Using Python, Partnerships, Standards and Web Services to provide Water Data for Texans

Dharhas Pothina, Andrew Wilson

✦

Abstract—Obtaining time-series monitoring data in a particular region often requires a significant effort involving visiting multiple websites, contacting multiple organizations and dealing with a variety of data formats. Although there has been a large research effort nationally in techniques to share and disseminate water related time-series monitoring data, development of a usable system has lagged. The pieces have been available for some time now, but a lack of vision,expertise, resources and software licensing requirements have hindered uptake outside of the academic research groups. The Texas Water Development Board is both a data provider and large user of data collected by other entities. As such, using the lessons learned from the last several years of research, we have implemented an expandable infrastructure for sharing water data in Texas. In this paper, we discuss the social, institutional and technological challenges in creating a system that allows discovery, access, and publication of water data from multiple federal, state, local and university sources and how we have used Python to create this system in a resource limited environment.

Index Terms—time-series, web services, waterml, data, wofpy, pyhis, HIS, hydrologic information system, cyberinfrastructure

Introduction

A wealth of physical, chemical and biological data exists for Texas' lakes, rivers, coastal bays and estuaries and the near-shore Gulf of Mexico, but unfortunately much of it remains unused by researchers and governmental entities because it is not widely disseminated and its existence not advertised. Historical data is typically stored by multiple agencies in a variety of incompatible formats, often poorly documented. Locating, retrieving and assembling these datasets into a form suitable for use in scientific studies often consumes a significant portion of many studies. Hence, despite having access to this information, much of the data remain underutilized in the hydrologic sciences due in part to the time required to access, obtain, and integrate data from different sources [Goodall2008] .

As both a consumer of water related data delivered by other entities and as a data provider whose data needs to be disseminated to external users the Texas Water Development Board (TWDB) has been researching the use of new technologies for data sharing for several years. The Texas Hydrologic Information System (Texas HIS) was born as a research project funded by the Texas Water Development Board (TWDB) and implemented by the University of Texas at Austin's Center for Research in Water Resources (UT-CRWR) as a working prototype to facilitate data discovery,
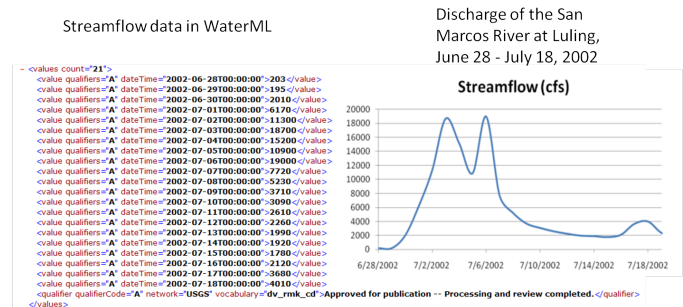

Figure 1: *An example of a WaterML format file and the data it contains.*

access, and publication of water related time-series monitoring data from multiple entities [Whiteaker2010]. It was built as an extension of the national Hydrologic Information System (HIS) that was developed by the Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) [Tarboton2009]. This prototype proved the viability of using web services along with xml based data format standards to reliably exchange data and showed how once the infrastructure was put in place, powerful standards based tools could be developed to facilitate data discovery and access [Ames2009].

Using Standards - WaterML, WaterOneFlow and the Observations Data Model

CUAHSI-HIS provides web services, tools, standards and procedures that enhance access to more and better data for hydrologic analysis [Tarboton2009]. CUAHSI-HIS has established a web service design called WaterOneFlow as a standard mechanism for the transfer of hydrologic data between hydrologic data servers (databases) and users. Web services streamline the often time-consuming tasks of extracting data from a data source, transforming it into a usable format and loading it in to an analysis environment [Maidment2009]. All WaterOneFlow web services return data in a standard format called WaterML (Figure 1). The specifics of WaterML are documented as an Open Geospatial Consortium, Inc., discussion paper [[Zaslavsky2007].

To publish data in CUAHSI-HIS, a data source provides access to their data via a WaterOneFlow web service. CUAHSI-HIS also includes mechanisms for registering WaterOneFlow web services so that users can discover and use them. Data sources often store their data locally in a CUAHSI-HIS
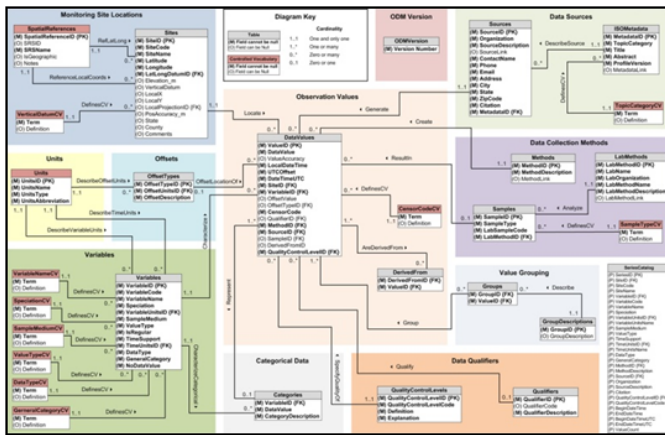
The corresponding author is with Texas Water Development Board, e-mail: dharhas.pothina@twdb.state.tx.us.

Figure 2: The Observations Data Model Schema.

Observations Data Model (ODM) database (2), where ODM is a database design for storing hydrologic time series data reported at discrete point locations [Horsburgh2008]. ODM databases, ODM data loaders, and a special version of the WaterOneFlow web service specifically designed to work with ODM as its underlying data source are all available for free on the HIS website (http://his.cuahsi.org).

In addition, CUAHSI-HIS provides several free community supported clients to search and retrieve data from WaterOne-Flow compliant services. These include a Microsoft Excel plugin called HydroExcel and a desktop GIS softare called HydroDesktop [Ames2010].

Barriers to Adoption

Although the technology seemed mature and had participation from several universities and a few large federal agencies like the United States Geological Service(USGS), uptake outside of this group was low. Data providers are often resource poor in staff, technical knowledge, time and money. In most cases, they already have a system for collecting, storing and disseminating data that works for their particular needs. In order to convince them to become part of a system like the CUAHSI-HIS, the cost of sharing their data has to be as low as possible and they have to be educated on the benifits their organization will receive through being part of the system. A review of the experiences from building the prototype Texas HIS system showed that there are significant barriers to wide scale adoption.

While there is nothing intrinsic to the CUAHSI-HIS that requires a particular software stack, for historical reasons all currently available software for both serving data and retrieving data from the CUAHSI-HIS system was built on a Microsoft .Net software stack and in some cases also needs commercial licenses. Hence, data providers who could not or did not want to use this software stack needed to write an in-house implementation of WaterOneFlow web services from the ground up. In addition, client side tools were also not cross-platform had were built for specific use cases and could not be easily adapted for alternate needs.

Changing Paradigms

Building a custom implementation of WaterOneFlow web services to attach to a datasource is a non trivial endeavour. It requires and understanding of the web services, XML and the particulars of the WaterML and WaterOneFlow. Hence, the paradigm followed by most participating data providers is to manipulate their data into an ODM database hosted on an MSSQL server. CUAHSI-HIS has a prebuilt WaterOneFlow implementation that can then be used to serve data. This approach requires that the data provider either adopt the ODM as their internal structure for storing data or they must build a translator and periodically dump data from their in-house database to the ODM database on a regular basis. The ODM schema is designed to hold data from multiple sources and hence is often much more complicated than most data providers in-house database schemas. It also excludes data providers that use non Microsoft operating systems.
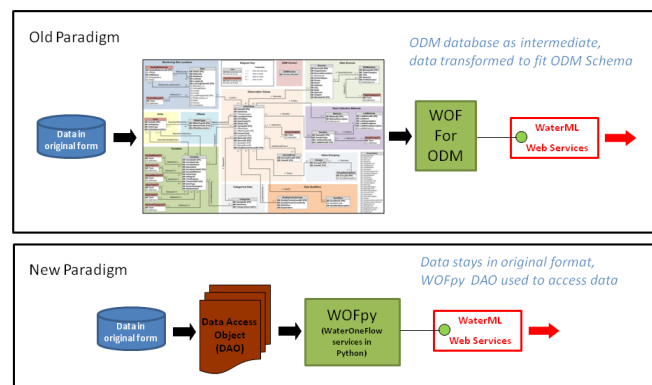

Figure 3: Comparison of changing paradigms.

Lowering these barriers requires flexible cross-platform software that can be relatively easily adapted to each organizations needs. In addition, participation in data sharing should not require large changes to an organizations internal data systems. Based on these requirements, two python modules were developed, WOFpy for serving data as WaterOneFlow services and pyhis as the basis for building customized data access tools.

Using Python to serve water data - WOFpy

WaterOneFlow in Python or WOFpy implements a reduced ODM data model that maps to WaterML objects. It has an implementation of both REST and SOAP web services that are compliant to the WaterOneFlow specification. This is done through the use of the Flask and SOAPlib python packages. On the backend, Data Access Objects (DAO's) are used to connect the services to the underlying database or storage mechanism. Through the use of the sqlalchemy python package DAO's can be written for any database backend that sqlalchemy supports. This allows a large degree of flexibility in attaching the web
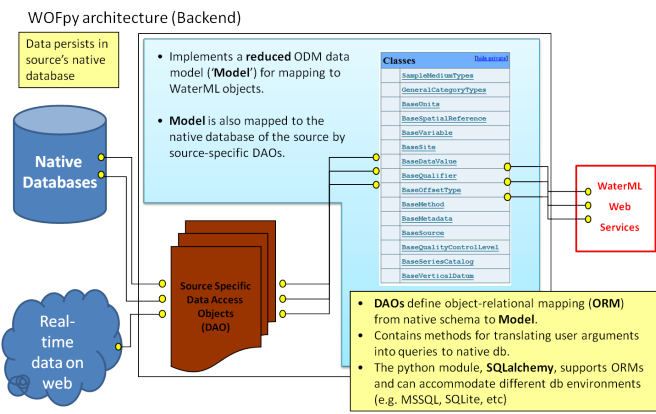
Figure 4: *Architecture of WOFpy.*

services to disparate data sources. Figure 4 shows the basic architecture of WOFpy.

WOFpy can be used to serve data from flat files, a variety of database backends and even as an on-the-fly translator of web services that use other standards.

Using python to retrieve data - pyhis

Existing CUAHSI-HIS clients are not cross-platform and are GUI based, pyhis is a command line python package that was developed to allow access to WaterOneFlow services with requiring knowledge of how the underlying web services architecture works. Pyhis uses suds to retrieve data and caches downloaded data to a local sqlite database using sqlachemy. Using pyhis more complicated scripts can be built to conduct spatial analysis or retrieve data automatically for use in real time forcast models.

Water Data For Texas

Although the development of WOFpy and pyhis has lowered the resource requirements of sharing data, a level of resources is still required. To overcome this, TWDB has formed partnerships with three agencies, The Texas Commission on Environmental Quality (TCEQ), the Texas Parks and Wildlife Department (TPWD) and the Conrad Blucher Institute for Surveying and Science(CBI) to serve their data as WaterOne-Flow services. This is being done either through scheduled data dumps or using web scapers or on-the-fly web service translations [Pothina2011]. In addition, TWDB is partnering with the Texas Natural Resource Information System (TNRIS) to build a web based map interface that can be used by the general public to find and download water data through a easy to use interface. A high level design schematic of the entire system is presented in Figure 7.

This system has been now branded *Water Data for Texas* and will reside at the url http://waterdatafortexas.org once completed.

Conclusions

Water Data for Texas is a community effort to build a robust, sustainable system for the sharing of water data across

Figure 5: *Example of using pyhis within ipython to retrieve data from the USGS National Water Information System.*

Figure 6: *Water Data for Texas logo.*

Federal, State and local entities. Parts of the system are live now with the rest expected to be completed by the Fall of 2011. Currently the system will provide access to all Nation CUAHSI-HIS datasets as well as data from the TCEQ, TPWD, CBI and TWDB. It is expected that new water related data sets will become available as more organizations choose to participate.

Python is an integral part of building this Texas-specific HIS that employs partnerships with Federal and Texas agencies to share water data. The system inherits the national CUAHSI-HIS technology and provides additional tools and services to provide ease of use and a level of quality control for partners and clients. In order to foster continued development and uptake of the technology in a community supported environment WOFpy and pyhis are being released under a BSD open source license. Development is currently taking
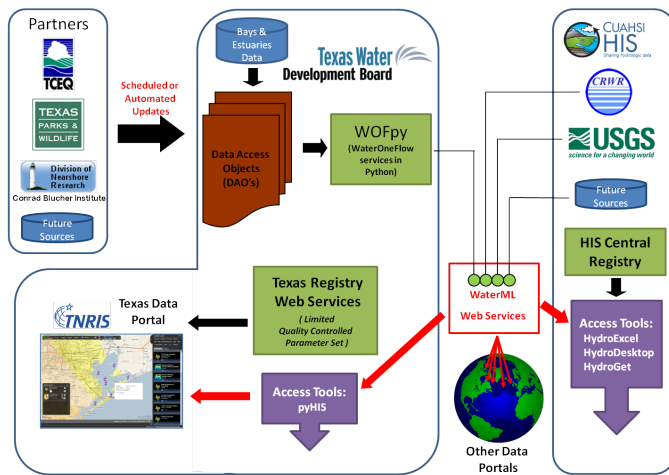
Figure 7: *Water Data for Texas Framework.*

place under the swtools organization on GitHub (https://github.com/organizations/swtools).

REFERENCES

[Zaslavsky2007] Zaslavsky, I., D. Valentine and T. Whiteaker, (2007), "CUAHSI WaterML," OGC 07-041r1, Open Geospatial Consortium Discussion Paper, http://portal.opengeospatial.org/files/?artifact_id=21743.

[Goodall2008] Goodall, J. L., J. S. Horsburgh, T. L. Whiteaker, D. R. Maidment and I. Zaslavsky, *A first approach to web services for the National Water Information System*, Environmental Modeling and Software, 23(4): 404-411, doi:10.1016/j.envsoft.2007.01.005.

[Horsburgh2008] Horsburgh, J. S., D. G. Tarboton, D. R. Maidment and I. Zaslavsky, (2008), "A Relational Model for Environmental and Water Resources Data," Water Resour. Res., 44: W05406, doi:10.1029/2007WR006392.

[Ames2009] Ames, D. P., J. Horsburgh, J. Goodall, T. Whiteaker, D. Tarboton and D. Maidment, (2009), *Introducing the Open Source CUAHSI Hydrologic Information System Desktop Application (HIS Desktop)*, 18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation, ed. R. S. Anderssen, R. D. Braddock and L. T. H. Newham, Modelling and Simulation Society of Australia and New Zealand and International Association for Mathematics and Computers in Simulation, July 2009, p.4353-4359, http://www.mssanz.org.au/modsim09/J4/ames.pdf.

[Maidment2009] Maidment, D. R., R. P. Hooper, D. G. Tarboton and I. Zaslavsky, (2009), "Accessing and Sharing Data Using CUAHSI Water Data Services," in Hydroinformatics in Hydrology, Hydrogeology and Water Resources, Edited by I. Cluckie, Y. Chen, V. Babovic, L. Konikow, A. Mynett, S. Demuth and D. A. Savic, Proceedings of Symposium JS4 held in Hyderabad, India, September 2009, IAHS Publ. 331, Hyderabad, India, p.213-223, http://iahs.info/redbooks/331.htm.

[Tarboton2009] Tarboton, D. G., J. S. Horsburgh, D. R. Maidment, T. Whiteaker, I. Zaslavsky, M. Piasecki, J. Goodall, D. Valentine and T. Whitenack, (2009), *Development of a Community Hydrologic Information System*, 18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation, ed. R. S. Anderssen, R. D. Braddock and L. T. H. Newham, Modelling and Simulation Society of Australia and New Zealand and International Association for Mathematics and Computers in Simulation, July 2009, p.988-994, http://www.mssanz.org.au/modsim09/C4/tarboton_C4.pdf.

[Ames2010] Ames, D. P., J. Kadlec, and J. Horsburgh, (2010), "HydroDesktop: A Free and Open Source Platform for Hydrologic Data Discovery, Visualization, and Analysis", Francisco Olivera (Editor), 2010 AWRA Spring Specialty Conference: Geographic Information Systems (GIS) and Water Resources VI. American Water Resources Association, TPS-10-1, ISBN 1-882132-82-3, http://his.cuahsi.org/documents/conference-awra2010/Ames_abs_13.pdf.

[Whiteaker2010] Whiteaker, T., D. Maidment, D. Pothina, J. Seppi, E. Hersh, and W. Harrison, (2010), "Tesas Hydrologic Information System", Francisco Olivera (Editor), 2010 AWRA Spring Specialty Conference: Geographic Information Systems (GIS) and Water Resources VI. American Water Resources Association, TPS-10-1, ISBN 1-882132-82-3, http://his.cuahsi.org/documents/conference-awra2010/DavidMaidment_9eb7f8b0_6581.pdf.

[Pothina2011] Pothina D., A. Wilson *Building a Coastal Geodatabasefor the State of Texas*, Report submitted to the Texas General LandOffice and the Mineral Management Service, Coastal Impact AssistanceProgram Grant Award #M09AF15208, July 2011.