2.2. When the WDE makes data request calls using the different WaterOneFlow web services methods supported in each SOA system, the structure of the response data follows the information model: Organization, Network, Sites, Data Series, and Observations. The WDE uses the response data from the different WaterOneFlow web services methods described in section 2.3 for different purposes. For example, the WDE is able to use the response data from the different methods to import catalogs, visualize time series plots, and get information about a specific site.

3.1.1 Pywaterml Python Package

We created the Pywaterml package to allow WDE to connect to any of the CUAHSI HIS Central and WHOS-broker customized datasets ‘WaterOneFlow’ web services and read any ‘WaterML’ time series data response. The Pywaterml python package provides users with simple

API calls to execute WaterOneFlow web services methods and receive responses in JSON, WaterML and CSV format as depicted in Figure 3-1 and outlined here:

1. The Pywaterml package sets up the connection with any WaterOneFlow web service using lightweight SOAP-based web service client for Python from third-party libraries.
2. Pywaterml enables six standard WaterOneFlow web services methods to retrieve different types of data over the established connection; two customized methods, extending the standard WaterOneFlow web services methods; three additional analysis methods to analyze the data retrieved from the WaterOneFlow web services methods.
3. Responses from the different methods are formatted in three different standards and used for data discovery, download, and visualization in the client component.

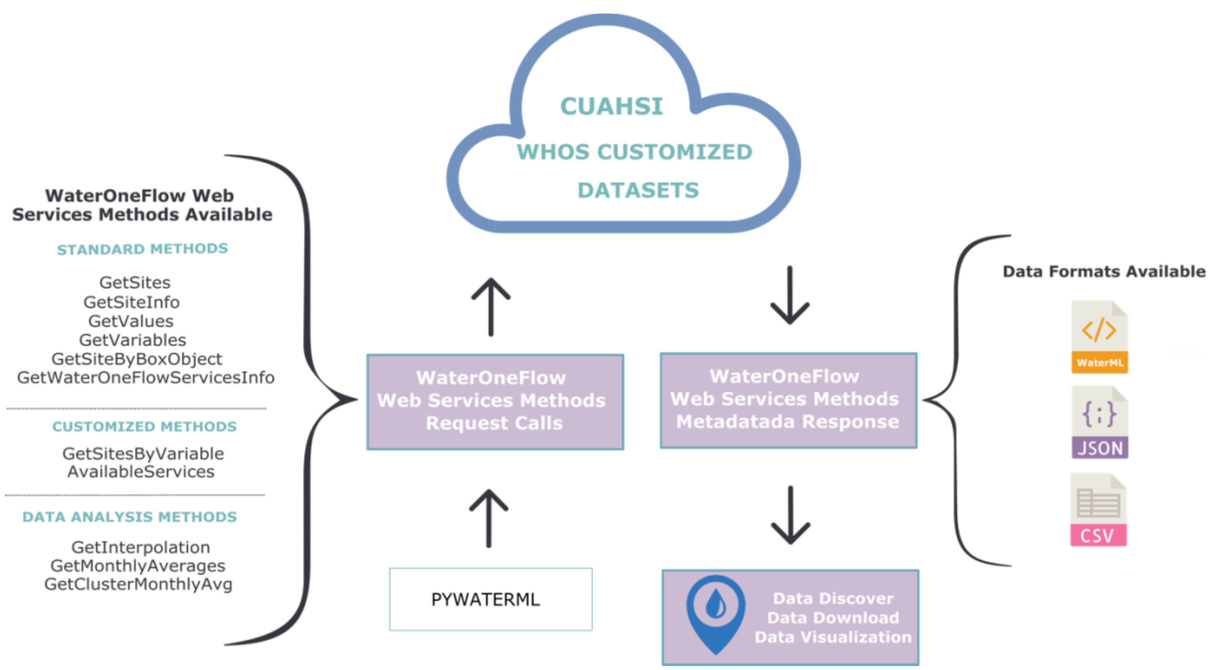

Figure 3-1: Pywaterml Package Functionality

There is a total of 11 methods in the Pywaterml. The different methods provided by the pywaterml python package can be divided into three categories: standard WaterOneFlow web services methods, customized WaterOneFlow web services methods, and analysis data methods. The methods in the first category are named after the standard WaterOneFlow web services methods from which they retrieve data. The methods in the second category provide functionality that is not found in the standard WaterOneFlow web services such as retrieving sites by a list of variables. The methods in the third category provides users with methods to perform interpolation, calculate monthly averages, and perform machine learning clustering in the sites.

The six standard WaterOneFlow web services methods provided by the pywaterml retrieved the same data that the methods explained in section 2.3, which are the following: i) GetSites, ii) GetSitesByBoxObject, iii) GetVariables, iv) GetSiteInfo, and v) GetValues, vi) GetWaterOneFlowServicesInfo. On the other hand, the other 5 remaining methods add functionality to the data retrieved from the standard WaterOneFlow web services, and they are the following:

1. GetSitesByVariable: Retrieve sites containing specific list of variables.
2. AvailableServices: Get the WaterOneFlow web services and metadata from the available dataset that are contained in a HIS catalog
3. GetInterpolation: Interpolates the data given by the GetValues function in order to fix datasets with missing values. Three types of interpolations are offered: mean, backward, forward. The default is the mean interpolation.
4. GetMonthlyAverages: Gets the monthly averages for a given variable, or from the response given by the GetValues function for a given site.

5. GetClustersMonthlyAvg: Gets "n" number of clusters for the time series data of a particular variable for the sites containing such variable. The clustering algorithm used is the k- means with the use of a Dynamic Time Warping (DTW) for the core metric in a *k*-means algorithm.

3.1.2 Tethys Application Framework for Customized Views

We developed the WDE using the Tethys Platform framework which allows developers to create customized versions of the WDE for different users and organizations. Tethys Platform consists of three major components: Tethys Software Suite, Tethys Software Development Kit (SDK), and Tethys Portal as depicted in Figure 3-2. Tethys Software Suite includes software for file dataset management, user account management, spatial database storage, geoprocessing, mapping and visualization, and distributed computing. The user account management and file dataset management make it possible to provide different privileges to administrators and regular users.

Tethys also brings the ability to have different databases. Allowing application such the WDE to download and store data retrieved from the SOA systems using WaterOneFlow web services. The Tethys SDK allows the WDE to be customized as needed using the Application Programming Interfaces (APIs) for each element of the software suite. The WDE used the App Settings and Persistent Storage API to customize a name, link a Web Mapping Service (WMS) to be a boundary, and to assign a database. As a result, multiple customizations can result from the different combinations of given names, WMS boundary layers, and databases assigned to the WDE. The Tethys Portal provides a website user interface that facilitates administrator users to access and customize the WDE in the settings page and assign a name, WMS boundary layer and database. In summary, the Tethys SDK serves to develop the WDE using the Tethys Software APIs for app custom settings and persistent storage while the Tethys Portal allowed users to install a generic version of the WDE and customize it through a settings page.

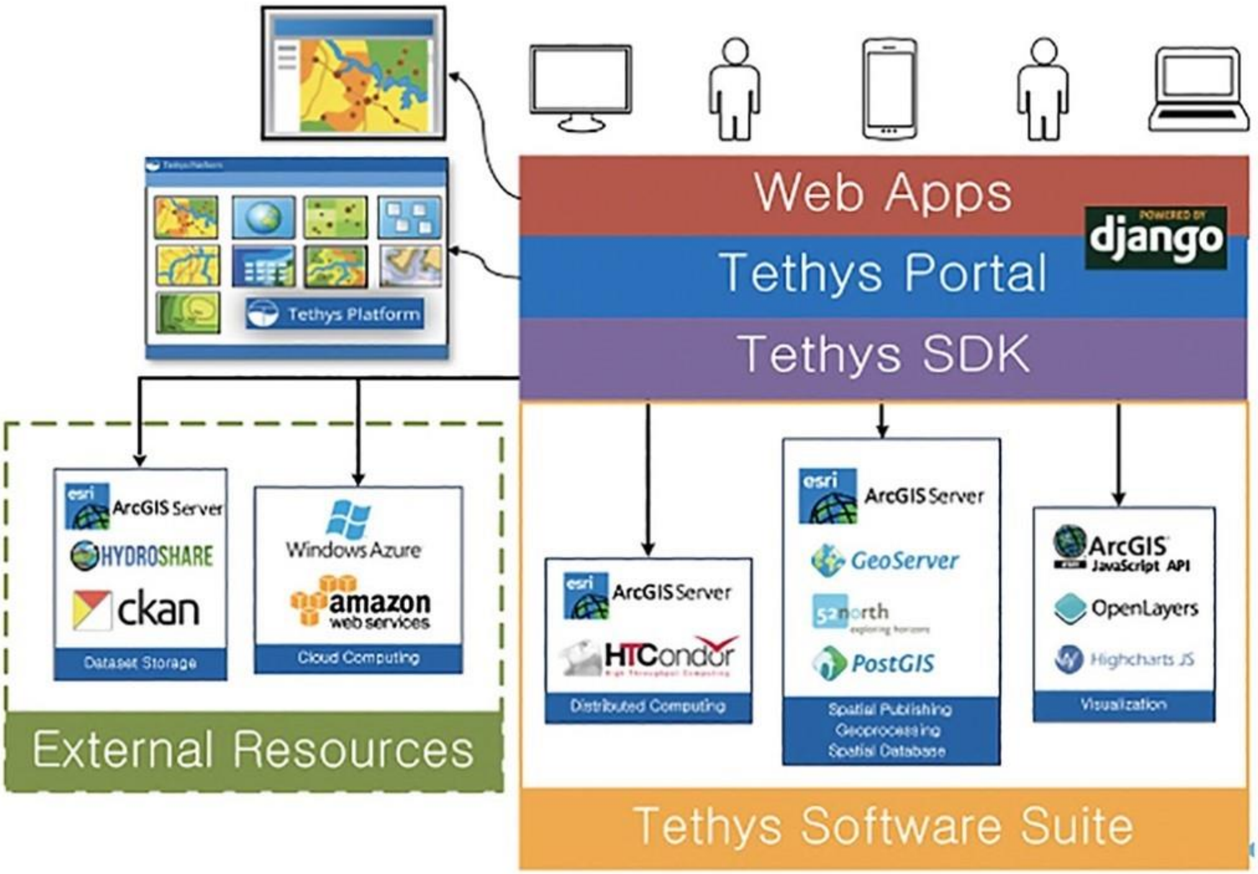

Figure 3-2: Tethys Framework Components (Nelson et al., 2019).

3.2 WDE Catalog Management

The WDE can serve as a client component for the CUAHSI HIS system because of its ability to retrieve data from systems that support WaterOneFlow web services. Similarly, The Water Data Explorer can leverage the WHOS “views” functionality by presenting the relevant “views” to users. The WDE is able to organize the different WaterOneFlow web services and their associated metadata by creating and managing catalogs, which are groups of WaterOneFlow web services containing the discovered metadata associated to them. The WDE can create catalogs from existing SOA WaterOneFlow web services containing other web services or create a catalog from diverse WaterOneFlow web services. For example, the CUAHSI HIS contains one central Catalog (HIS Central) containing different registered WaterOneFlow web services. The WDE is able to create a catalog from the selected WaterOneFlow web service provided by the HIS Central or from the entire catalog. The created WDE catalog will contain the selected WaterOneFlow web services and metadata associated to the imported HIS catalog. Similarly, another catalog can be created from the WaterOneFlow web services from the CUAHSI HIS and WHOS WMO. This implies that the WDE can create catalogs containing any combination of WaterOneFlow web services. WaterOneFlow web services can belong to one SOA system or can be a combination of different WaterOneFlow web services from other SOA systems.

Figure 3-3 depicts the process of creating a catalog for the CUAHSI HIS and WHOS WMO, and the different possible WaterOneFlow web services that can be part of this new catalog. First, the WDE creates an empty catalog. Second, the WDE request the CUAHSI HIS or WHOS WMO for metadata through WaterOneFlow web services method calls. Third, the necessary response metadata related to the selected WaterOneFlow web services methods are retrieved to the WDE. Fourth, the catalog is saved in the Catalog schema from the WDE database, and it is available in the WDE list of catalogs for management.

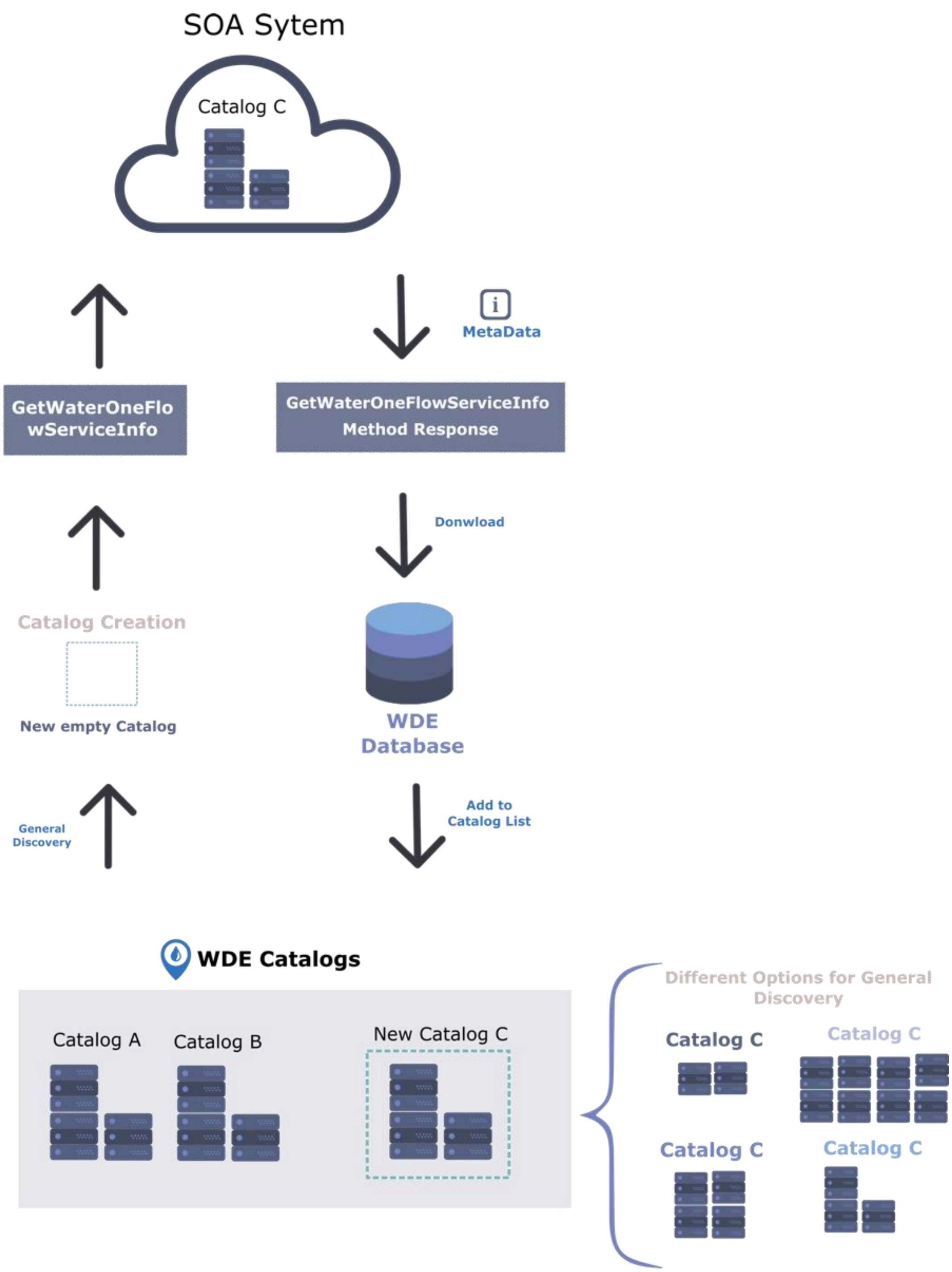

Figure 3-3: WDE Use of Catalogs in SOA Systems

New web services can be added to or deleted from the WDE catalog after its creation in order to allow users to customize the number of WaterOneFlow web services that each catalog contains. In addition, empty catalogs can be created, and WaterOneFlow web services added later by the users. In summary, the ability to create, manage, and customize catalogs allows the use of multiple SOA systems in one single client component for visualization and data discovery.

3.3 WDE Levels

The WDE uses Catalogs to organize and manage the different WaterOneFlow web services. The WDE can create catalogs from existing SOA WaterOneFlow web services containing other web services. This means that there are WaterOneFlow web services for SOA HIS Catalogs such as the HIS Central Registry or a WHOS view containing information of other web services. However, there are also WaterOneFlow web services that do not serve the purpose of a HIS Catalog, providing metadata for other WaterOneFlow web services. The difference between these two types of WaterOneFlow web services can be found in the termination of the endpoint. For example, WaterOneFlow web services representing a HIS catalog will have the termination “hiscentral.asmx”, and the services representing a single HIS server have the termination “cuahsi_1_1.asmx”.

The following two endpoints for the HIS Central catalog and Calvin_HSS show the difference in the endpoint termination:

1. https://hiscentral.cuahsi.org/webservices/hiscentral.asmx (HIS Central)
2. http://hydroportal.cuahsi.org/CALVIN_HHS/cuahsi_1_1.asmx (Calvin_HSS)

As a result, the WDE refers to the WaterOneFlow web services that do not represent a catalog as servers. These servers representing WaterOneFlow web services, and their associated metadata are contained in the catalogs. In other words, the catalogs and servers are levels inside the WDE, being the catalog a level higher than the servers because multiple servers can be part of a catalog. Similarly, servers representing WaterOneFlow web services are composed of multiple observation sites. SOA systems can provide a WaterOneFlow service for individual station as well, but metadata from sites can be accessed through the server level WaterOneFlow web services methods. For example. methods such as GetSiteInfo can retrieve metadata from a specific site given a site code. In order to represent a specific site and its metadata, the WDE uses a level called "station". Therefore, the WDE structure manages three different levels: catalog, server, and station.

Different WaterOneFlow web services methods retrieve metadata for the different WDE levels. Figure 3-4 depicts the different WaterOneFlow web services methods that are handled at each WDE level, and the databases containing the downloaded metadata coming from the responses of these methods. The catalog level handles the GetWaterOneFlowServiceInfo WaterOneFlow web service method to retrieve information from a HIS catalog from a SOA system to later store the metadata in the catalog scheme from the WDE local database. Similarly, the server level handles two methods: GetSites and GetVariables to retrieve data from a SOA HIS server to later store the metadata in the server scheme from the WDE local database. On the other hand, the station level retrieves metadata from two methods: GetSiteInfo and GetValues, but do not store the retrieved metadata, but allows users to download the content to their personal computer.

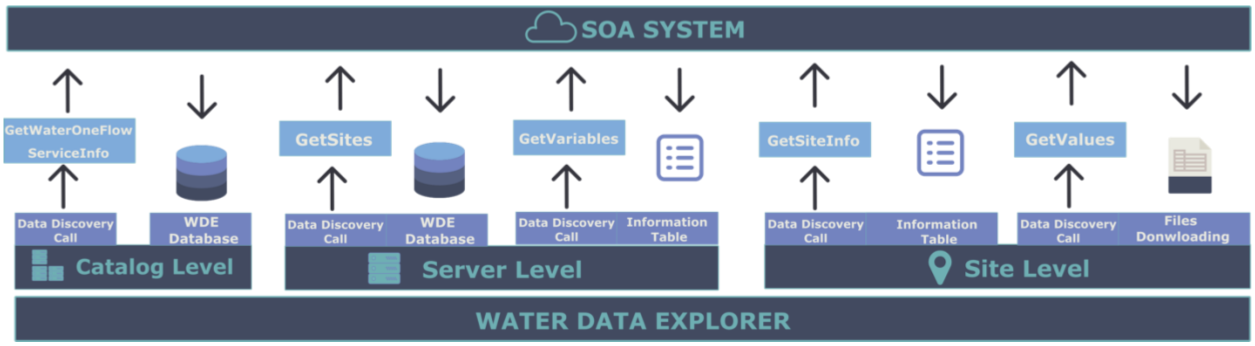

Figure 3-4: WDE Structure Levels

The WDE stores the responses received from the WaterOneFlow web services methods at the Catalog and server level because the WDE uses the metadata stored in order to avoid multiple requests to load the metadata to the user at the startup of the application. For example, geospatial visualization of the different sites in the WDE will require calling the GetSites for each WaterOneFlow service contained in each Catalog of the application. However, having the response saved in the WDE database reduces the time of loading and takes out the need to request data to the SOA systems every time the app starts. In contrast, the WDE does not store data at the station level for the WaterOneFlow web service methods that this level handles because storing the metadata from these methods will require making a GetSiteInfo for each site and a GetValues for each variable in a site making the retrieval of metadata slow. This means that every time the user request for metadata related to a specific site or time series observation values, the WDE will make a new request, and the user will have the ability to download the metadata to his local computer for further analysis.

3.4 WDE User Interface

The WDE interface is similar to other web-based applications developed by using Tethys Platform. Figure 3-5 depicts the User Interface (UI) of the WDE and indicates the location of the different WDE levels in the interface. The UI was designed to be a single web page that includes modal windows and buttons to provide tools for searching, obtaining, and managing data. The WDE includes a left panel displaying all the WDE catalogs and the different servers associated to each catalog. WDE also includes modals and tools for each of the three levels of information. Finally, a map displaying the stations from the different WDE catalogs and servers is also included in the right portion of the UI.

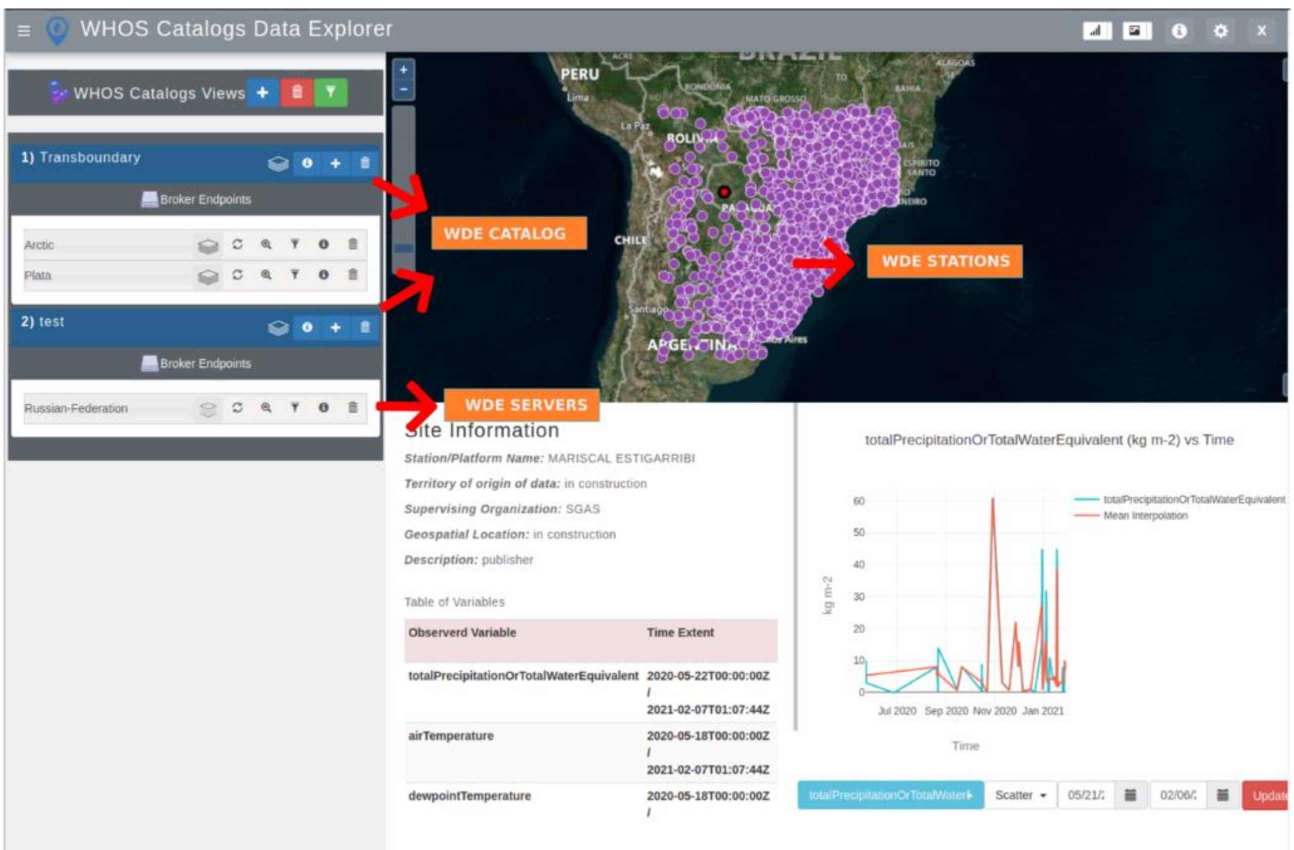

Figure 3-5: WDE User Interface

The UI components in the WDE have the objective to provide information at each of the three WDE levels: catalog, server, site. For example, the right- bottom portion of the WDE contains the site panel information which shows the users the following metadata: station name, supervising organization, geospatial location, territory of origin, and the measured variables at that station. Similarly, the right portion of the site information panel contains time series plots of variables belonging to the selected site. Section 5 will explore in more detail different UI when showing the results and UI for the two case studies: CUAHSI HIS and WHOS WMO. In summary, the objective of the WDE UI is to provide components and tools at each WDE component allowing the visualization, querying and downloading of data.

4 WDE DATA FUNCTIONALITIES

4.1.1 Data Discovery

WDE presents two different types of data discovery: 1) across all the WaterOneFlow web-services that have been registered to any HIS Catalog in a SOA system such as HIS Central in the CUAHSI HIS system; 2) within a single WaterOneFlow web-service that has not been registered to any HIS Catalog in a SOA system. The first type of data discovery, managed at the WDE catalog level, allows the user to do a complete discovery of the catalog metadata using any of the WaterOneFlow web services methods associated to a HIS catalog from a SOA system. The second type of data discovery, managed at the WDE server level, allows the user to make discovery calls to any WaterOneFlow web services in a SOA system. In the case of the CUAHSI HIS system, both types of discovery allow WDE to offer datasets that have been documented at HIS Central or databases that are stored in individual or regional HydroServers. Similarly, in the case of the WHOS, both types of discovery allow the WDE to offer the different customized datasets (views) from the WHOS broker.

Data discovery at the WDE catalog level involves two different discovery criteria: general discovery and a country-based discovery of the different servers representing WaterOneFlow web services contained in the WDE catalogs. General discovery at the catalog level is related to the discovery of WaterOneFlow web-services that have been registered to any

HIS Catalog while a country-based discovery involves data discovery of WaterOneFlow web services within selected countries.

Figure 3-3 in section 3.2 can be used to depict the process for the general discovery at the WDE catalog level. The WDE general discovery at the catalog level allows users to be able to import all the WaterOneFlow web services metadata or only from a selected number making different possible combinations of WaterOneFlow web services metadata in a WDE catalog. General discovery example can be the discovery of all the 98 web services registered at HIS Central Catalog or the partial discovery of selected WaterOneFlow web services. Similarly, the complete or partial discovery of WaterOneFlow web services in a WHOS view containing multiple web services is associated to the general discovery at the catalog level.

The WDE Catalog level also allows users to do a country-based search over all the WaterOneFlow web services in the catalogs. Figure 4-1 depicts the country-based search at the catalog level, which allows users to retrieve the servers, representing WaterOneFlow web services, associated with a country of interest. The country-based discovery is perform using latitude/longitude polygons to serve as representation for the selected countries, then each one of the sites associated to the selected servers are filtered using the latitude and longitude of the sites to see if the sites lie within the polygons of the countries. The country-based discovery is performed on the local WDE database over the WaterOneFlow web services metadata of the sites contrary to the general discovery use of WaterOneFlow web services methods on the SOA systems.

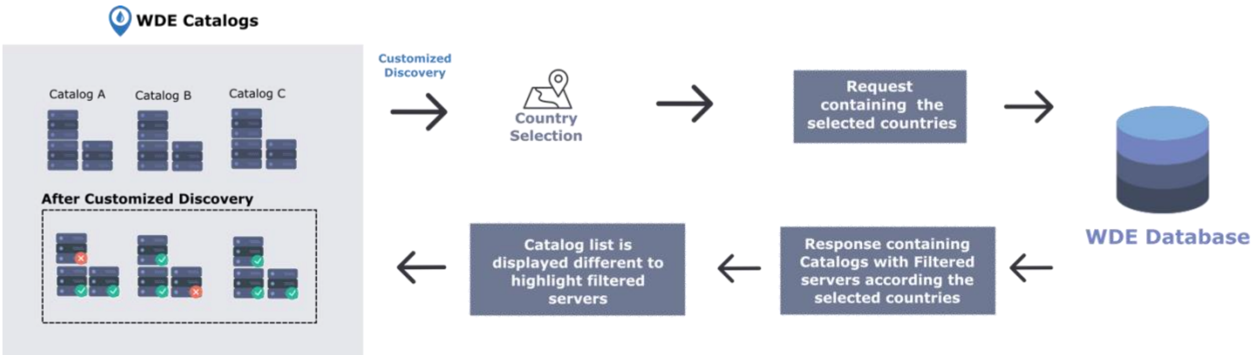

Figure 4-1: Catalog Level Country-Based Discovery

Similarly, to the retrieval of information at the WDE Catalog level, the WDE Server level also has two different types of data discovery methods: general and variable discovery for the WaterOneFlow web services represented as servers. The general discovery allows users to discover new WaterOneFlow web services in SOA systems that can be part or not from a HIS catalog, while the variable data discovery is related to the ability to discover the variables that are available at each WaterOneFlow web service represented by the different servers.

The general discovery at the server level is depicted by Figure 4-2 in which the user is able to discover the metadata associated to sites in the discovered WaterOneFlow web service using web services methods. The general discovery procedure at the server level is similar to the one performed at the Catalog level because it involves the creation of a server; request of information via WaterOneFlow web services methods; and storing of retrieved metadata in the WDE database. The general discovery procedure can be summarized in four steps. First, the WDE creates a new server at the server level. Second, the WaterOneFlow web service method GetSites is used to request the sites metadata. Third, the sites metadata are retrieved and stored in the WDE database. Fourth, the metadata is displayed in the newly created server.

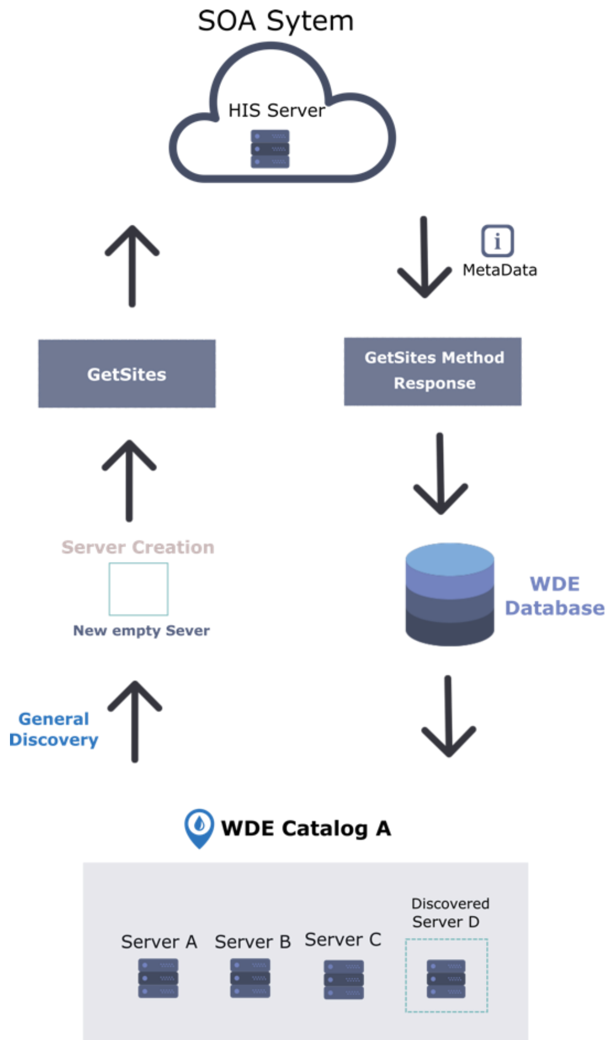

Figure 4-2: Server Level General Data Discovery

The server level general discovery allows users to import metadata of a selected WaterOneFlow web service to a server at the WDE server level allowing to expand the size of servers that are part of catalog as the last portion of Figure 4-2 depicts. In addition, the general discovery allows the site metadata from WaterOneFlow web services from different SOA systems to be added to a created server allowing WDE catalogs to be composed of multiple SOA system WaterOneFlow web services.

The WDE server level also allows a variable data discovery for the discovered WaterOneFlow web services that are represented at each server in the server level. However, in contrast to the general discovery at the server level, the retrieved metadata is not saved in the WDE database, and the data discovery is over already discovered WaterOneFlow web services. Figure 4-3 depicts the process of data discovery for the variables associated to a discovered WaterOneFlow web service as follows. First, the WaterOneFlow web service method GetVariables is used to request the metadata from the variables associated to the discovered WaterOneFlow web service Second, metadata is retrieved and showed into the user using a variable information table.

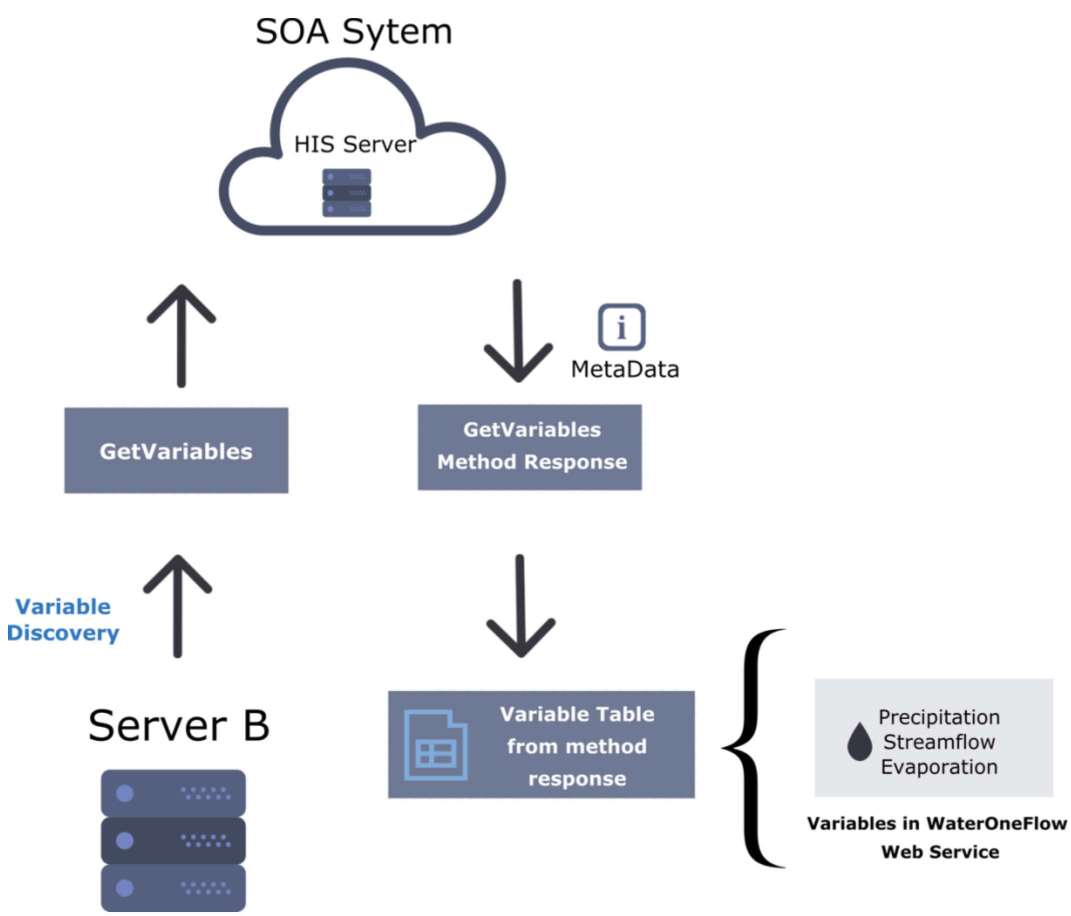

Figure 4-3: Server Level Variable Discovery

Similarly, to the retrieval of information at the WDE catalog and server level, the WDE Site level also has two different types of data discovery methods: general and time series discovery for each site at the WaterOneFlow web services represented as servers. The general discovery allows users to discover the metadata of a specific site such as: site name, supervising organization and measured variables, while the time series data discovery is related to the ability to discover the time series observations of the different variables at a selected site. However, the time series data discovery depends on the general discovery at the selected site in order to discover the variables that are measured by the site. Therefore, the WDE server level handles the general discovery of a site before discovering the time series observation values of a particular variable.

The general discovery at the site level is depicted by Figure 4-4 in which the user is able to discover the metadata containing the information of a specific site belonging to a discovered WaterOneFlow web service at the server level. The general discovery procedure at the site level is not similar to the one performed at the catalog and server level because it does not involve storing of retrieved metadata in the WDE database, but the ability for the user to download a file containing the site information metadata. The general discovery procedure can be summarized in three steps. First, a site is selected for general discovery and the WaterOneFlow web service method GetSiteInfo is used to request the site information metadata from the WaterOneFlow web service containing the site. Second, metadata from the site information is retrieved to the WDE. Third, the metadata is displayed in site information table at the site information panel for the following metadata: site name, supervising organization, location, measured variables, territory of origin.

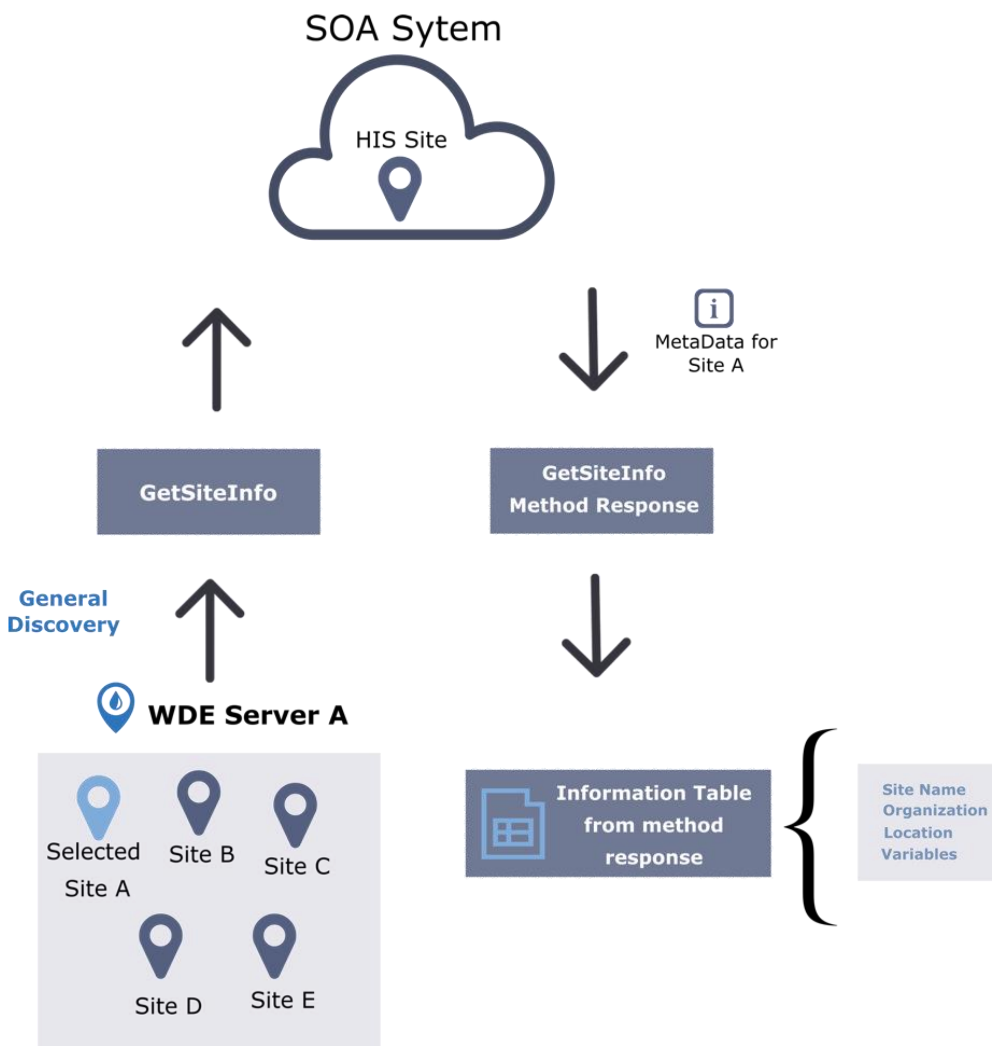

Figure 4-4: Site Level General Data Discovery

The WDE site level also allows a time series data discovery for any measured variable in the selected site in a discovered WaterOneFlow web service. However, in contrast to the general discovery at the site level, the retrieved metadata containing the time series observation values is not downloaded to the WDE database as explained in Section 3.3. Instead, the retrieved time series metadata can be downloaded by the user in two formats: CSV and XML.

Figure 4-5 depicts the process of time series data discovery for the time series values of a variable in a site associated to a discovered WaterOneFlow web service. First general discovery is performed in the selected site to find the measured variables available in the selected site. Second, a measured variable is selected, and the WaterOneFlow web service method GetValues is used to request the time series metadata. Third, the metadata containing the time series observation values of the measured variable is retrieved to the WDE. Finally, the metadata time series values are plotted at the site information panel, and metadata is available for downloading.

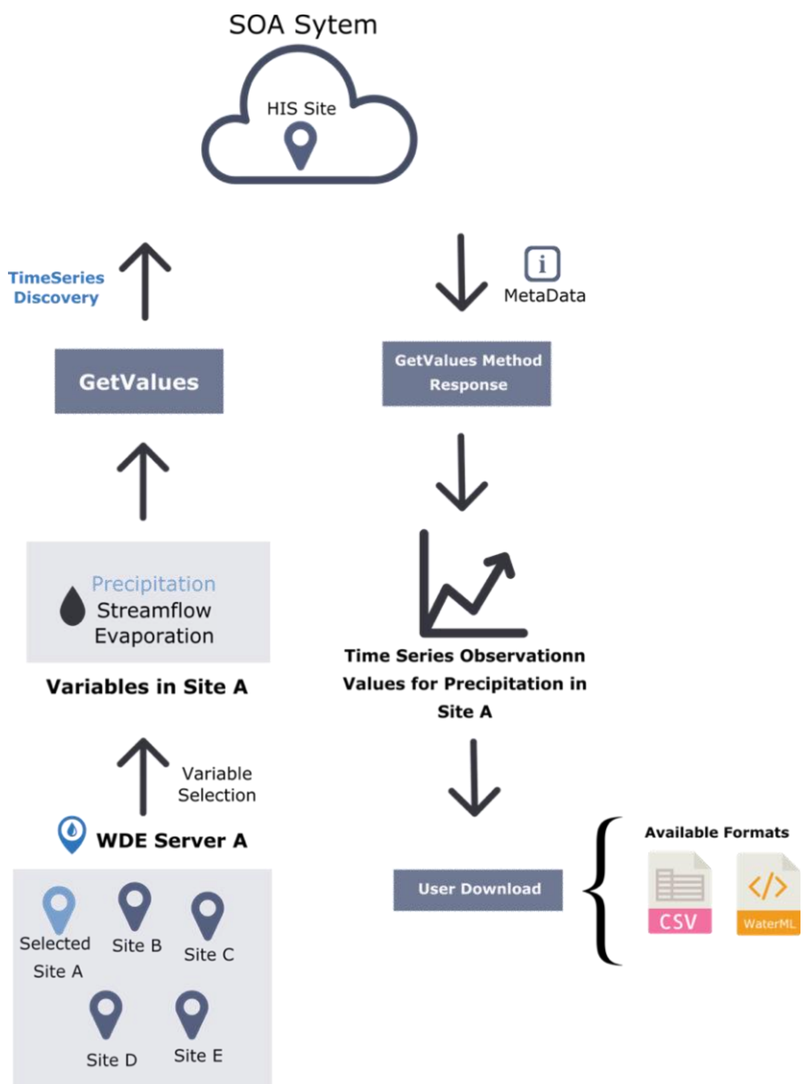

Figure 4-5: Site Level Time Series Data Discovery

4.1.2 Data Download

As explained in Sections 3.3 and 3.3 introducing the WDE levels and data discovery respectively, the WDE has a local database that is used to store the metadata response for the general data discovery at the WDE catalog and server level. In contrast, the variables data discovery at the server level, and the general and time series data discovery at the site level do not use the WDE database to store the metadata coming from the responses of the WaterOneFlow web services methods but offer the user the ability to download the data in different formats.

The general discovery at the WDE Catalog level stores the following metadata from the GetWaterOneFlowServicesInfo WaterOneFlow web service method response: i) name, ii) description, iii) URL of the WaterOneFlow web services. On the other hand, the metadata stored from the general discovery at the WDE server level from the GetSites WaterOneFlow web service method consist of the following information: i) sites names, ii) sites codes, iii) sites geospatial location, iv) sites network. WDE creates a local copy of the data mentioned above in a local PostgreSQL database. Data is stored in two schemes one for the different WDE catalogs and the other for servers as depicted in Figure 4-6. Having the Catalog and Server scheme allows the WDE to save and present data that have already been discovered when the user starts the application. As a result, loading the data into a database saves the discovered data of the user and make it available when the user opens the web application. For example, in the case of the CUAHSI HIS system, the Server scheme contains all the metadata response from the GetSites WaterOneFlow web service method associated to individual HydroServers and HIS central registered WaterOneFlow web services. For the WHOS system the Server scheme contains the metadata response from the same GetSites WaterOneFlow web service method associated to any

WaterOneFlow web service from the different data providers of the WHOS. Similarly, the catalogs scheme contains metadata from the GetWaterOneFlowServicesInfo WaterOneFlow web service method associated to HIS Central (CUAHSI HIS) or any WHOS customized view containing different web services.

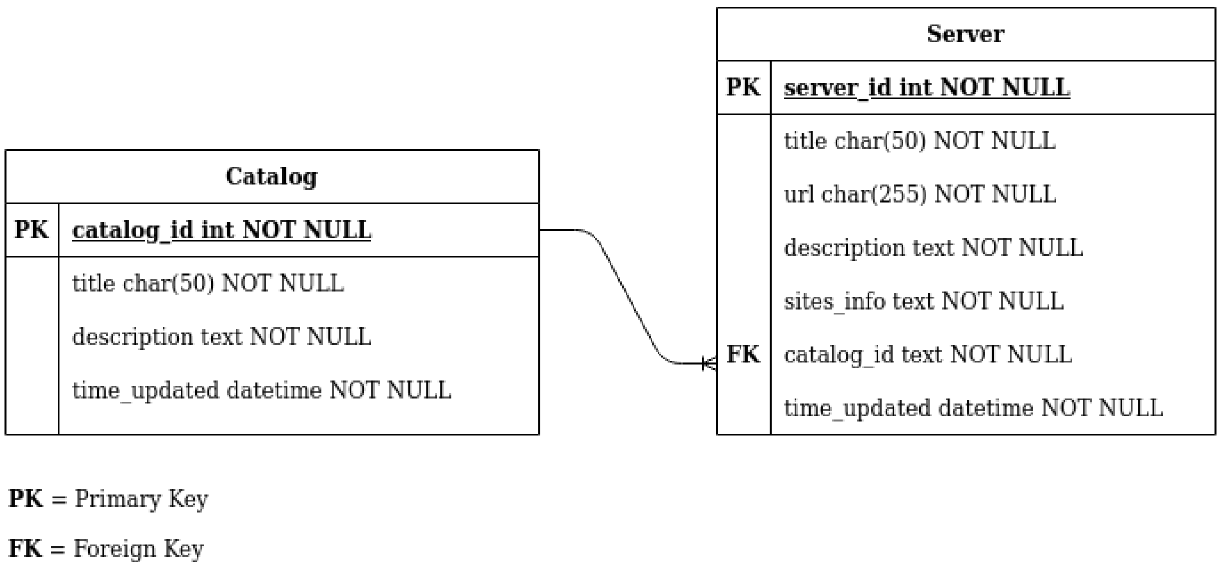

Figure 4-6: WDE PostgreSQL Database Structure

4.1.3 Data Visualization

WDE data visualization includes tools for visualizing geospatial, site information and time series metadata. The WDE User Interface (UI) includes a map on the right-hand side which displays the sites from the WaterOneFlow services. The geospatial visualization of the different sites associated to a WaterOneFlow web service at the server level starts after the metadata response from the GetSites WaterOneFlow web services is stored in the WDE database. Once the metadata is stored in the server scheme, WDE uses the Open Layers JavaScript library to create a vector (point) layer from the site's geospatial location attribute. After the sites are displayed on the map, the display can be turned on/off as needed.

Visualization of site information and time series metadata in the WDE occurs at the lower part of the UI in the site information panel when a site is selected. The site information and time series observation metadata visualization are displayed in the lower portion of the map and includes metadata and plots for data exploration. The site information metadata visualization occurs immediately after the metadata response from the GetSiteInfo WaterOneFlow web service method is retrieved. Users are able to visualize the following information: i) station, ii) territory of origin, iii) supervising organization, and iv) geospatial location (latitude and longitude). In addition, a table is presented with the following fields: i) observed variables, ii) units, iii) and temporal extent. Once the metadata is displayed, visualization of time series observation metadata is performed after choosing the variable of interest of the selected site. The WDE produces the time series metadata visualization after the metadata response from the GetValues WaterOneFlow web service method retrieved. The Plotly JavaScript library is used for the time series metadata visualization, and it also provides options between time series and box and whisker plots.

4.2 Experimental Case Study

The experimental case study focused on testing the different WaterOneFlow web services associated to the WDE catalog and server level. In addition, we customized the WDE for the different experimental case studies, such that each case had a customized version of the WDE in which it was being tested. In the case of the CUAHSI HIS system, data discovery at the catalog and server level were performed for the different WaterOneFlow web. Data discovery for the CUAHSI HIS Central catalog WaterOneFlow web service was tested at the catalog level while data discovery of the WaterOneFlow web services from individual HydroServers registered or not at the CUAHSI HIS Central was tested at the server level. Similarly, WaterOneFlow Web services from WHOS customized data view cataloging other WaterOneFlow services or not will be tested in the WDE catalog and server level respectively.

Designing the case study for the CUAHSI HIS Central catalog was focused on the performance and speed of the WDE to discover the large number of HydroServers registered at the CUAHSI HIS Central registry. The CUAHSI HIS Central catalog provides 98 different WaterOneFlow web services. As a result, the WDE data discovery for the CUAHSI HIS Central catalog needed to be able to permit discovery only of the HydroServer WaterOneFlow web services of interest. On the other hand, in the case study of the WHOS system the design was more focused on the ability of visualizing the site information and downloading the time series data of a variable of interest in different file formats: XML (WaterML 1.0, WaterML 2.0) and CSV.

5 RESULTS

5.1 Overview

The following section presents the WDE regional customization and results for the WHOS and CUAHSI HIS systems. We created two different WDE regional customizations to test the HIS Central catalog and the WHOS customized dataset's views. The first WDE regional customization was called “WHOS Catalogs”, and it was responsible for testing the WHOS customized dataset's views containing the transboundary and country regions in the WHOS system. The second WDE regional customization was called “HIS Central”, and it was responsible for the testing access to the HIS Central catalog WaterOneFlow web services. In this case study, the goal was to successfully customize the WDE regionally while being able to discover, download, visualize data from the different WaterOneFlow web services associated with the WDE Catalog and server levels. Next, visualizing geospatially the points and time series data from the different sites at the WDE station level belonging to WaterOneFlow web services is also an important part of the case study

5.2 Regional Customization

One of the advantages of the WDE is the ability to have regional customization. Regionally customization of the WDE is made in the custom settings of the newly installed WDE application as depicted in Figure 5-1. The custom settings of the WDE allows users to change the display title in the upper-left corner and to add a Web Mapping Service (WMS) layer to serve as a boundary in the application map. Figure 5-1 displays the different options for the customization of the app title and WMS boundary as follows:

1. View Names: title of displayed for the application
2. Boundary GeoServer Endpoint: endpoint of the GeoServer WMS layer.
3. Boundary Workspace Name: name of the WMS layer workspace
4. Boundary Layer Name: name of the WMS layer
5. Boundary Movement: Boolean to allow users to move outside the established boundary
6. Boundary Color: boundary layer color
7. Boundary Width: boundary layer width

CUSTOM SETTINGS					
NAME	DESCRIPTION	TYPE	VALUE	REQUIRED	ERRORS
Views Names	Name of the region holding the views (e.g. La Plata Basin)	String			
Boundary Geoserver Endpoint	Geoserver endpoint for the hydroshare resource containning the layer (e.g:"https://geoserver.hydroshare.org/geoserver/layerID")	String			
Boundary Workspace Name	workspace and layer name (e.g workspace:layername)	String			
Boundary Layer Name	layer name (e.g workspace:layername)	String			
Boundary Movement	Block or Allow movement outside the map boundary layer (True/False)	Boolean			
Boundary Color	The color style for the boundary (e.g #ffcc33)	String			
Boundary Width	Width of the boundary. A number from 1 to 10	String			

Figure 5-1: Custom Settings Panel

Figure 5-2 depicts an example of regional customization using the WMS boundary layer for La Plata basin area in which the title of the app appears as. "La Plata Data Explorer", and the yellow WMS layer represents the boundary of La Plata basin. The two different customized

WDE versions used for the test cases are depicted in Figure 5-3 and Figure 5-4. Both Figures show each one of the customized views: a) WHOS Catalogs, and b) HIS Central respectively. The two different customized views do not have the WMS layer for the boundary as Figure 5-2 depicts because the geographic extent of them can cover multiple countries in different continents. In addition, it is important to notice that the only difference within the two customized views for the CUAHSI HIS and WHOS WMO is the appearance. Both applications have the same data discovery, download, and visualization functionalities. This means that users of the CUAHSI HIS customized version of the WDE can perform a general discovery of WaterOneFlow web services that do not belong to the HIS Central. Therefore, the regional customization of both case studies pretends to demonstrate that users can have custom version of the WDE for their geographic area of intervention.

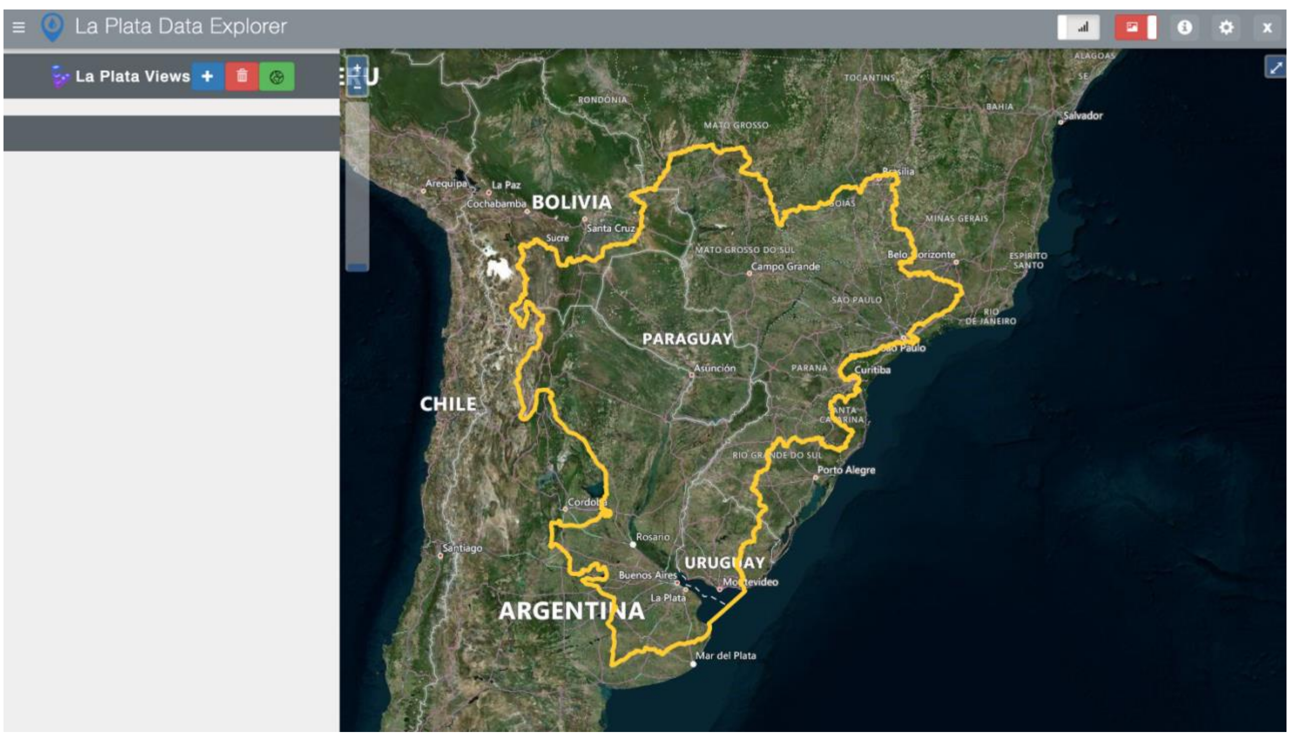

Figure 5-2: Regional Customization Example for La Plata Basin

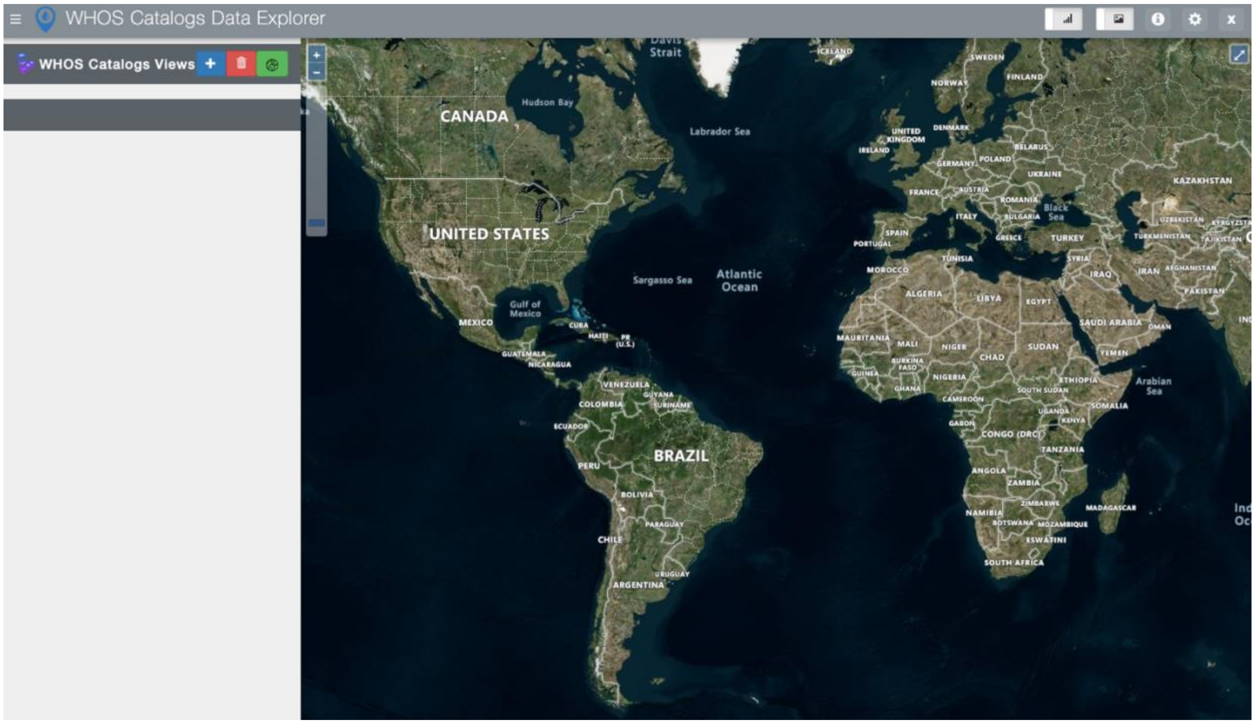

Figure 5-3: WHOS regional customized version

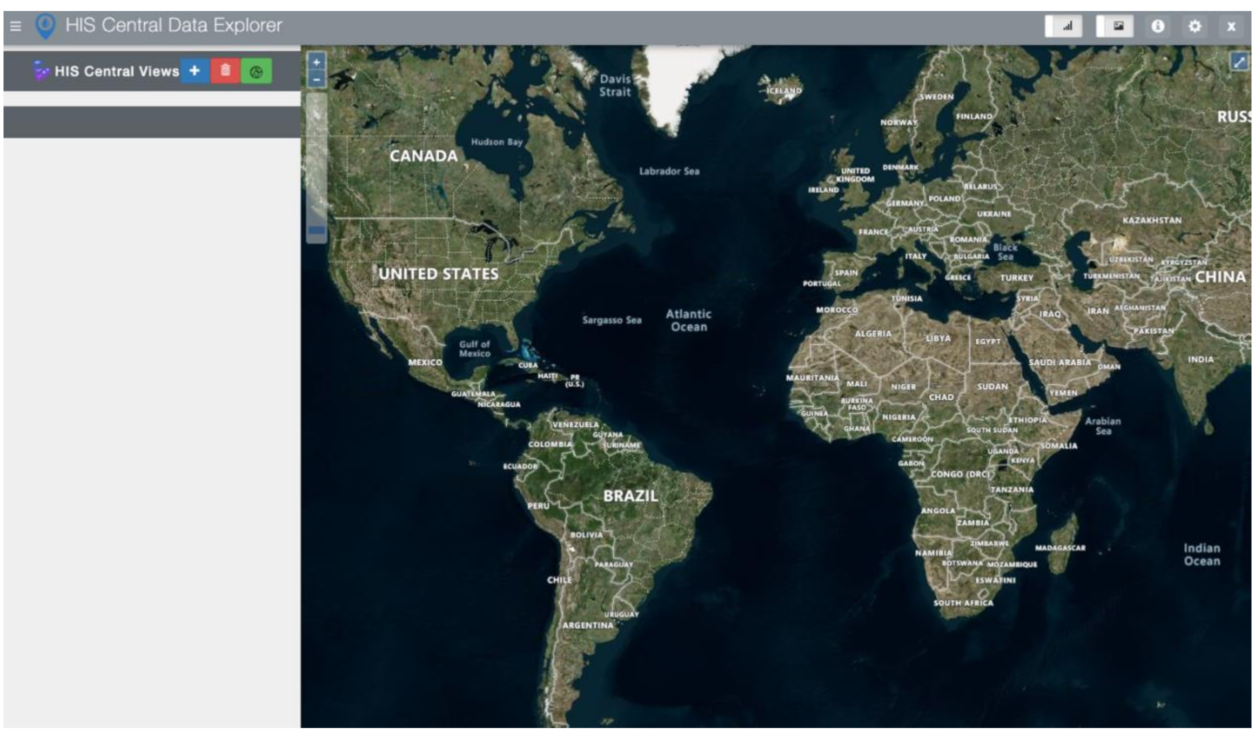

Figure 5-4: HIS Central Regional Customized Version

5.3 Data Discovery

The WDE homepage is presented with a map displaying the Bing terrain base map alone in order to avoid the confusion of recognizing different sites when the user starts the application. However, control is given to turn off and on the display of the different sites associated to the WDE Catalog and server levels. General data discovery at the server level was performed on the WaterOneFlow web services for the HIS Central catalog and the countries and transboundary customized WHOS views. General discovery was performed creating a WDE catalog, making the GetWaterOneFlowWebServicesInfo call, storing the metadata response in the WDE database, and displaying it in the WDE catalog list. The input required for the general discovery of the HIS catalogs is the following: i) Catalog Name, ii) Catalog Description, iii) URL (optional) as depicted in Figure 5-5.

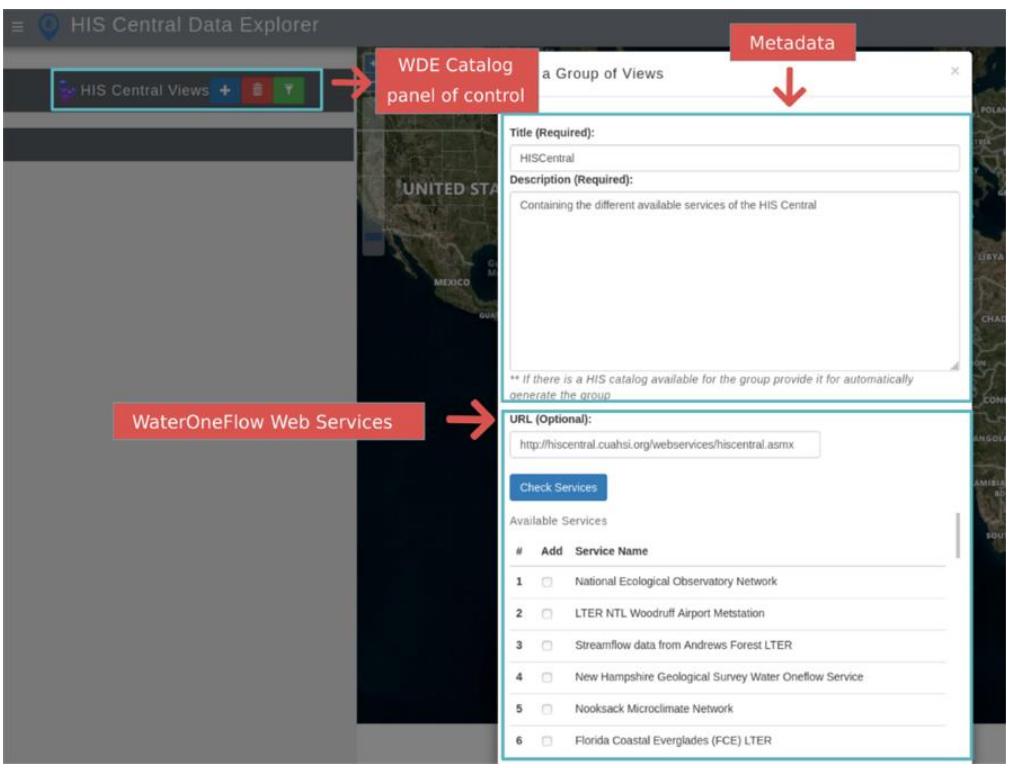

Figure 5-5: WDE Catalog Data Discovery User Interface

In the “WHOS Catalogs Data Explorer” WDE customized version, the transboundary and countries WaterOneFlow web services were discovered with all their available web services respectively as Figure 5-6 depicts. On the other hand, in the “HIS Central Data Explorer” customized version not all the WaterOneFlow web services were discovered from the HIS Central catalog because some WaterOneFlow web services contains big amounts of metadata that slow down the performance of the data discovery such as NWIS daily values, NWIS Unit values WaterOneFlow web services from the U.S. Geological Survey (USGS). Therefore, only six WaterOneFlow web services were chosen to be discovered without a specific selection criterion as Figure 5-7 depicts.

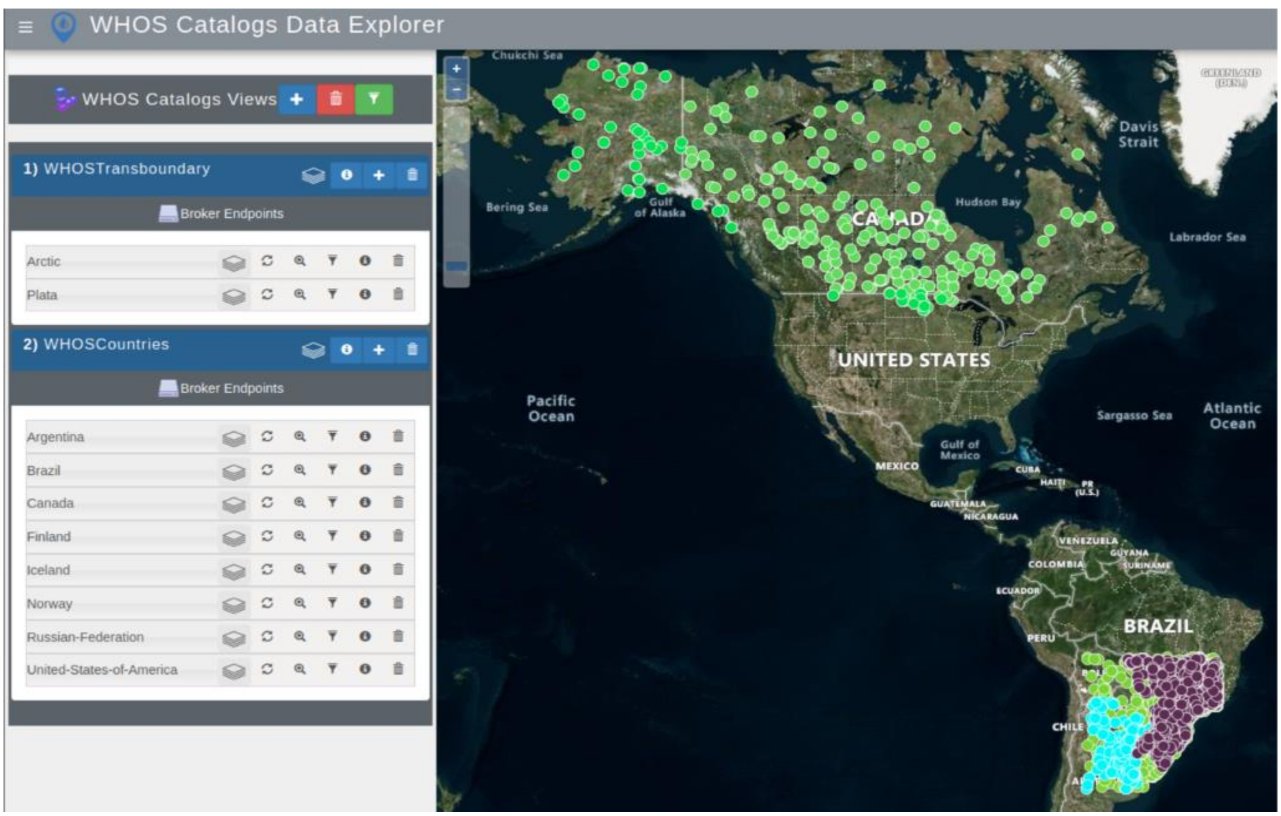

Figure 5-6: WDE Customized Versions for the WHOS System

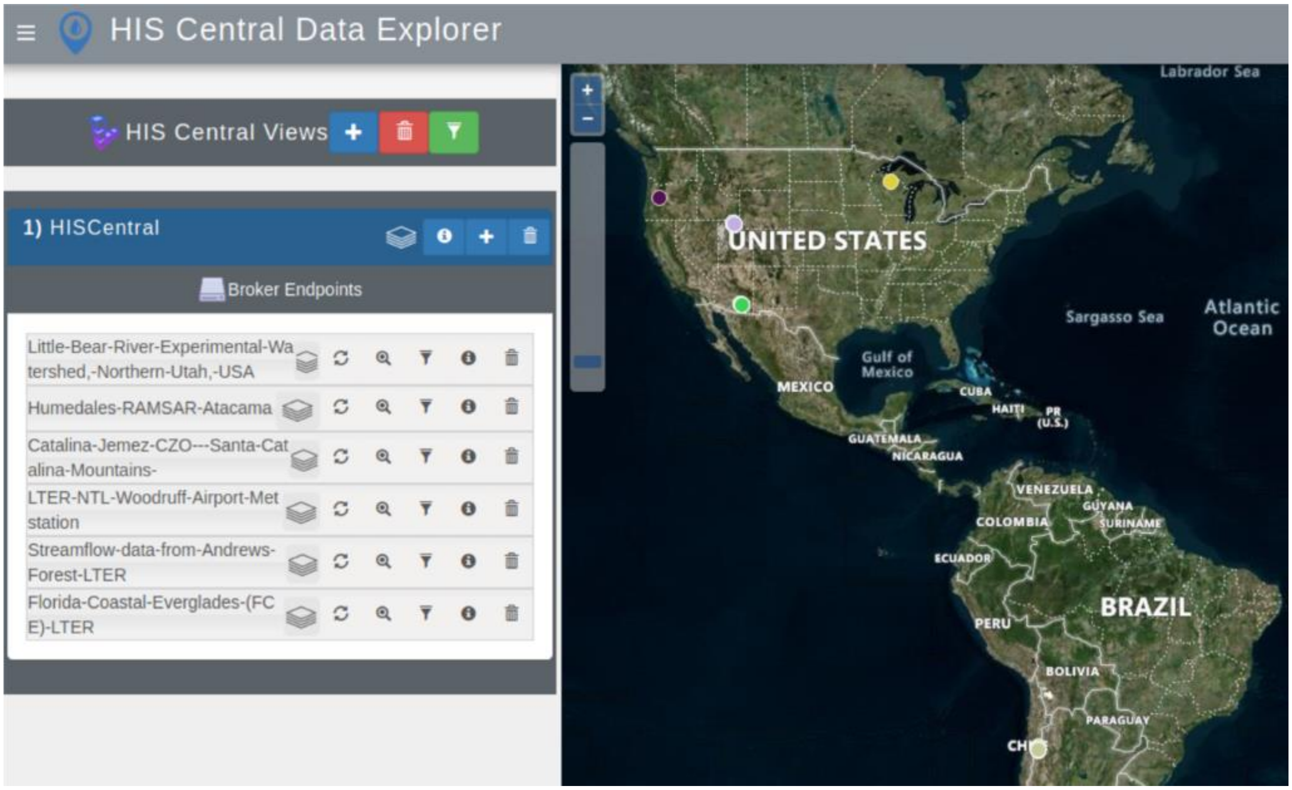

Figure 5-7: WDE Customized Versions for the CUAHSI Central

Country-based Discovery was also tested in the two case studies. The panel of control containing the tool for the general discovery also contains the button for the country-based discovery modal menu as Figure 5-8 depicts. A search bar is available to facilitate the selection of a country if the WDE would contain services involving multiple countries. The menu shows all the different countries in which the discovered WaterOneFlow web services are present. In the case of the WHOS case study, the countries available were Canada, Iceland, Brazil, Russia, Argentina, Bolivia, Paraguay, Finland, Uruguay, Norway, and United States of America. In the case of the CUAHSI HIS, the only two available countries were: Chile and United States of America.

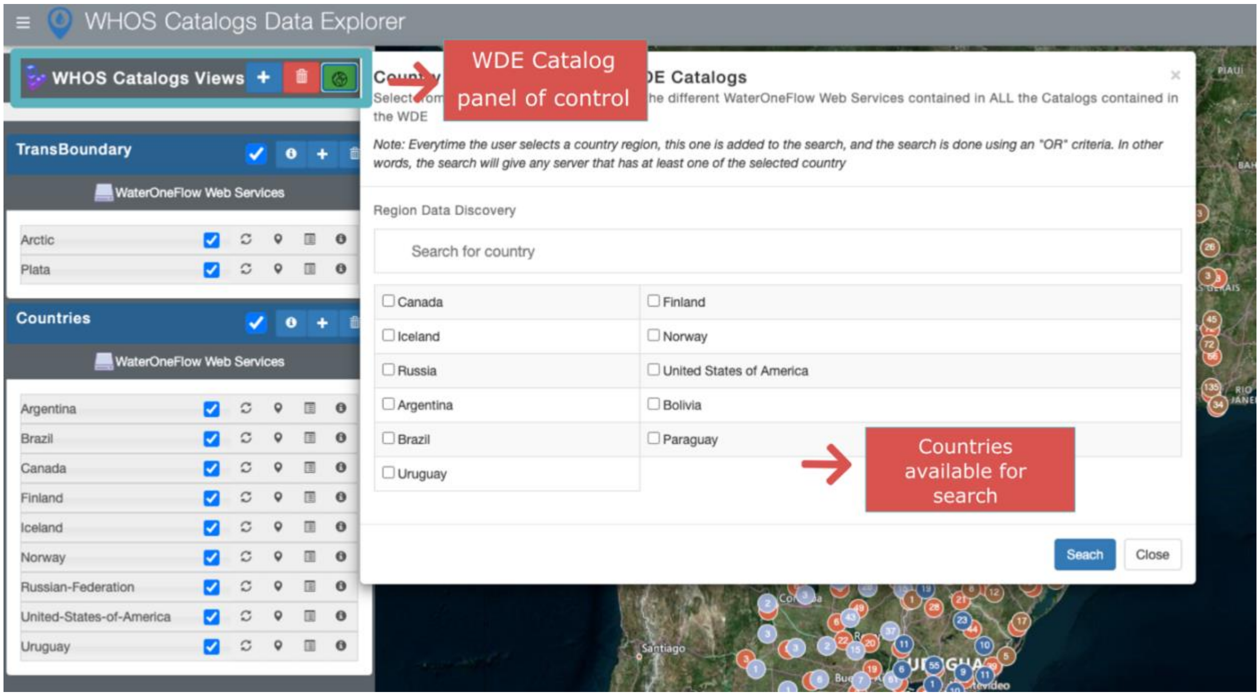

Figure 5-8: WDE Country-based Discovery Menu

Once the user selects the Countries of interest, then the WDE performs the country-based discovery as depicted in Figure 5-9 where Brazil was selected for the WHOS case study. Figure 5-9 also shows that the country-based discovery for the WaterOneFlow web services that lies within Brazil reports "La Plata" WaterOneFlow web services as the only WDE server that its sites are within the geographic extension of Brazil. Figure 5-9 also depicts that the country based search will only shows the different sites that lies within Brazil even though the WaterOneFlow web service for La Plata cover multiple countries. The list of servers in the Transboundary Catalog also shows the Plata WaterOneFlow web service outlined in red to distinguish it from the WaterOneFlow web services that do not contain stations that lies within Brazil.

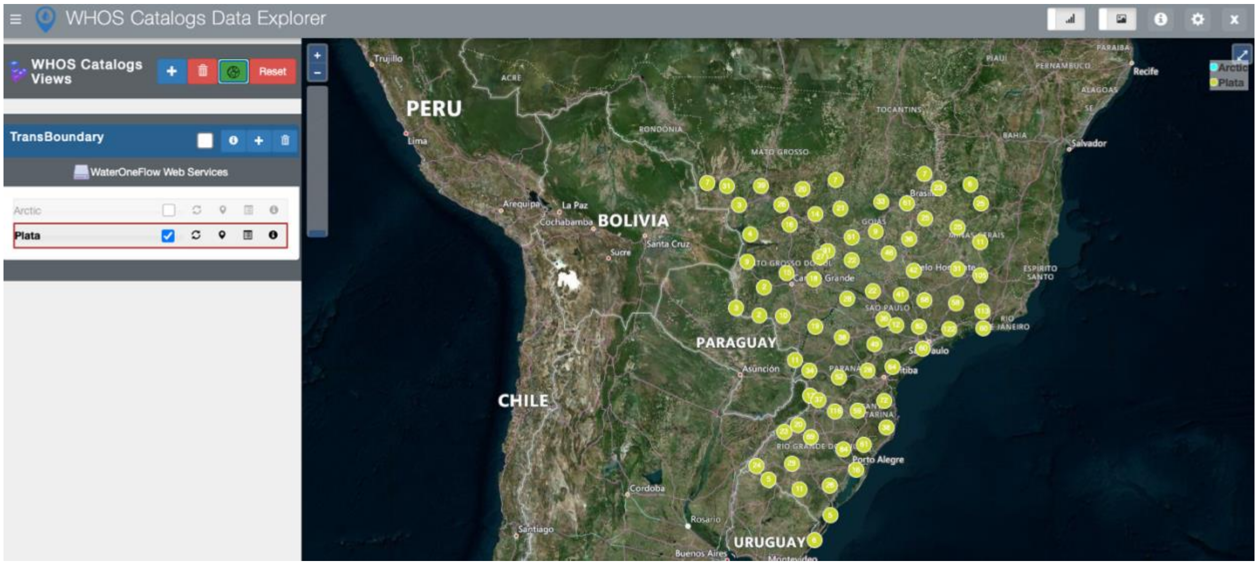

Figure 5-9: Country search for WaterOneFlow Web Services in Brazil

In addition, other WaterOneFlow web services can be added at the WDE server level for the two different case studies as explained in Section 4.1.1. Figure 5-10 depicts the modal menu that allows the user to perform a general discovery at the server level. Input data required for the general discovery at the server level for the CUAHSI HIS and WHOS was the same: title, URL, and description of the WaterOneFlow web service. In order to test the general discovery at the server level, input data for the Humedales Ramsar Atacama WaterOneFlow web service registered at CUAHSI HIS Central was entered into the modal depicted in Figure 5-10 to perform a general discovery at the server level in the WHOS case study for the WDE transboundary Catalog. Results of the server general discovery are depicted at Figure 5-11 for the WHOS case study. Five different sites were discovered from the Humedales Ramsar Atacama WaterOneFlow web service, and the WDE map zoom level was changed to focus in the sites when the general discovery at the server level was finished.

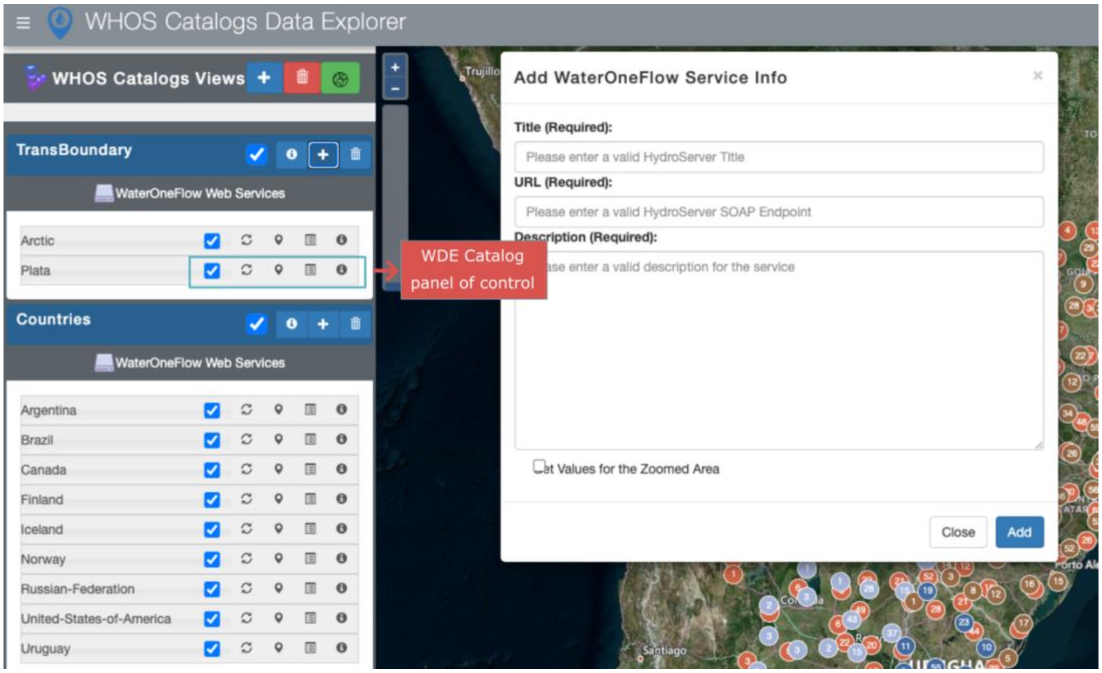

Figure 5-10: WDE Data Discovery User Interface at Server Level

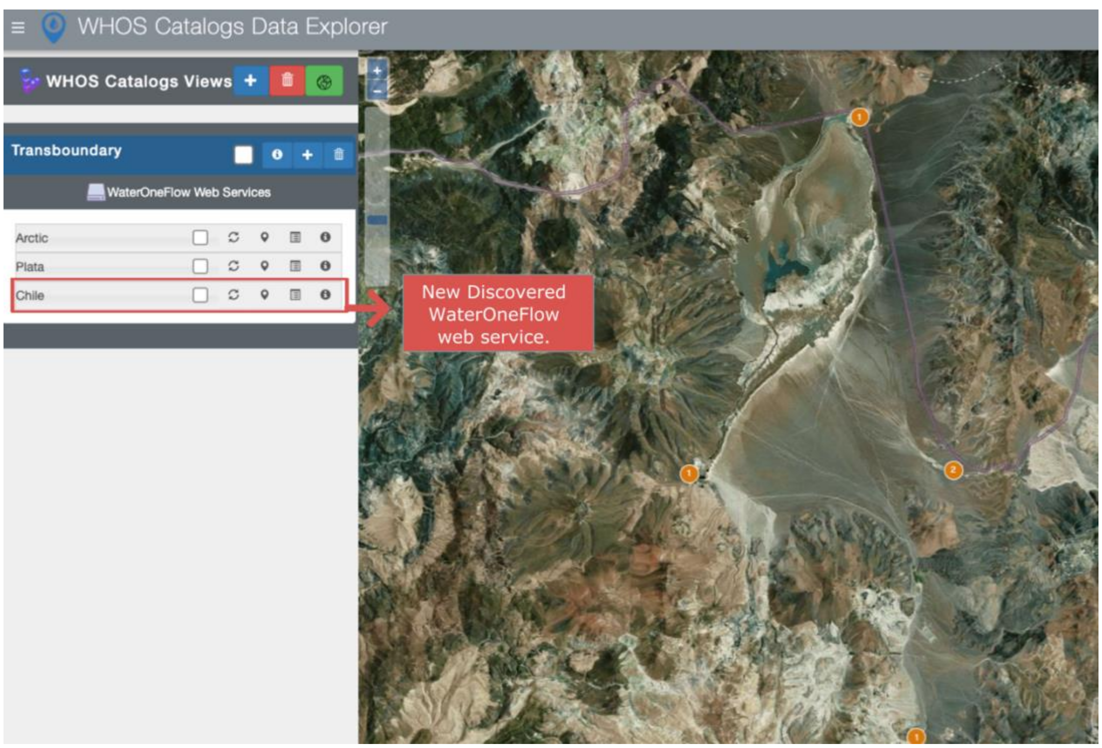

Figure 5-11: Server Level General Discovery for WHOS Catalogs

Similarly, variable discovery was conducted at the WDE server level by clicking the server panel of control for each WaterOneFlow web service discovered in the server level. Variable discovery was tested in the WHOS case study, namely in the La Plata server in the Transboundary catalog, as depicted in Figure 5-12. The metadata from the different variables belonging to the La Plata WaterOneFlow web service is displayed in a table containing: variable name, measurements units, and WHOS variable code. At the WDE site level, discovery of data was conducted by clicking in an individual site. Once a site is clicked the WDE Site Information panel appears with the site metadata. Results for data discovery at the site level are depicted in the next section related to visualization of hydrologic data because of the visualization of the timeseries and site information data discovery.

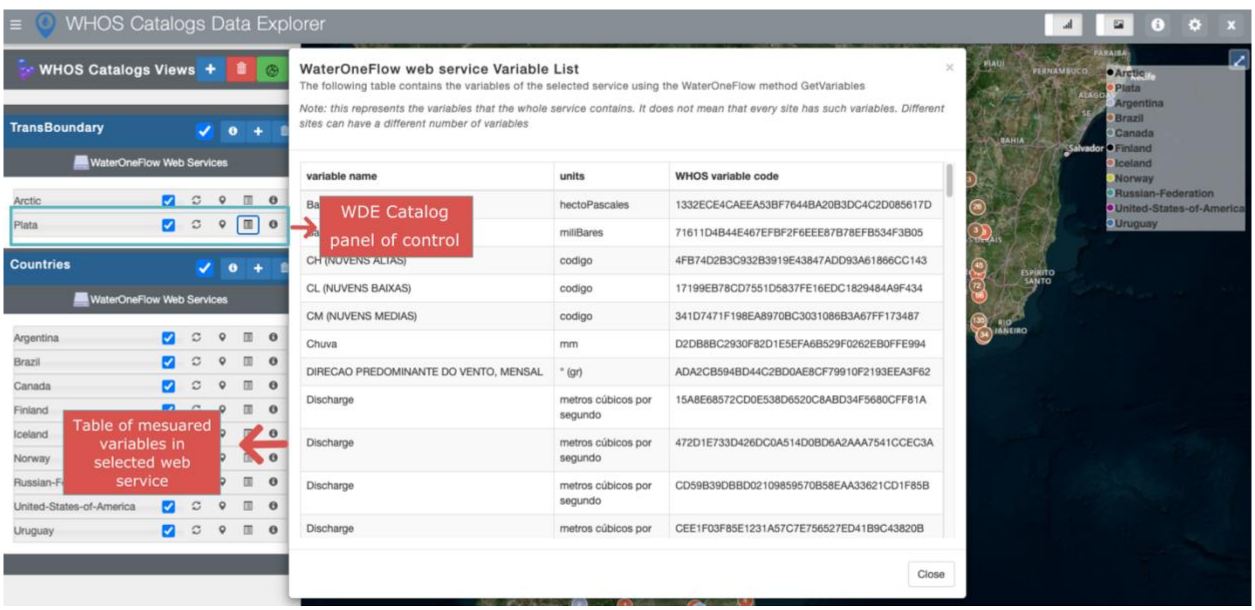

Figure 5-12: WDE Variable Discovery at Server Level

5.4 Visualizing Hydrologic Data

Visualizing Hydrologic data in the two different case studies was conducted in two different places: in the Site Information panel for time series visualization, and in the map and Server information panel for geospatial visualization. Time series visualization for the observation data happened in the site information panel after the user selected the variable of interest. The period for the time series can be customized by changing the temporal extent of the variable of interest at the selected site by using the date-time menus in the plot. Figure 5-13 depicts time series visualization for the air temperature variable in the “WHOS Catalogs Data Explorer" case study for the La Plata WaterOneFlow web service (http://gs-service-production.geodab.eu/gs-service/services/essi/view/whos-plata/cuahsi_1_1.asmx?WSDL) associated to the site BAHIA NEGRA - SGAS. Similarly, Figure 5-14 depicts time series visualization for the reservoir storage variable in the “HIS Central Data Explorer" case study for the CALVIN_HHS WaterOneFlow web service (http://hydroportal.cuahsi.org/CALVIN_HHS/cuahsi_1_1.asmx?WSDL) associated to the site SR-BBL.

Finally, the WDE geospatial visualization at the WDE Catalog and Server level can be tested in the home page map (when all the different sites are displayed) and in the Server information modal window respectively. The server information modal can be accessed by using the information button at the server panel of controls, and it shows the different sites associated to the HIS server. Figure 5-15 depicts the geospatial visualization in the “WHOS Catalogs Data Explorer" case study for the La Plata WaterOneFlow web service, and it also introduces the Pywaterml official documentation and python notebook templates using the package for the selected WaterOneFlow web service.

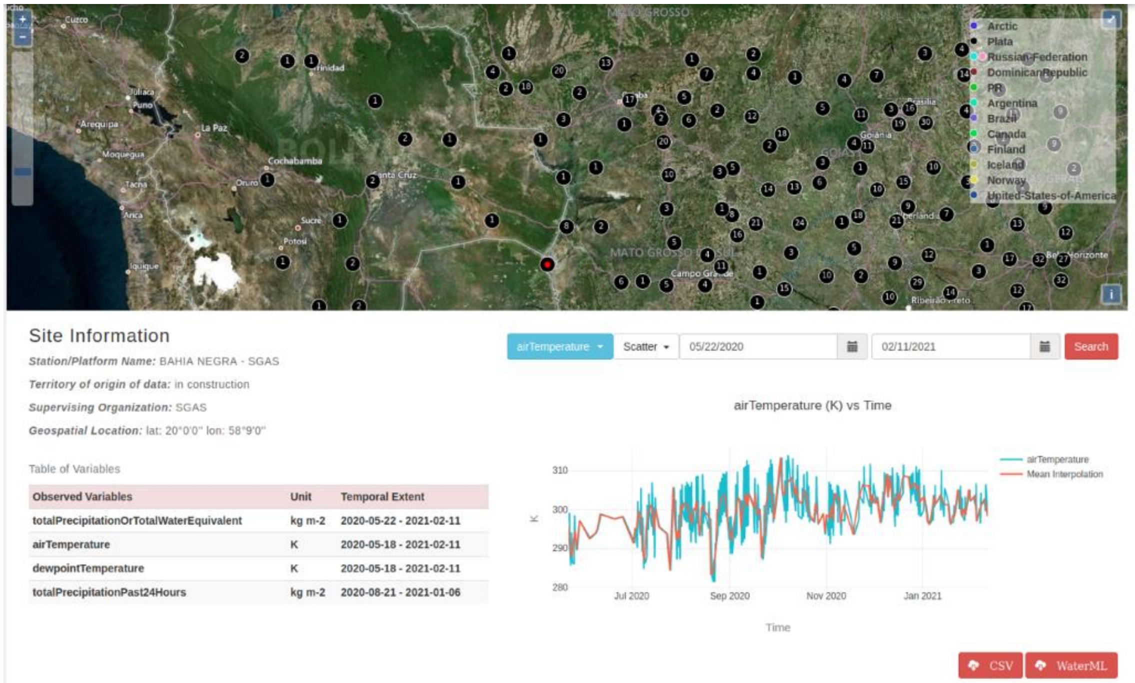

Figure 5-13: Visualization for the Air Temperature

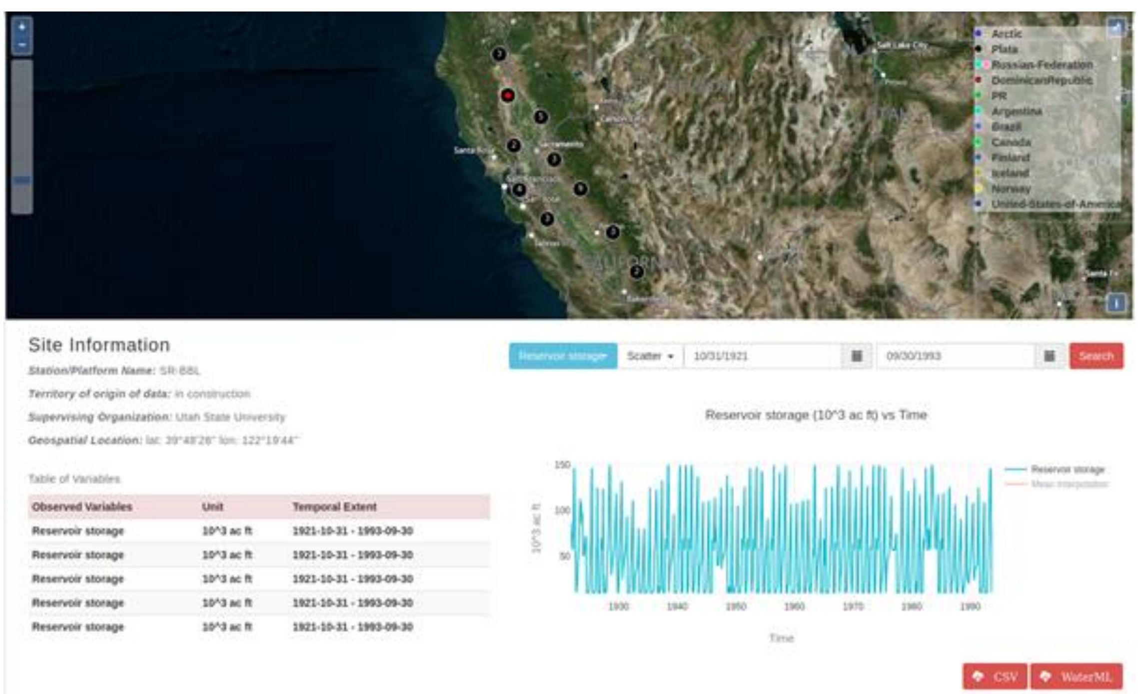

Figure 5-14: Visualization for the Reservoir Storage Variable

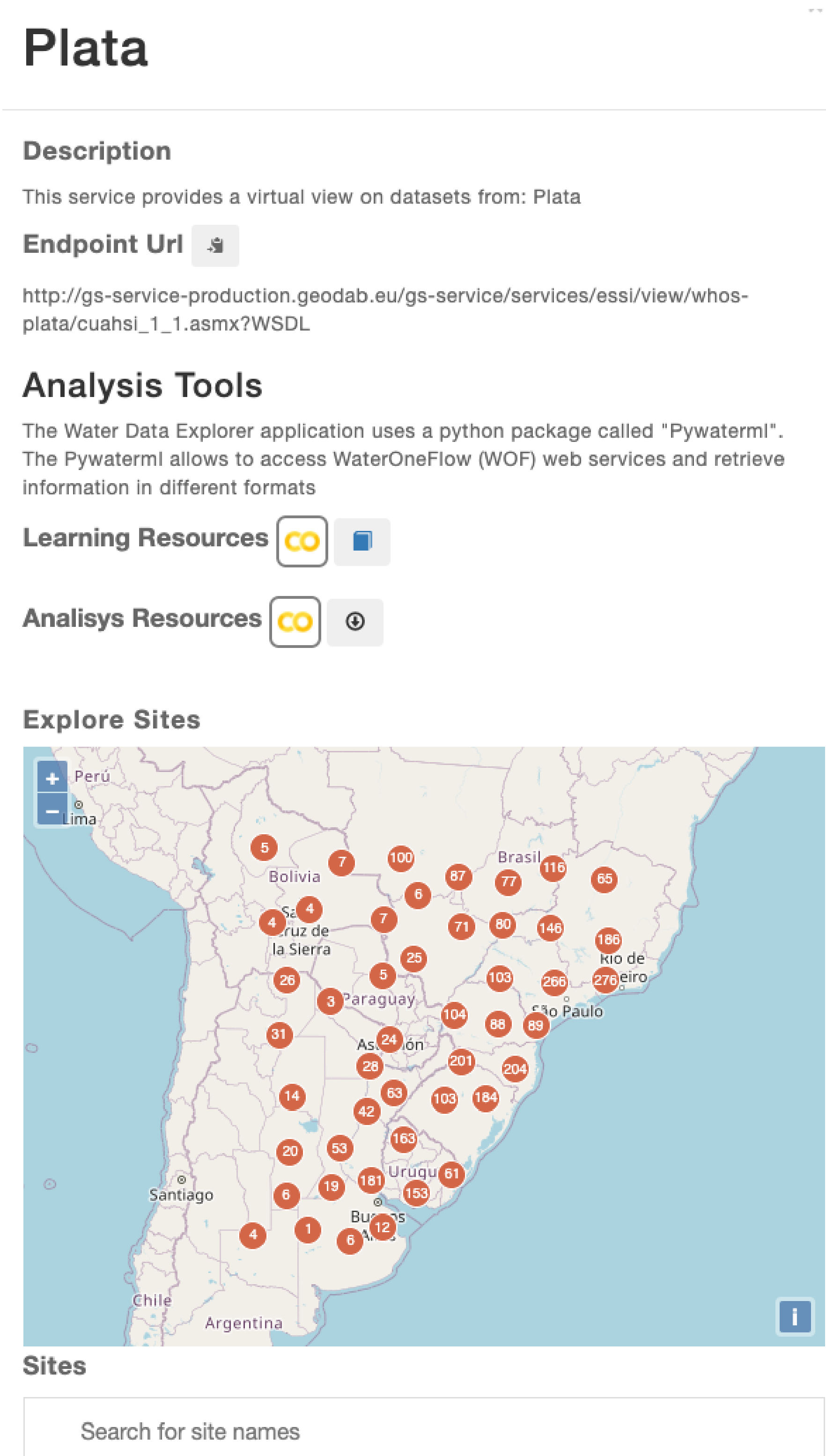

Figure 5-15: WDE Geospatial Visualization Modal for La Plata Web Services

5.5 Downloading Data

Downloading the data in the two different case studies of the WDE were conducted at two different moments. The first moment of data downloading was tested when the user discover data from the HIS Central (CUAHSI HIS) and from the countries and transboundary dataset customized views (WHOS) at the WDE catalog level, when metadata for the WaterOneFlow services was downloaded to PostgreSQL. The second moment that was tested happens when the user download time series observational data for a variable in a site. Figure 5-16 shows the download of a CSV file associated to La Plata Basin from the WHOS transboundary region. The total precipitation variable of the site Mariscal Estigarribia is downloaded in a CSV format for testing, but also WaterML format is available.

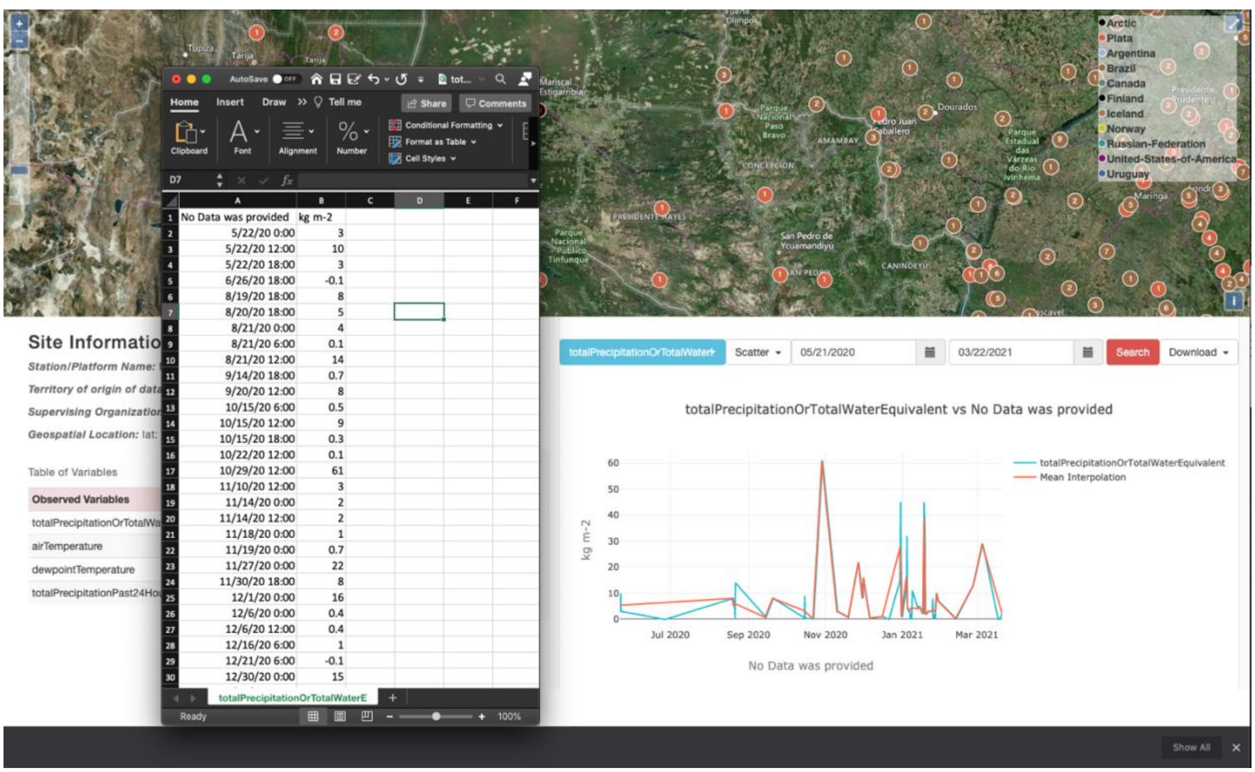

Figure 5-16: Data Downloading in CSV Format in the WDE Site Panel

6 DISCUSSION AND RESULTS

The case studies illustrate the ease in customizing of the WDE, which allows creation and customization of different versions of the same base application. Further customization of the WDE such as change of the logo, theme colors, and application name is possible by customizing the source code of the application which is available from its open-source repository. The results for the two different case studies for the CUAHSI HIS and WHOS systems also tested the three main data functionalities: discovery, download and visualization of the WDE.

Data discovery in the CUAHSI HIS system included the 98 different WaterOneFlow web services that the HIS Central contains while the transboundary and country regions customized data views from the WHOS contained two and nine WaterOneFlow services respectively. Discovering 98 different web services consumed more time and demonstrated that the responses from the WaterOneFlow methods were consistent in the structure of the responses the majority of the time. Similarly, the two customized data views from WHOS also show consistency in their responses. However, consistency between the response of the WaterOneFlow web services methods between the different SOA systems is still a limiting factor in the WDE as a client component. The pywaterml package tries to use the common metadata that can be found among WaterOneFlow web services responses as the baseline for the WDE data discovery. However, depending on the WaterOneFlow web service and SOA system the WDE might need to be customized to suit different response structures. Data visualization and downloading among different WaterOneFlow services also depends on the consistency of the structure of the responses and might need further customization. As a result, the WDE serves primarily as a client component for a single SOA system, and it can be customized to serves as a client component for more SOA systems.

Future work related to the data download from the WaterOneFlow web services methods involves updating automatically the local database every time the WaterOneFlow web service data is updated on the data provider side. Currently, the WDE has an option to manually update the data through the UI, but having the data of the local database synchronized automatically with the database of the provider will ensure the WDE to have observational data up to date without having the user to check for updates manually.

The WDE, as a web application developed in the Tethys Platform framework, provides users the ability to customize the application for different regions of WaterOneFlow web services, which allows different versions of the WDE regionally. The WDE also provides three main functionalities as a client component for different SOA systems: data discovery, data visualization, and data downloading. The WDE's design consisted of three different levels of information, including catalog, server, and stations that allow the application to organize the data response of the WaterOneFlow web services methods. The design of the WDE also included the creation of the Pywaterml package to connect to the different WaterOneFlow web services from the SOA systems, and to execute their different methods to discover, visualize, and download data. As such the methods can be extended to other applications with general or specific needs for accessing observational time series data.

REFERENCES

Ames, D.P., Horsburgh, J.S., Cao, Y., Kadlec, J., Whiteaker, T., Valentine, D., 2012. HydroDesktop: Web services-based software for hydrologic data discovery, download, visualization, and analysis. Environmental Modelling & Software 37, 146–156. https://doi.org/10.1016/j.envsoft.2012.03.013

Boldrini, E., Mazzetti, P., Nativi, S., Santoro, M., Papeschi, F., Roncella, R., Olivieri, M., Bordini, F., Pecora, S., 2020a. WMO Hydrological Observing System (WHOS) broker: implementation progress and outcomes (No. EGU2020-14755). Presented at the EGU2020, Copernicus Meetings. https://doi.org/10.5194/egusphere-egu2020-14755

Boldrini, E., Mazzetti, P., Nativi, S., Santoro, M., Papeschi, F., Roncella, R., Olivieri, M., Bordini, F., Pecora, S., 2020b. WMO Hydrological Observing System (WHOS) broker: implementation progress and outcomes, in: EGU General Assembly Conference Abstracts. p. 14755.

Boldrini, E., Mazzetti, P., Nativi, S., Santoro, M., Papeschi, F., Roncella, R., Olivieri, M., Bordini, F., Pecora, S., 2019. WMO Hydrological Observing System (WHOS): a collaborative implementation approach 21, 13620.

Challenges to Hydrological Observations [WWW Document], 2015. . World Meteorological Organization. URL https://public.wmo.int/en/bulletin/challenges-hydrological-observations (accessed 3.21.20).

Conner, L.G., Ames, D.P., Gill, R.A., 2013. HydroServer Lite as an open source solution for archiving and sharing environmental data for independent university labs. Ecological Informatics 18, 171–177. https://doi.org/10.1016/j.ecoinf.2013.08.006

Crawley, S., Ames, D., Li, Z., Tarboton, D., 2017. HydroShare GIS: Visualizing Spatial Data in the Cloud. Open Water Journal 4.

Gregersen, J.B., Gijsbers, P.J.A., Westen, S.J.P., 2007. OpenMI: Open modelling interface. Journal of hydroinformatics 9, 175–191.

Horsburgh, J., Tarboton, D., Maidment, D., Zaslavsky, I., 2008a. A Relational Model for Environmental and Water Resources Data. Water Resour. Res 44. https://doi.org/10.1029/2007WR006392

Horsburgh, J., Tarboton, D., Maidment, D., Zaslavsky, I., 2008b. A Relational Model for Environmental and Water Resources Data. Water Resour. Res 44. https://doi.org/10.1029/2007WR006392

Horsburgh, J., Tarboton, D., Schreuders, K., Maidment, D., Zaslavsky, I., Valentine, D., 2010. HydroServer: A Platform for Publishing Space-Time Hydrologic Datasets.

IOSS, 2010. Data Management Planning and Coordination [WWW Document]. The U.S. Integrated Ocean Observing System (IOOS). URL https://ioos.noaa.gov/data/contribute-data/data-management-planning-coordination/ (accessed 3.9.21).

Lehnert, K., Walker, J., Carlson, R., Hofmann, A., Sarbas, B., 2004. Building the EarthChem System for Advanced Data Management in Igneous Geochemistry. AGU Fall Meeting Abstracts.

Lehnert, K.A., Carbotte, S.M., Ryan, W.B.F., Ferrini, V., Block, K., Arko, R.A., Chan, C., 2011. IEDA: Integrated Earth Data Applications to support access, attribution, analysis, and preservation of observational data from the ocean, earth, and polar sciences, in: Geophysical Research Abstracts.

Lehnert, K.A., Walker, D., Block, K.A., Ash, J.M., Chan, C., 2009. EarthChem: Next Developments to Meet New Demands (Invited). AGU Fall Meeting Abstracts 12, V12C-01.

Lubchenco, J., 2010. Ocean Observations: Essential for Good Stewardship. Marine Technology Society Journal 44, 6–9. https://doi.org/10.4031/MTSJ.44.6.23

Nelson, E.J., Pulla, S.T., Matin, M.A., Shakya, K., Jones, N., Ames, D.P., Ellenburg, W.L., Markert, K.N., David, C.H., Zaitchik, B.F., Gatlin, P., Hales, R., 2019. Enabling Stakeholder Decision-Making With Earth Observation and Modeling Data Using Tethys Platform. Front. Environ. Sci. 7. https://doi.org/10.3389/fenvs.2019.00148

Organization (WMO), W.M., WIGOS, 2019. Guide to the WMO Integrated Global Observing System, 2019 edition. ed, WMO. WMO, Geneva.

Pecora, S., Lins, H.F., 2020. E-monitoring the nature of water. Hydrological Sciences Journal 0, 1–16. https://doi.org/10.1080/02626667.2020.1724296

Swain, N., 2015. Tethys Platform: A Development and Hosting Platform for Water Resources Web Apps. Theses and Dissertations.

Swain, N.R., Christensen, S.D., Snow, A.D., Dolder, H., Espinoza-Dávalos, G., Goharian, E., Jones, N.L., Nelson, E.J., Ames, D.P., Burian, S.J., 2016. A new open source platform for lowering the barrier for environmental web app development. Environmental Modelling & Software 85, 11–26. https://doi.org/10.1016/j.envsoft.2016.08.003

Tarboton, D., Idaszak, R., Horsburgh, J., Heard, J., Ames, D., Goodall, J., Band, L., Merwade, V., Couch, A., Arrigo, J., Hooper, R., Valentine, D., Maidment, D., 2014. HydroShare: Advancing Collaboration through Hydrologic Data and Model Sharing, in: International Congress on Environmental Modelling and Software.

Tarboton, David G, Horsburgh, J.S., Maidment, D.R., Whiteaker, T., Zaslavsky, I., Piasecki, M., Goodall, J., Valentine, D., Whitenack, T., 2009. Development of a community hydrologic information system, in: 18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation, Modelling and Simulation Society of Australia and New Zealand and International Association for Mathematics and Computers in Simulation. pp. 988–994.

Tarboton, David G., Horsburgh, J.S., Maidment, D.R., Whiteaker, T., Zaslavsky, I., Piasecki, M., Goodall, J., Valentine, D., Whitenack, T., 2009. Development of a community hydrologic information system, in: 18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation, Modelling and Simulation Society of Australia and New Zealand and International Association for Mathematics and Computers in Simulation. pp. 988–994.

Tarboton, D.G., Horsburgh, J.S., Schreuders, K., Maidment, D.R., Zaslavsky, I., Valentine, D.W., 2010. The HydroServer Platform for Sharing Hydrologic Data. AGU Fall Meeting Abstracts 53, H53H-03.

UN-WATER, 2008. Transboundary waters: sharing benefits, sharing responsibilities [online]. Thematic paper. Available from: https://www.unwater.org/app/uploads/2017/05/UNW_TRANSBOUNDARY.pdf [accessed 2.17.21].

Whitenack, T., Valentine, D., Zaslavsky, I., Piasecki, M., Tarboton, D., Horsburgh, J., Whiteaker, T., Ames, D., Maidment, D., 2010. Hydrologic Metadata Catalog and Semantic Search Services in CUAHSI HIS. 2010 AWRA Spring Specialty Conference Geographic Information Systems (GIS) and Water Resources VI.

Zaslavsky, I., Whitenack, T., Williams, M., Tarboton, D., Schreuders, K., Aufdenkampe, A.K., 2011. The initial design of data sharing infrastructure for the Critical Zone Observatory, in: Stroud Water Research Center. Presented at the Information Management Conference, M. B. Jones and C. Gries (editors)., University of California, Santa Barbara, California., pp. 145–150.