**A method to Publish and Share Water Management Data through Web Services**

A short draft proposal

Adel Abdallah, Pabitra Dash Erfan Goharian, David Rosenberg, and Steven Burian

**Problem and Motivation**

There is a growing interest to share and publish water management data that are used in modeling water resources systems. Scientists are encouraged to publish their data along with their journal articles to allow reproducibility, promote transparency, and support reuse of data ([Easterbrook, 2014](#_ENREF_6); [Goodman et al., 2014](#_ENREF_10); [Hey et al., 2009](#_ENREF_12); [Rosenberg and Madani, 2014](#_ENREF_23)). The value of publishing scientific water data in a standardized format cannot be over emphasized ([Ames et al., 2012](#_ENREF_2); [Argent, 2014](#_ENREF_3); [Campbell et al., 2013](#_ENREF_5); [Elmroth et al., 2010](#_ENREF_7); [Goodman et al.](#_ENREF_10); [Govindaraju et al., 2009](#_ENREF_11); [Hey et al., 2009](#_ENREF_12); [Horsburgh et al., 2011](#_ENREF_15); [Horsburgh et al., 2009](#_ENREF_16); [Laniak et al., 2013](#_ENREF_19); [Piasecki and Beran, 2009](#_ENREF_21); [Rosenberg and Madani, 2014](#_ENREF_23)). Thus, there is a need for a generic approach to publish water management data that is domain and application independent.

Sharing and publishing water management data requires overcoming two fundamental data challenges: semantic (i.e., terminology) and syntax (i.e., data structure) heterogeneity across data sources and providers ([Argent, 2014](#_ENREF_3); [Tolk and Diallo, 2005](#_ENREF_24)). The Water Management Data Model (WaM-DaM), provides a consistent representation of water management data with supportive metadata and overcomes the mentioned two challenges ([Abdallah and Rosenberg, 2014](#_ENREF_1)). Thus, there is an opportunity to use WaM-DaM and web services to publish and share water management and modeling data. Users can then search and discover disparate water management modeming data over the web in one federated source.

**Previous work**

The Consortium of the Universities for the Advancements of the Hydrologic Science Inc. has led the efforts to standardize and manage hydrologic time series data and publish it to the broader community using web-services ([Goodall et al., 2008](#_ENREF_9); [Horsburgh et al., 2014](#_ENREF_13); [Horsburgh et al., 2008](#_ENREF_14); [Horsburgh et al., 2009](#_ENREF_16)). Other efforts to publish water management data has been either for particular regions or application specific ([Kämpgen et al., 2014](#_ENREF_17)). In addition, there are data publishing methods that are used for specific aspects of water management data (e.g., water planning and use data) like the Water Data Exchange (WaDE) project of the Western States Water Council ([Larsen and Young, 2014](#_ENREF_20)). Decision support systems also publish models and allow users to interact with them by changing some of model’s input data ([Goharian, 2015](#_ENREF_8)). There is also a method under development to share water resources networks among collaborators through web-services with loose metadata requirements like HydraPlatform ([Knox et al., 2014](#_ENREF_18)). Here we propose to use WaM-DaM as a persistent data model that supports water management data with minimum mandatory metadata to publish and share modeling data through web services.

**Contribution**

We propose a web services method to exchange and publish generic water management data based on a common data model (WaM-DaM). The web services are location and application independent. So ideally, users can share and publish any water management data in the context of systems analysis without being constraint to a specific software environment (e.g., WEAP) or a geographic area.

**Research Questions**

* How to share and publish water management data that originates with different semantics and syntax in disparate data sources and models?
* What are the required and optional metadata fields to publish water management data?
* How to search and discover water management data for models through web-services?

**Design and Methods**

First, we will determine the required and optional metadata for publishing data. For example, a data value should have an attribute, unit, source, method. We will determine the required and optional fields for the tables of these metadata items. For example, a data source must have a source name, citation, and a person name that provides the data.

Second, we will develop a Representational State Transfer (REST) Application Programming Interface (API) for WaM-DaM database using Flask open-source web development framework ([Ronacher, 2015](#_ENREF_22)). Flask is a Python based software that makes it easier to develop REST API for databases. Users submit data queries to the API using the Hypertext Transfer Protocol (HTTP) and the API responds with an HTTP response that contains data in the JavaScript Object Notation (JSON) format as a lightweight data-interchange format ([Bray, 2013](#_ENREF_4)). We already developed and tested a prototype of this API at Utah State University. Figure 1 shows the steps of publishing data using WaM-DaM data model. A WaM-DaM REST API Client which makes it easier to consume WaM-DaM REST API can be used by users to see the JSON data in tabular form and then export it to CSV files. Such a Client can be written in Python or any other programming language.

In Figure 1, we first implement WaM-DaM in a relational database (PostgreSQL). Then we import water management data to the database using Matlab. We use CSV file as a common middle data storage between different data sources and the WaM-DaM database. Importing data requires mapping each attribute and its metadata from the data source to WaM-DaM. We add appropriate metadata that describes data sources and we introduce controlled vocabularies to allow semantic interoperability between different terms. We demonstrate the process of disparate different data sources and models data to WaM-DaM. We will implement WaM-DaM REST API as a Resource in HydroShare system to demonstrate a proof of concept to publish water management data and exchange it over the web to allow users to query data across disparate data sources in a federated source. Here are the data services we plan to implement and demonstrate: GETDataStructures, GETObjectTypes, GETNetworks, GetScenarios, GetInstances, GetAttributeInfo, and GetDataStorage. The reader here is encouraged to refer to WaM-DaM design specifications for details about these data objects: https://github.com/amabdallah/WaM-DaM1.0

**Uses Cases**

* Share and publish data for the Parley’s Canyon model in GoldSim software through WaM-DaM “standard” REST API. Allow users to query data Parley’s Canyon model. Demonstrate key quires for data and metadata about the model.
* Share and publish national data for dams, waterbodies, and stream networks. Search and discover data across these data sources for the Parley’s Canyon model ([Goharian, 2015](#_ENREF_8)) through controlled vocabularies.

**Added Value and** **Benefits to the Water Resources Community**

Interested water managers, scientists, or the public can access and use the self-describing data of the Parley’s Canyon model to answer different data questions about the system. Users may use Parleys Canyon data for other models or they can use it to reproduce the models results. They also can search and discover national water management data across different sources using the same controlled vocabulary.

C:\Users\Adel\Desktop\API_WaM-DaM.emf

**Figure 1:** Schematic of the steps to publish water management data from multiple sources using the WaM-DaM method

**References:**

Abdallah, A.M., Rosenberg, D.E., 2014. WaM-DaM: A Data Model to Organize and Synthesize Water Management Data, In: Ames, D.P., Quinn, N., Rizzoli, A.E. (Eds.), 7th International Congress on Environmental Modelling and Software. International Environmental Modelling and Software Society (iEMSs).

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(0) 146-156.

Argent, R.M., 2014. Information modelling in water resources: an Australian perspective. Stochastic Environmental Research and Risk Assessment 28(1) 137-145.

Bray, T., 2013. The JSON Data Interchange Format. Ecma International.

Campbell, J.L., Rustad, L.E., Porter, J.H., Taylor, J.R., Dereszynski, E.W., Shanley, J.B., Gries, C., Henshaw, D.L., Martin, M.E., Sheldon, W.M., Boose, E.R., 2013. Quantity is Nothing without Quality: Automated QA/QC for Streaming Environmental Sensor Data. BioScience 63(7) 574-585.

Easterbrook, S.M., 2014. Open code for open science? Nature Geosci 7(11) 779-781.

Elmroth, E., Hernández, F., Tordsson, J., 2010. Three fundamental dimensions of scientific workflow interoperability: Model of computation, language, and execution environment. Future Generation Computer Systems 26(2) 245-256.

Goharian, E., Burian, S., Bardsley, T., Strong, C., 2015. A Vulnerability Assessment to Integrate Flooding and Shortage of Water Supply Systems under Climate Change Conditions. ournal of Water Resources Planning and Management.

Goodall, J.L., Horsburgh, J.S., Whiteaker, T.L., Maidment, D.R., Zaslavsky, I., 2008. A first approach to web services for the National Water Information System. Environmental Modelling & Software 23(4) 404-411.

Goodman, A., Pepe, A., Blocker, A.W., Borgman, C.L., Cranmer, K., Crosas, M., Di Stefano, R., Gil, Y., Groth, P., Hedstrom, M., 2014. Ten Simple Rules for the Care and Feeding of Scientific Data. PLoS computational biology 10(4) e1003542.

Govindaraju, R., Engel, B., Ebert, D., Fossum, B., Huber, M., Jafvert, C., Kumar, S., Merwade, V., Niyogi, D., Oliver, L., Prabhakar, S., Rochon, G., Song, C., Zhao, L., 2009. Vision of Cyberinfrastructure for End-to-End Environmental Explorations (C4E4). Journal of Hydrologic Engineering 14(1) 53-64.

Hey, A., Tansley, S., Tolle, K., 2009. The Fourth Paradigm: Data-Intensive Scientific Discovery. Microsoft Research, Redmond, Wash.

Horsburgh, J.S., Tarboton, D.G., Hooper, R.P., Zaslavsky, I., 2014. Managing a community shared vocabulary for hydrologic observations. Environmental Modelling & Software 52(0) 62-73.

Horsburgh, J.S., Tarboton, D.G., Maidment, D.R., Zaslavsky, I., 2008. A relational model for environmental and water resources data. Water Resour. Res. 44(5) W05406.

Horsburgh, J.S., Tarboton, D.G., Maidment, D.R., Zaslavsky, I., 2011. Components of an environmental observatory information system. Computers & Geosciences 37(2) 207-218.

Horsburgh, J.S., Tarboton, D.G., Piasecki, M., Maidment, D.R., Zaslavsky, I., Valentine, D., Whitenack, T., 2009. An integrated system for publishing environmental observations data. Environmental Modelling & Software 24(8) 879-888.

Kämpgen, B., Riepl, D., Klinger, J., 2014. SMART Research using Linked Data - Sharing Research Data for Integrated Water Resources Management in the Lower Jordan Valley, In: Alexander García Castro, Christoph Lange, Phillip Lord, Robert Stevens (Eds.), Proceedings of the 4th Workshop on Semantic Publishing co-located with the 11th Extended Semantic Web Conference (ESWC 2014): Anissaras, Greece.

Knox, S., Meier, P., Harou, J., 2014. Web service and plug-in architecture for flexibility and openness of environmental data sharing platforms, In: Ames, D.P., Quinn, N., Rizzoli, A.E. (Eds.), 7th International Congress on Environmental Modelling and Software. San Diego, California, USA.

Laniak, G.F., Olchin, G., Goodall, J., Voinov, A., Hill, M., Glynn, P., Whelan, G., Geller, G., Quinn, N., Blind, M., Peckham, S., Reaney, S., Gaber, N., Kennedy, R., Hughes, A., 2013. Integrated environmental modeling: A vision and roadmap for the future. Environmental Modelling & Software 39(0) 3-23.

Larsen, S.G., Young, D., 2014. WaDE: An Interoperable Data Exchange Network for Sharing Water Planning and Use Data. Journal of Contemporary Water Research & Education(153) 33-41.

Piasecki, M., Beran, B., 2009. A semantic annotation tool for hydrologic sciences. Earth Science Informatics 2(3) 157-168.

Ronacher, A., 2015. Flask: Web Development, one drop at a time.

Rosenberg, D., Madani, K., 2014. Water Resources Systems Analysis: A Bright Past and a Challenging but Promising Future. Journal of Water Resources Planning and Management 140(4) 407-409.

Tolk, A., Diallo, S.Y., 2005. Model-based data engineering for Web services. Internet Computing, IEEE 9(4) 65-70.