8-2020

Advancing Water Resources Systems Modeling Cyberinfrastructure to Enable Systematic Data Analysis, Modeling, and Comparisons

Adel Mohammad Kheir Abdallah
Utah State University

ADVANCING WATER RESOURCES SYSTEMS MODELING CYBERINFRASTRUCTURE TO ENABLE SYSTEMATIC DATA ANALYSIS, MODELING, AND COMPARISONS

by

Adel Mohammad Kheir Abdallah

A dissertation submitted in partial fulfillment
of the requirements for the degree
of

DOCTOR OF PHILOSOPHY
in
Civil and Environmental Engineering

Approved:

David E. Rosenberg, Ph.D. Major Professor	David G. Tarboton, Sc.D. Committee Member
Jeffery S. Horsburgh, Ph.D. Committee Member	Steven J. Burian, Ph.D. Committee Member
Sarah E. Null, Ph.D. Committee Member	Richard S. Inouye, Ph.D. Vice Provost for Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2020

ABSTRACT

Advancing Water Resources Systems Modeling Cyberinfrastructure To Enable Systematic Data Analysis, Modeling, and Comparisons

by

Adel Mohammad Kheir Abdallah, Doctor of Philosophy

Utah State University, 2020

Major Professor: Dr. David E. Rosenberg
Department: Civil and Environmental Engineering

Since its emergence half a century ago, the water resources systems analysis community has made significant advancements to improve the modeling of interrelated natural and built water resources infrastructure and inform decisions regarding systems planning and management. Despite modeling advances, modelers face three basic technical challenges to i) identify, organize, and analyze data used in models that are stored and described in different formats and vocabularies, ii) prepare and populate data to models, and iii) visualize system model networks, plot, and compare input and output for different management scenarios. Existing tools to store, query, and visualize modeling data are model, location, and dataset-specific, and developing such tools is time-consuming and requires programming experience.

This dissertation contributes a novel software architecture and tools that generalize data management used in modeling water systems to enable systematic data and modeling comparisons and reuse across many models and datasets. First, the Water Management

Data Model (WaMDaM) is designed to help modelers organize, store, and compare water management data from multiple sources and models. WaMDaM uses metadata to help interpret and relate values and controlled vocabulary across models. Second, an open-source Python software is designed to automate the process to prepare and load large input data into the Water Evaluation and Planning system (WEAP) model or extract for already existing WEAP models outside its proprietary database using its Application Programming Interface. Third, a software interoperability among WaMDaM and other existing independently developed, state-of-the-art, generalized tools is designed to visualize water resources systems modeling data. The software connected Hydra Platform, OpenAgua, and HydroShare web-based tools to visualize, compare, edit, publish, discover, and analyze model networks, input, and output data for many models.

The dissertation software architecture was guided and demonstrated by use cases that represent common tasks performed by modelers and water managers over a dozen of different water resources datasets and four models in three watersheds located in the USA and Mexico. The use cases show a fundamental significance to the science of water management by enabling comparisons that generate insight across datasets and models within or across study locations.

(183 Pages)

PUBLIC ABSTRACT

Advancing Water Resources Systems Modeling Cyberinfrastructure to Enable Systematic Data Analysis, Modeling, and Comparisons

Adel Mohammad Kheir Abdallah

Water resources systems models aid in managing water resources holistically considering water, economic, energy, and environmental needs, among others. Developing such models require data that represent a water system's physical and operational characteristics such as inflows, demands, reservoir storage, and release rules. However, such data is stored and described in different formats, metadata, and terminology. Therefore, Existing tools to store, query, and visualize modeling data are model, location, and dataset-specific, and developing such tools is time-consuming and requires programming experience. This dissertation presents an architecture and three software tools to enable researchers to more readily and consistently prepare and reuse data to develop, compare, and synthesize results from multiple models in a study area: (1) a generalized database design for consistent organization and storage of water resources datasets independent of study area or model, (2) software to extract data out of and populate data for any study area into the Water Evaluation and Planning system, and (3) software tools to visualize online, compare, and publish water management networks and their data for many models and study areas. The software tools are demonstrated using dozens of example and diverse local, regional, and national datasets from three watersheds for four models; the Bear and Weber Rivers in the USA and the Monterrey River in Mexico.

DEDICATION

To my Parents Mohammad Kheir and Monira,

To my siblings Samira, Khalid, Sumaya, Sulaiman, Alaa

To my wife Allia and our son Kareem

To my in-laws, Maher, Sharifa, Anisa, Rashad, and Jenna

To my extended family, uncles, aunts, cousins

To my teachers and mentors and fellow students

To my friends

To the generous donors who supported my scholarships and awards

To Palestine and its oppressed people and their struggle to justice and freedom

To my fellow Arabs

To my brothers and sisters in faith, Islam

To my fellow Americans and taxpayers who funded this research

To humanity

You made me who I am and this dissertation was only possible because of you

ACKNOWLEDGMENTS

Up until my senior year in my Civil Engineering studies, I was not interested in pursuing a degree in graduate school, but passion and support can lead into unexpected pathways. Being the first generation in my family to have the opportunity to earn a college degree, I would like to acknowledge a few people who helped me go further and earn this Ph.D.

My mom only attended school until fifth grade and my dad went to study in college in Amman, Jordan back in the 1970s, only to learn that his first college class did not have enough students to launch the program. It must have been very disappointing to my dad that he returned home without getting that college degree. My parents cannot be prouder of my accomplishments and I am very excited to finish this degree which was possible because of their daily sacrifices to raise me and my five siblings in the unstable and occupied Palestinian territories. My mom, dad, and extended family utmost believe that education has been critical to my journey throughout life. I have no doubt that education and hard work can help overcome many obstacles.

I owe a sincere gratitude to my first mentor, Dr. Mohammad Almasri at An-Najah National University in Palestine, who is also an alumnus of Utah State University. Mohmmad has not only believed in my abilities but also encouraged and supported me to do research as an undergraduate student. I was his first undergraduate mentee and I am sure that earning this degree means a great deal to him. I owe Dr. Laurie McNeill a debt, as she also took a chance to co-mentor my undergraduate capstone project for a full year when she was spending her sabbatical at An-Najah University in 2007 and 2008. Laurie made

research fun and exciting and opened my eyes further to Utah State University, and ultimately and critically supported my application to the USU Graduate School. I am glad to set an example to professors thinking about sabbaticals overseas where they can have an incredible impact in mentoring students and change their lives, for good.

Then in 2010, Dr. David Rosenberg took the torch in believing in me, not only in a master's degree under his mentorship, but also for this PhD. I came to admire David's utmost dedication to science, education, and mentoring. David encouraged and guided me to persist in achieving concreate contributions to both the science of water management and hydroinformatics. David has given me probably a great deal of freedom to chart this PhD dissertation, which in probably has contributed to my long journey. His reminders that "science has no deadline" and "there is no rush in science" were helpful, yet harder to believe for people around me after many years in graduate school. Besides my deep self-motivation, receiving many of scholarships and awards along the way was very rewarding and encouraging to keep going.

During my graduate school journey, I had unwavering feelings and dedication to challenge myself to go above and beyond what I could normally do and stay faithful to the opportunity that I was given. It is an opportunity that thousands of students around the world dream of, just like my dad's dream to get a college degree. The subject matter of hydroinformatics as a relatively new field was very challenging, not necessarily because it is difficult, but because it is a non-traditional interdisciplinary field that is harder to make the case for its contributions to either water resources engineering or informatics communities.

During my long learning journey, I came to realize that education and science transcend political, geographical, and social boundaries. I believe I am a living example of this fact and I will be its champion for others. I am a first-generation college graduate who grew up in Palestine speaking Arabic as a native language, yet with the support of countless people, I have earned this PhD degree at Utah State University in the United States. I am thankful to countless people who contributed to my success and I aspire to do the same and help the next generation of students to realize their potential regardless of their nationality or whomever they are. I am very satisfied with my contributions and my time in graduate school has very well prepared me to my career. I hope others look into my research and find it useful.

Adel M. Abdallah

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

CHAPTER I

Introduction

Data synthesis and analysis are necessary in developing water resources management models (Loucks et al., 2005), and the way data are organized can enable or inhibit the analysis that water managers and researchers perform (Horsburgh et al., 2008). Current practices to organize, manipulate, compare, prepare, and visualize water resources data in developing water resources systems models are specific to the data sources, models, and study location (Brown et al., 2015). Water resources systems models represent the natural and built environment and their interactions as networks of nodes and links. Source-, model-, and study area-specific practices arise because models have varying data requirements for their components, store data in different file formats, have varying spatial coverage, use inconsistent metadata to describe methods, sources, and units, and use different semantic terms to name similar system components and their attributes (Miller et al., 2004; Laituri and Sternlieb, 2014; Maidment, 2016). This heterogeneity hampers synthesis of information from multiple studies (Brown et al., 2015), and source-, model- and study area-specific practices often require considerable effort and time to develop models (Ridley and Stoker, 2001; Draper et al., 2003; Miller et al., 2004; CUAHSI, 2005; Michener, 2006; Maidment, 2008; Hey et al., 2009; Beniston et al., 2012; Leonard and Duffy, 2013; Watkins, 2013). Most of the published studies in the broad field of hydrology do not have their data published which inhibits data reuse, synthesis, and study reproducibility (Stagge et al., 2019; Rosenberg et al., 2020). Modelers would benefit from generalized tools that work for multiple datasets, models, and study locations to i) organize and store data with consistent metadata and terminology, ii) automate loading data to models, iii) visualize and compare results in a web-browser, and publish modeling data (Bajcsy, 2008; Govindaraju et al., 2009; Brown et al., 2015; Vogel et al., 2015). These tools should be reusable, independent of any systems modeling specific software, and require minimal programming to increase the chance of their uptake by the water resources community.

Existing methods for organizing water management data provide limited capabilities across many systems models and their different data types. Data used in systems modeling include: 1) representations of different water resources systems components in space through nodes and links, including hydrology, infrastructure, and demand sites, and 2) multiple data types that represent quantitative and qualitative attributes of the system components like time series and multi-column arrays. As an example system, the Data System Storage of the Hydrologic Engineering Center of the U.S. Army Corps of Engineers (HEC-DSS) organizes and retrieves large sequential datasets, like time series and paired tabular data to support hydrologic and hydraulic modeling using HEC models (HEC, 2009). In some cases, the HEC-DSS is used to manage and query a water resources model's time series data through its propriety software, but users organize any data about the network's nodes and links in spreadsheets or Microsoft Access (Jenkins et al., 2004).

As another example, the Arc Hydro Framework data model organizes hydrologic data with limited metadata for hydrologic system components, including stream networks, monitoring points, and watersheds within the propriety ArcGIS environment (Maidment, 2002). Organizing time series data for system components like monitoring sites along with its metadata in Arc Hydro requires users to adopt either a rudimentary representation of time series metadata or to pair Arc Hydro with other data models like the Observations Data Model (ODM1 and ODM2) for spatially discrete locations of environmental and earth observations (Horsburgh et al., 2008; Horsburgh et al., 2016). ODM uses metadata to describe monitoring sites, observed variables, units, sources, and methods used to collect and measure observations at a site. ODM also uses controlled vocabularies to reconcile the use of different terms for synonymous variables.

Other data systems like HydraPlatform have functionality to organize, visualize, and export systems water management data to simulation and optimizations solvers like the General Algebraic Modeling System (GAMS) (Harou et al., 2010; Knox et al., 2014; Rosenthal, 2014). HydraPlatform uses a binary data storage format for its time series and multi-column arrays, which needs third party software to access and compare its stored data. Other water resources simulation and optimization models, such as RiverWare (Zagona et al., 2001), and Water Evaluation and Planning (WEAP) (Yates et al., 2005), manage input and output data using their own specific and sometimes propriety data storage systems.

The different and, in some cases, proprietary data management environments for systems models hampers efforts to prepare input data and reuse their output data as input to other aggregate models. Researchers often need to write specific scripts or use manual methods to prepare input data for models or reuse output data of a small scale model as input for another regional model (Wurbs, 2005). Existing methods to facilitate exchanging specific output data from one model as input to another are more prominent in hydrologic models, and they are intended to exchange data during simulation, like the Open Modeling

Interface (OpenMI) (Moore and Tindall, 2005) and the Community Surface Dynamics Modeling System (CSDMS) (Peckham et al., 2013). Such methods are mainly used to couple components of hydrologic models to execute in sequence without archiving either of the models' data. For water resources systems models, we focus on archiving models input and output to allow for their reuse in other models and data synthesis after the study is completed.

This dissertation presents a framework to advance cyberinfrastructure in three software tools to enable systematic data analysis, modeling, and cross-comparisons between overlapping datasets and models. The cross-comparisons demonstrate a fundamental scientific activity that is needed and used by water resources systems modelers in developing models. Comparisons also show how the same software tools work for many datasets and models opposed to existing tools that often focus on a single model or dataset. The dissertation focuses on developing open source tools to enable their progress by the water resources community with no cost barrier to use. I use Python as the main programming language due to its powerful open source capabilities for data manipulations and visualizations. The dissertation is presented in the following three standalone chapters. Chapter 5 summarizes the dissertation and suggests future work.

1. A Data Model to Manage Data for Water Resources Systems Modeling

Limitations with model and dataset-specific methods to identify, organize, analyze, and serve data to water resources systems models are addressed by designing a generalized database design and supporting software tools to organize and store water management data from multiple sources and models. The overarching motivating question is: how can data from multiple sources be organized and described in a semantically and syntactically consistent way to facilitate data query, comparison, joining, and analysis that will ultimately help modelers choose input data to build and run water resources systems models? The main contributions of this work include:

- Design of the Water Management Data Model (WaMDaM) that allows modelers to use metadata and controlled vocabularies to link water systems terms across different datasets and models.
- Prototype software tools that enable modelers to manage shared controlled vocabularies online and help them load datasets into an instance of the WaMDaM relational data model.
- Demonstrate five use cases with thirteen overlapping datasets and models focused in the Bear River Watershed, United States to show how a user can identify, compare, and choose from multiple types of data, networks, and scenario elements then serve data to models.

2. Open Source Python Software to Manage, Populate, and Compare WEAP Models and Scenarios

Limitations in study-specific methods to prepare and populate the world-wide used Water Evaluation and Planning system (WEAP) model with input data and perform sensitivity analysis are addressed by designing an open-source Python software that generalizes and automates the process to prepare and load large input data into WEAP, or extract its network and data for many already existing models and scenarios. In one application, input data are often needed for sensitivity analysis that quantify the effect of changes in systems operation, physical, or socio-economic factors on the system performance such as meeting demand. The overarching motivating question is: how to automate the process to extract data out of WEAP and populate it with input data to enable reusable, comparative data and scenarios analysis across WEAP models? The main contributions of this chapter include:

- Design of generic data workflows to first allow modelers to extract networks and data for WEAP models and load them into WaMDaM to then publish data to the HydroShare online repository. Second, design data workflows to allow modelers to prepare and populate WEAP models with input data from WaMDaM as a single source of consistent data that originates from multiple disparate datasets.
- Allow modelers to programmatically query input data of the two different WEAP models extracted into WaMDaM database to compare and benchmark how regulated their river basins against others.
- Allow modelers to perform automated sensitivity analysis and compare how water system's demand reliability in two different WEAP models in response to changes in changes in reservoir capacity, demand, evaporation, and river headflows.

3. An Interoperable Software Ecosystem to Store, Visualize Online, and Publish Water Resources Systems Modelling Data

Limitations in model-specific software tools to store, visualize, edit, run, and publish systems modeling data are addressed by coupling WaMDaM with three existing independently developed, state-of-the-art, generalized software tools into a software ecosystem. The tools are Hydra Platform web service for systems modeling data, OpenAgua for visualizing systems modeling data online, and HydroShare to publish modeling data and enable their discovery and analysis. The overarching motivating question is: how can data of multiple systems models be stored, visualized, and published using existing interoperable software tools to facilitate systems modeling and scenario comparisons? The main contributions of this chapter include:

- Couple data transfer between WaMDaM with Hydra Platform, OpenAgua, and HydroShare to allow modelers to store data, visualize it and publish it online.
- Three use cases that show how modelers can systematically reuse software ecosystem tools and web services to visualize three different models in the Bear River Watershed, United States and Monterrey, Mexico, set up scenarios, update input data, and compare model outputs. The use cases offer comparison insights into similarities and differences across the three models in different regions.

References

Abdallah, A.M., 2018. The WaMDaM Wizard Source Code. GitHub.

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).

Bajcsy, P., 2008. A Perspective on Cyberinfrastructure for Water Research Driven by Informatics Methodologies. Geography Compass 2(6) 2040-2061.

Beniston, M., Stoffel, M., Harding, R., Kernan, M., Ludwig, R., Moors, E., Samuels, P., Tockner, K., 2012. Obstacles to data access for research related to climate and water: Implications for science and EU policy-making. Environmental Science & Policy 17(0) 41-48.

Brown, C.M., Lund, J.R., Cai, X., Reed, P.M., Zagona, E.A., Ostfeld, A., Hall, J., Characklis, G.W., Yu, W., Brekke, L., 2015. The future of water resources systems analysis: Toward a scientific framework for sustainable water management. Water Resources Research.

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.

Cantor, A., Michael Kiparsky, Rónán Kennedy, Susan Hubbard,, Roger Bales, L.C.P., Kamyar Guivetchi,, Christina McCready, a.G.D., 2018. Data for Water Decision Making: Informing the Implementation of California’s Open and Transparent Water Data Act through Research and Engagement.

Carata, L., Akoush, S., Balakrishnan, N., Bytheway, T., Sohan, R., Seltzer, M., Hopper, A., 2014. A primer on provenance. Commun. ACM 57(5) 52-60.

Chen, P.P.-S., 1976. The entity-relationship model - Toward a unified view of data. ACM Trans. Database Syst. 1(1) 9-36.

Codd, E.F., 1970. A relational model of data for large shared data banks. Commun. ACM 13(6) 377-387.

Connolly, T.M., Begg, C.E., 2010. Database systems : a practical approach to design, implementation, and management. Addison-Wesley, Boston, Mass.

Consortium of Universities for the Advancement of Hydrologic Science Inc (CUAHSI), 2005. Hydrologic Information System Status Report, In: Maidment, D.R. (Ed.).

Couch, A., Hooperb, R., Pollakb, J., Martinb, M., Seulb, M., 2014. Enabling Water Science at the CUAHSI Water Data Center, iEMSs 2014 Conference, p. 1.

Dodd, B., 2016. The Open and Transparent Water Data Act, In: California, S.o. (Ed.), AB-1755.

Draper, A., Jenkins, M., Kirby, K., Lund, J., Howitt, R., 2003. Economic-Engineering Optimization for California Water Management. Journal of Water Resources Planning and Management 129(3) 155-164.

Dublin Core Metadata Initiative (DCMI), 2013. Dublin Core Metadata Element Set, Version 1.1 ANSI/NISO Z39.85-2012
National Information Standards Organization (NISO): Baltimore, MD.

Duval, E., Hodgins, W., Sutton, S., Weibel, S.L., 2002. Metadata principles and practicalities. D-lib Magazine 8(4) 16.

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

Gil, Y., David, C.H., Demir, I., Essawy, B.T., Fulweiler, R.W., Goodall, J.L., Karlstrom, L., Lee, H., Mills, H.J., Oh, J.-H., Pierce, S.A., Pope, A., Tzeng, M.W., Villamizar, S.R., Yu, X., 2016. Toward the Geoscience Paper of the Future: Best practices for documenting and sharing research from data to software to provenance. Earth and Space Science 3(10) 388-415.

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.

Gray, J., Liu, D.T., Nieto-Santisteban, M., Szalay, A., DeWitt, D.J., Heber, G., 2005. Scientific data management in the coming decade. SIGMOD Rec. 34(4) 34-41.

Harou, J.J., Pinte, D., Tilmant, A., Rosenberg, D.E., Rheinheimer, D.E., Hansen, K., Reed, P.M., Reynaud, A., Medellin-Azuara, J., Pulido-Velazquez, M., Matrosov, E., Padula, S., Zhu, T., 2010. An open-source model platform for water management that links models to a generic user-interface and data-manager In: David A. Swayne, W.Y., A. A. Voinov, A. Rizzoli, T. Filatova (Ed.), International Congress on Environmental Modelling and Software, Modelling for Environment's Sake ed. International Environmental Modelling and Software Society (iEMSs) Ottawa, Ontario, Canada.

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

Horsburgh, J.S., Aufdenkampe, A.K., Mayorga, E., Lehnert, K.A., Hsu, L., Song, L., Jones, A.S., Damiano, S.G., Tarboton, D.G., Valentine, D., Zaslavsky, I., Whitenack, T., 2016. Observations Data Model 2: A community information model for spatially discrete Earth observations. Environmental Modelling & Software 79 55-74.

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.

HydroLogics, 2009. User Manual for OASIS WITH OCL.

Jenkins, M., Lund, J., Howitt, R., Draper, A., Msangi, S., Tanaka, S., Ritzema, R., Marques, G., 2004. Optimization of California's Water Supply System: Results and Insights. Journal of Water Resources Planning and Management 130(4) 271-280.

Knox, S., 2018. Hydra Platform handling of units and dimensions. http://umwrg.github.io/HydraPlatform/devdocs/techdocs/units_and_dimensions.html?highlight=units

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.

Laituri, M., Sternlieb, F., 2014. Water Data Systems: Science, Practice, and Policy. Journal of Contemporary Water Research & Education 153(1) 1-3.

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.

Leonard, L., Duffy, C.J., 2013. Essential Terrestrial Variable data workflows for distributed water resources modeling. Environmental Modelling & Software 50(0) 85-96.

Loucks, D.P., Van Beek, E., Stedinger, J.R., Dijkman, J.P., Villars, M.T., 2005. Water resources systems planning and management: an introduction to methods, models and applications. Paris: UNESCO.

Maidment, D.R., 2002. Arc hydro : GIS for water resources. ESRI Press, Redlands, Calif.

Maidment, D.R., 2008. Bringing Water Data Together. Journal of Water Resources Planning and Management 134(2) 95-96.

Maidment, D.R., 2016. Open Water Data in Space and Time. JAWRA Journal of the American Water Resources Association n/a-n/a.

Michener, W.K., 2006. Meta-information concepts for ecological data management. Ecological Informatics 1(1) 3-7.

Miller, R.C., Guertin, D.P., Heilman, P., 2004. Information Technology in Watershed Management Decision Making. JAWRA Journal of the American Water Resources Association 40(2) 347-357.

Moore, R.V., Tindall, C.I., 2005. An overview of the open modelling interface and environment (the OpenMI). Environmental Science & Policy 8(3) 279-286.

Morsy, M.M., Goodall, J.L., Castronova, A.M., Dash, P., Merwade, V., Sadler, J.M., Rajib, M.A., Horsburgh, J.S., Tarboton, D.G., 2017. Design of a metadata framework for environmental models with an example hydrologic application in HydroShare. Environmental Modelling & Software 93(Supplement C) 13-28.

Peckham, S.D., Hutton, E.W.H., Norris, B., 2013. A component-based approach to integrated modeling in the geosciences: The design of CSDMS. Computers & Geosciences 53(0) 3-12.

Pokorný, J., 2006. Database architectures: Current trends and their relationships to environmental data management. Environmental Modelling & Software 21(11) 1579-1586.

Refsgaard, J.C., Nilsson, B., Brown, J., Klauer, B., Moore, R., Bech, T., Vurro, M., Blind, M., Castilla, G., Tsanis, I., Biza, P., 2005. Harmonised techniques and representative river basin data for assessment and use of uncertainty information in integrated water management (HarmoniRiB). Environmental Science & Policy 8(3) 267-277.

Ridley, M., Stoker, C., 2001. Data Management Tools, In: Shelke, D.P.G. (Ed.), Bridging the Gap: Meeting the World's Water and Environmental Resources Challenges. American Society of Civil Engineers: Orlando, Florida, United States, pp. 1-10.

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.

Rosenberg, D.E., Fillion, Y., Teasley, R.L., Sandoval-Solis, S., Hecht, J.S., Zyl, J.E.v., McMahon, G.F., Jeffery S. Horsburgh, Kasprzyk, J.R., Tarboton, D.G., 2020. The Next Frontier: Making Research More Reproducible. Journal of Water Resources Planning and Management.

Rosenthal, R.E., 2014. General Algebraic Modeling System (GAMS) | A User's Guide. .

Rossman, L.A., 2000. EPANET 2: Users Manual. US Environmental Protection Agency: Cincinnati, Ohio.

Sarle, W.S., 1995. Measurement theory: Frequently asked questions. Disseminations of the International Statistical Applications Institute 1(4) 61-66.

Sears, R., Ingen, C.v., Gray, J., , 2006. To BLOB or Not To BLOB: Large Object Storage in a Database or a Filesystem? , Microsoft Research Microsoft Microsoft Corporation One Microsoft Way Redmond, WA 98052

Stagge, J.H., Rosenberg, D.E., Abdallah, A.M., Akbar, H., Attallah, N.A., James, R., 2019. Assessing data availability and research reproducibility in hydrology and water resources. Scientific Data 6 190030.

Szalay, A.S., Blakeley, J.A., 2009. Gray's laws: database-centric computing in science.

US Army Corps of Engineers Hydrologic Information Center (HEC), 2009. HEC Data Storage System, 2.0 ed. US Army Corps of Engineers Institute for Water Resources Hydrologic Engineering Center (HEC) Davis, CA.

Vogel, R.M., Lall, U., Cai, X., Rajagopalan, B., Weiskel, P.K., Hooper, R.P., Matalas, N.C., 2015. Hydrology: The interdisciplinary science of water. Water Resources Research 51(6) 4409-4430.

Watkins, D.W., 2013. Water resources systems analysis through case studies data and models for decision making. Environmental Water Resources Institute. American Society of Civil Engineers., Reston, Virginia.

Wurbs, R.A., 1993. Reservoir-System Simulation and Optimization Models. Journal of Water Resources Planning and Management-Asce 119(4) 455-472.

Wurbs, R.A., 2005. Comparative evaluation of generalized river/reservoir system models. Texas Water Resources Institute: College Station, TX.

Wurbs, R.A., 2012. Reservoir/River System Management Models. The Texas Water Journal 3(1) 16.

Yates, D., Sieber, J., Purkey, D., Huber-Lee, A., 2005. WEAP21—A Demand-, Priority-, and Preference-Driven Water Planning Model. Water International 30(4) 487-500.

Zagona, E.A., Fulp, T.J., Shane, R., Magee, T., Goranflo, H.M., 2001. RiverWare: A Generalized Tool for Complex Reservoir System Modeling. JAWRA Journal of the American Water Resources Association 37(4) 913-929.

Zeiler, M., 1999. Modeling our world : the ESRI guide to geodatabase design. Environmental Systems Research Institute (ESRI) Press, Redlands, CA.

CHAPTER II

A DATA MODEL TO MANAGE DATA FOR WATER RESOURCES SYSTEMS MODELING[1]

Abstract

Current practices to identify, organize, analyze, and serve data to water resources systems models are typically model and dataset specific. Data are stored in different formats, described with different vocabularies, and require manual, model-specific, and time-intensive manipulations to find, organize, compare, and then serve to models. This paper presents the Water Management Data Model (WaMDaM) implemented in a relational database. WaMDaM uses metadata, controlled vocabularies, and supporting software tools to organize and store water management data from multiple sources and models and allow users to more easily interact with its database. Five use cases use thirteen datasets and models focused in the Bear River Watershed, United States to show how a user can identify, compare, and choose from multiple types of data, networks, and scenario elements then serve data to models. The database design is flexible and scalable to accommodate new datasets, models, and associated components, attributes, scenarios, and metadata.

[1] Abdallah, Adel M., and David E. Rosenberg. "A data model to manage data for water resources systems modeling." Environmental Modelling & Software 115 (2019): 113-127.

Keywords

Data management, systems analysis, systems modeling, data fusion, water resources, open-source

Highlights

- We present a data model to organize water resources systems data and models
- Controlled vocabularies link native terms across different datasets and models
- Software tools manage controlled vocabularies and help load datasets
- Modelers can identify and compare available data then serve data to models

Software availability

Name of software: Water Management Data Model (WaMDaM)

Developer: Adel M. Abdallah

Contact: Adel M. Abdallah; 8200 Old Main Hill, Logan, UT 84322, USA; Email amabdallah@aggiemail.usu.edu

Year first available: 2018

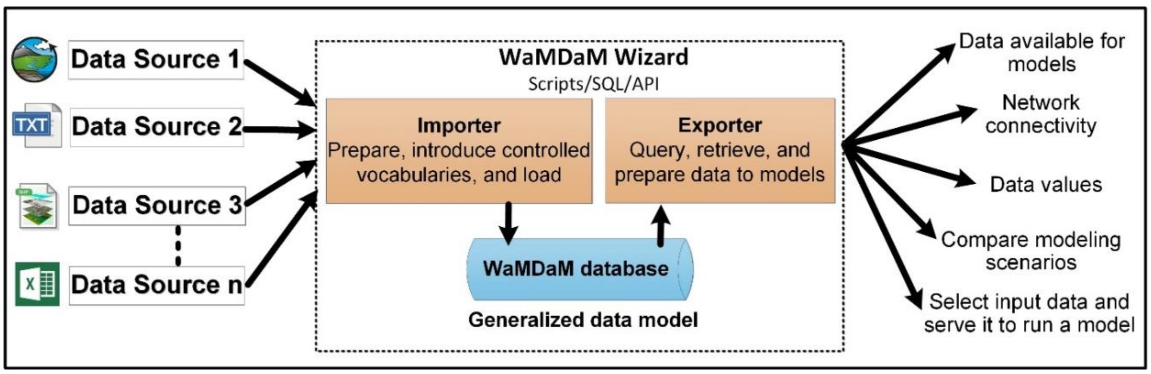
Required hardware and software: The WaMDaM data model can be used within any relational database management system or platform. The WaMDaM Wizard executable (.exe) is available for use with Microsoft Excel (2007 and later) and SQLite3 on Windows 64-bit computers.

Input data and directions: Documentation of all source code, datasets, use cases, and instructions to use WaMDaM and replicate results are available on GitHub and facilitated by Jupyter Notebooks at Abdallah (2019), “WaMDaM Use Cases Repository” Zenodo doi: http://doi.org/10.5281/zenodo.1484581

Programming languages: Python 2.7 and Structured Query Language (SQL)

Cost and license: Free. Software and source-code are released under the New Berkeley Software Distribution (BSD) 3-Clause License, which allows for liberal reuse.

Graphical Abstract



2.1 Introduction

Data analysis and synthesis are fundamental in developing water resources management models (Loucks et al., 2005). Data organization enables or inhibits the analysis that water managers and modelers perform (Horsburgh et al., 2008; Brown et al., 2015). Well organized data can help modelers prepare data for models while poorly organized data can make the process time-consuming and frustrating. Current practices to organize, manipulate, and compare multiple water resources datasets and develop water systems models are typically specific to the data sources, models, and study location (Brown et al., 2015). Source-, model-, and study area-specific practices arise because models have different data requirements for their components, store data in different file formats, have varying spatial and temporal coverage, use inconsistent metadata to describe methods, sources, and units, and use different vocabularies to name similar system components and their attributes (Miller et al., 2004; Laituri and Sternlieb, 2014; Maidment, 2016). These practices limit managers' and modelers' ability to reuse datasets and models in other applications. To reuse, practitioners often spend up to 75% of their overall modeling time to modify, subset, transform, convert, and restructure data (Ridley and Stoker, 2001; Draper et al., 2003; Miller et al., 2004; CUAHSI, 2005; Michener, 2006; Maidment, 2008; Hey et al., 2009; Beniston et al., 2012; Leonard and Duffy, 2013; Watkins, 2013). A common database design to organize and manage water resources system data can help modelers and managers spend less time to wrangle with data formats and structures and more effort on analysis to learn about and model systems.

Water management data describe natural and built water system components like water supply, infrastructure, and demand sites, and these components are typically represented water systems models as networks of nodes and links (Loucks et al., 2005; Rosenberg and Madani, 2014; Brown et al., 2015). Each node and link are described with properties that represent observed values and input data, or variables that store model results. Data can be organized in time series, as seasonal parameters, as multi-variable arrays, or in other types.

In current practice, a water resources system modeler selects a water management modeling method and then searches for input data that meets the model's requirements (Brown et al., 2015). Modelers often manually search for, download, synthesize, and compare data from disparate datasets to populate input data (Rosenberg and Madani, 2014). In their data search, modelers often use a combination of existing methods to manually gather input data for the different supply and demand system components and their connectivity from local, state, and federal agencies. Searches can also use national data services like the Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) Water Data Services (Goodall et al., 2008; Couch et al., 2014). Each dataset has a particular file-format, organizational structure, syntax, and descriptive terminology. Some datasets also come with modeling scenarios that represent changes to values of physical, operational, network topology, or socio-economic attributes of the system. Modelers must reconcile structure and terminology heterogeneities in potential input data.

Many water resources modelers use the U.S. Army Corps of Engineers Hydrologic Engineering Center Data Storage System (HEC-DSS) (HEC, 2009) to store and manage paired variables and time series data. Modelers also use Hydra Platform (Knox et al., 2014) and ArcHydro (Maidment, 2002) for network connectivity. Others may also use the Observations Data Model (ODM) for organizing and storing site-specific time series data (Horsburgh et al., 2008). Other modelers simply organize data into one or many spreadsheets within a Microsoft Excel workbook with consistent column headers (e.g., variables) and units. Still other modelers store data that describe the water system and its operations in proprietary modeling software systems like the Water Evaluation and Planning system (WEAP) (Yates et al., 2005), RiverWare (Zagona et al., 2001), OASIS, ModSim, and others (Wurbs, 1993; Loucks et al., 2005; Wurbs, 2012). Although models like RiverWare (Zagona et al., 2001) and WEAP (Yates et al., 2005) are not strictly used for data management purposes, we consider them data management systems because they contain large amounts of data that describe water systems and house the data used for numerous river basin management studies around the world.

To identify, analyze, or compare water management data stored in one or many of the above systems, modelers often develop source- and model-specific workflows to manipulate, join, pivot, sort, aggregate (in time and/or space), and visualize data. Simultaneously, modelers must keep track of metadata, if present, that describe the source of data, methods used for creating the data, and methods used to transform data to a format appropriate for a particular model. These metadata elements are typically specific to the data source and model. Adding a data source, expanding a study area, or changing the underlying model means the modeler must modify the data preparation workflow or create a new workflow. Modelers then must manually repeat data manipulations and analyses.

Thus, there is a need for a generalized method to more readily and consistently organize, store, join, query, and compare multiple types of water management data and metadata across datasets, models, and study areas (Bajcsy, 2008; Govindaraju et al., 2009; Brown et al., 2015; Vogel et al., 2015). This need arises because of two fundamental data management challenges related to how data is structured (i.e., syntax) and how key data components are named and described (i.e., semantics). An example of different syntaxes is the number and order of headers and rows in a spreadsheet. Examples of different semantics include hydrologic system component names (e.g., “reservoir” versus “storage facility”), attribute names (e.g., “storage” versus “volume”), and system component names (e.g., “Hyrum Reservoir” versus “HYRUM”).

In reviewing more than 40 existing systems to organize water management data (Appendix A, Table A1), we found all systems incompletely support structure and syntax issues. Systems have different and limited capabilities to query and compare multiple datasets and models, no software standards, or no guidelines to organize water management data. Differences include how data is represented in space and time, how data is organized within structures (i.e., data type) (DCMI, 2013), the physical means used to store data (i.e., database, text file, or other formats) (DCMI, 2013), and software technology. The heterogeneity in methods reveals why modelers spend considerable time preparing and transferring data across different models, formats, and technologies.

Several recent efforts to increase data consistency and transparency, such as the Open Water Data Initiative (Blodgett et al., 2016), Observations Data Model 2 (Horsburgh et al., 2016), the Open and Transparent Water Data Act (Dodd, 2016; Cantor et al., 2018), and the Water Data Exchange program (Larsen and Young, 2014) have recommended data standards to integrate fragmented water information data into consistent and interoperable data systems. Such integrations and requests for them aim at improving access to water information to help quantify its availability and use at different scales in the present and future. Here, we contribute a generalized data model called the Water Management Data Model (WaMDaM) to help organize, join, compare, and analyze multiple water resources datasets and models. We also introduce software tools that demonstrate key functionalities of the design. The WaMDaM design helps answer the overarching research question of: how can data from multiple sources be organized and described in a semantically and syntactically consistent way to facilitate data query, comparison, joining, and analysis that will ultimately help modelers choose input data to build and run water resources systems models? A successful WaMDaM database design must have: 1) modular and extensible components, 2) networks of nodes and links, 3) scenarios and version control, 4) reusable contextual metadata, 5) support for multiple data types used by systems models, 6) extensible controlled vocabularies, 7) direct access to subsets of data and metadata, and 8) an open-source environment.

Next, we describe the motivation and design requirements for the WaMDaM system. Section 2.3 presents the WaMDaM data model design and physical implementations. Section 2.4 introduces companion software tools. In Section 2.5, we use WaMDaM to join 13 overlapping local, regional, and national models and datasets. We demonstrate the utility of the data model in five use cases. The use cases help modelers to identify, compare, and select water supply and demand data, connectivity between engineered infrastructure and natural systems components, model scenario data, and serve selected data to a WEAP model for the Bear River Watershed of Utah. Section 2.6 discusses how modelers can use WaMDaM, limitations, future work, and an invitation to use and improve the design. Section 2.7 concludes.

2.2 Design Motivation

WaMDaM focuses on the essential steps to organize, join, compare, analyze, and serve multiple datasets to build a water resources model. Because modelers often use multiple systems to gather, organize, store, join, and query the water management data they need to build models (Figure 2.1-A), they repeat that effort for each new model, data set, scenario, system component, and element. Modelers would benefit from a general approach that only requires doing the work once but allows others to re-use their effort in their other endeavors (Figure 2.1-B). Five use cases guided the WaMDaM design by answering key water management data questions. These use case questions sidestep less important aspects that may overcomplicate the design (Szalay and Blakeley, 2009). The use case questions are:

1. What data entered by others can be used to develop a model in a study area?
2. Which network connectivity should be used in a model?
3. How do data values differ across datasets, and which values should be chosen for a model?
4. How do scenarios differ, and which scenarios should be chosen in a model?
5. How do the input data developed in earlier use cases affect model outputs?

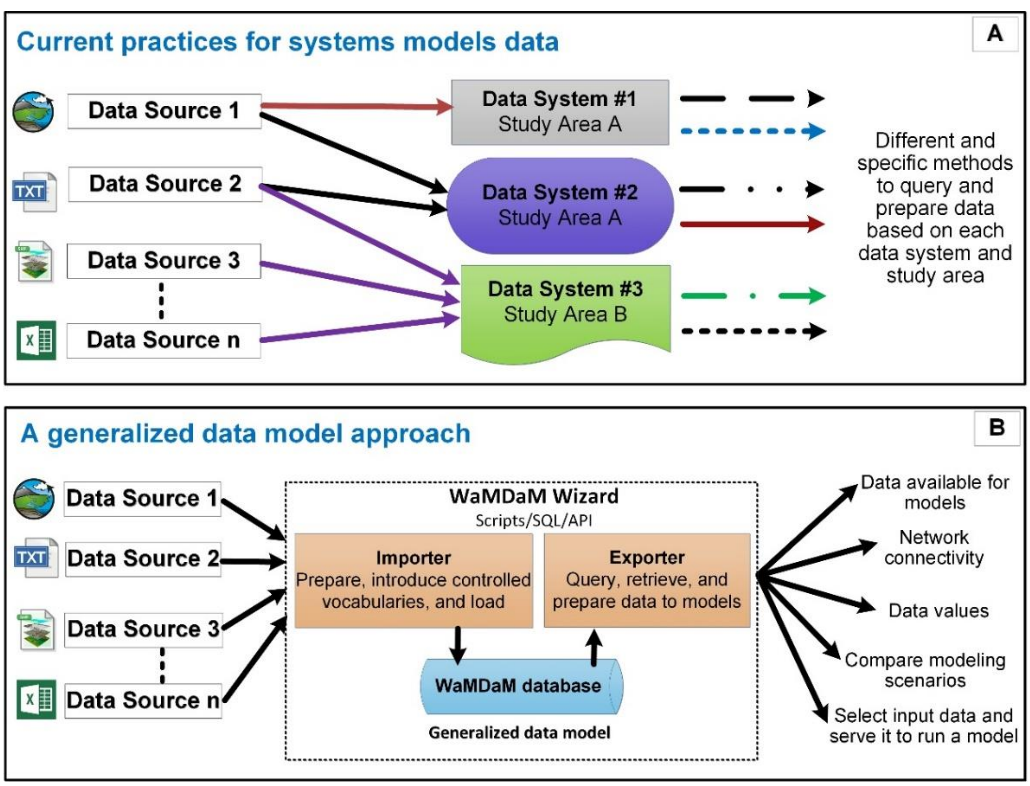


Figure 2.1 (A): Current data practices use different systems and data manipulation methods for each data source and study area while (B) a generalized data model integrates across the structure and syntax of data sources. The WaMDaM Wizard with scripts, SQL, and APIs allow modelers to undertake multiple efforts, such as load data, identify data for models, compare networks, data values, and scenarios, and serve data to models.

2.2.1 Synthesis of design requirements

We synthesized eight design requirements for an integrative data system from 40 prior data management approaches (Appendix A, Table A1). Below, we define each design requirement and then discuss how the functionality that satisfies these requirements improves over prior approaches.

The first requirement for a modular and extensible design will allow inclusion of multiple model types and their system components (e.g., reservoirs, demand sites, canals) as reusable data objects (i.e., as classes or modules) with properties or attributes (Zagona et al., 2001; Connolly and Begg, 2010; Wurbs, 2012; Knox et al., 2014). Attributes may apply to all network components globally or to individual components. For example, a time series of inflow applies to one reservoir component, while a budget parameter applies to a network. To improve storage efficiency and enable consistent reuse of data, the design must be able to share the same value of an attribute across many water resources system components.

Modular and extensible design is supported in most existing data systems and water management models such as Hydra Platform and the ODM (Harou et al., 2010; Knox et al., 2014). Other systems, such as Arc Hydro and WEAP (Maidment, 2002; Yates et al., 2005) allow adding new data objects (as in Arc Hydro), but users are still forced to use core components and attributes that might not be needed for a case study.

The second requirement is to represent the spatial configuration of system components as networks of nodes (junctions or points) and links between nodes (arcs, connections, curves, lines, or edges of a directed graph) (Zeiler, 1999; Rossman, 2000; HydroLogics, 2009). Networks help modelers organize and search for system components that are related in purpose (e.g., flow of water through connected pipes), use (e.g., drinking water supply), or in a spatial boundary (e.g., Bear River Watershed) (Loucks et al., 2005). Networks also represent connectivity, which is a key principle of water mass-balance fundamental to most systems models. Although most existing data systems support networks, each system uses different data organization methods and terms to manage the connectivity of nodes and links. Such different structures require different methods to query network data. While the ODM (Horsburgh et al., 2008) stores time series data for individual nodes or links, ODM was not designed to describe how the nodes relate to each other (upstream, downstream, etc.). A consistent method to represent networks will allow users to consistently retrieve information about how nodes are connected to each other through links.

Third, the data system must describe and store scenarios that represent changes to the physical, operational, infrastructure, and socio-economic model input data. Scenarios allow modelers to test and run current and proposed water management alternatives. The scenario requirement also includes the ability to track and manage versions of changes from a baseline network. A scenario can be created by one or two potential changes to a water system network: i) change network topology like to add or remove an infrastructure component and ii) change data for one or more attributes of a component such as to expand the capacity of a reservoir or update metadata such as the method or data source. Many existing systems (e.g., WEAP) use scenarios to track changes in input data but cannot track changes in the network components.

Fourth, the data system must allow users to add contextual metadata; the additional information to help modelers interpret data. Metadata also helps modelers maintain the data provenance needed to track the history and context of sources, methods, people, and organizations that contributed to create the data (Gray et al., 2005; Pokorný, 2006; Horsburgh et al., 2008; Campbell et al., 2013; DCMI, 2013; Carata et al., 2014; Goodman et al., 2014). Some existing systems store metadata in one table that accepts user-specified key-value metadata pairs (e.g., (Refsgaard et al., 2005; Knox et al., 2014). HEC-DSS manages and retrieves large sequential datasets, such as time series and paired tabular data. Support to describe each time series is limited to six metadata parameters that include the variable name, location, and time step. Each parameter must be described in less than 80 characters (HEC, 2009). The ODM uses contextual metadata to describe units, sources, and methods for collecting observational data variables at a site. This requirement mandates explicit support for the following fundamental metadata elements the unit, source, method, people, and their organization that contributed to creating data. The support to explicit metadata elements guides users to populate, reuse, and later to directly query them.

Fifth, the data system must be able to store and describe multiple data types that modelers use to represent physical, operational, and descriptive attributes of system components: time series, multi-attribute series (e.g., multi-variable for a reservoir bathymetry), numeric, categorical values (e.g., gate open or closed), and seasonal parameters (e.g., values that are the same for months across the years). Many existing systems support multiple data types, but store them as binary data objects, which limits users' ability to access stored data outside the software system (Harou et al., 2010; Knox et al., 2014). Supporting multiple data types allows modelers to store, access, and reuse different types of data for properties of water systems components.

Sixth, the data system must support controlled vocabularies (CVs) as sets of terms with definitions for object types, attributes, and names of nodes and links. CVs allow modelers to retain the native terms they are familiar with but simultaneously relate native terms to consistent names that can be reused across datasets and models (Laniak et al., 2013). For example, the following native terms are related to a single CV term (e.g., Reservoir): reservoir (WEAP), storage reservoir (RiverWare), Reservoir Node (Bear River Systems Dynamic Model), reservoir (US Bureau of Reclamation). The CV term then links all the fundamentally similar native terms together. Thus, a query for “Reservoir” returns all related native terms.

Seventh, the data system must support direct access to subsets of data and metadata that enable search and filtering based on a schema. In contrast, unstructured data storage known as the Binary Large OBject (BLOB) formats (Sears et al., 2006) do not allow direct access to subsets of stored values but rather to the entire block of data. Although storing BLOB data such as blocks of time series or arrays as in Hydra Platform and HEC-DSS (HEC, 2009) can be efficient and fast, users must use custom functions to decode and access subsets of the content. In a structured data storage, modelers can load and retrieve subsets of data based on selected water system components, attributes, metadata, networks, scenarios, and data types in space and time without being limited to a custom method.

The eighth requirement is to develop the WaMDaM implementations using free and open-source software tools, to allow access via an open-source code repository, promote reproducibility, and help others further advance the method (Easterbrook, 2014; Goodman et al., 2014; Gil et al., 2016). At the same time, we recognize that open-source software requires documentation to be reusable. Many existing data systems like WEAP, RiverWare, and HEC-DSS are proprietary and require specific tools to access their data. Those proprietary approaches contrast with other customized systems models that use a mix of spreadsheets, text files, and the General Algebraic Modeling System (GAMS) file formats to organize their data and metadata.

2.2.2 Support for Design Features

To date, existing water resources systems software tools incompletely support the eight requirements (Table 2.1). Thus, we designed WaMDaM to support all eight requirements. The next section describes how WaMDaM was designed and implemented to support the eight requirements, answer four use case questions, and complete a fifth use case that serves data to a model.

Table 2.1***:*** Support for the identified requirements by select data systems and water resources models. An "X" indicates that the system supports the requirement.

	Select Data System / Model					
Data Management Requirement	ODM	Hydra Platform	HEC-DSS	ArcHydro	RiverWare	WEAP
Modular and extensible design	X	X				
Supports networks of nodes & links		X		X	X	X
Supports scenarios & version control		X	X		X	X
Reusable contextual metadata	X					
Multiple data types for system models		X	X		X	X
Extensible controlled vocabularies	X					
Direct access to subsets of data	X			X		
Open-source environment & license	X	X				

2.3 WaMDaM Design

We used the eight requirements described in Section 2.2 to design the WaMDaM data model and its physical implementations to organize, manage, join, query, and compare water resources datasets and models. We aimed for a parsimonious design that minimizes the number of data and metadata entities needed to satisfy the eight requirements and answer the use case questions (Hey et al., 2009). The criteria for a successful design was a design that satisfies the eight requirements and answers the use case questions. Below we present the conceptual design, then show the logical design using an Entity Relationship Modeling (ERM) diagram. Afterwards, we describe physical implementations.

2.3.1 WaMDaM Conceptual Design

The WaMDaM conceptual design has multiple, hierarchal one-to-many relationships; color-coded grouped entities represent key design requirements (Figure 2.2). In general, the color-coded groups define the steps a modeler would follow to populate a physical implementation of the design with data.

The first group of blue entities supports a modular and extensible design by allowing the modeler to define the resource type (e.g., a WEAP model), one or many object types (e.g., reservoir, river reach, diversion, etc.) for each resource type, and one or many attributes (e.g., storage or diversion capacity, head flow, etc.) for each object type (Requirement #1). A resource type represents the types of data (input or output) used in a data provider such as a “Model Program” as defined in Morsy et al. (2017), independent of implementation. For example, a WEAP model resource type has 21 object types (e.g., reservoir, demand site, transmission link, etc.), and each object type has many attributes (e.g., “Storage Capacity”, “Net Evaporation”). The resource type entity can also be used for datasets. For example, the U.S. Major Dams Inventory shapefile has a list of 18 attributes that have values for the “Dam” object type. An object type is a system component with typologies such as node or link (e.g., reservoir, canal, water source, or demand site) and can have one or more quantitative or qualitative properties or attributes with units.

The second group of green entities supports networks and scenarios by allowing modelers to define a master network with many scenarios where each scenario can have one or many instances that are either node or links (Requirements #2 and #3). To specify connectivity among instances, links must have start and end nodes.

The third group of orange entities allows modelers to use reusable, contextual metadata where a modeler affiliates people to an organization and specifies methods and sources that generate data (Requirement #4). The fourth group of red entities allows modelers to store seven distinct types of data values such as time series or categorical data (Requirement #5). Within a scenario, an attribute for an instance has a source, method, and data type. The fifth group of controlled vocabulary (purple) entities allows modelers to relate native terms for object types, attributes, and instances (Requirement #6).

We satisfied direct access to all data and metadata (Requirement #7) by using relational database theory (also referred to as the Relational Model) to implement the data model entities as interrelated tables (Codd, 1970; Chen, 1976) as further described in Section 2.3.2. We developed a physical implementation of the data model and software tools in an open-source physical database system (Requirement #8; see Section 2.3.3). Next, we explain how and why the relationships are implemented to form the WaMDaM Logical Data Model.

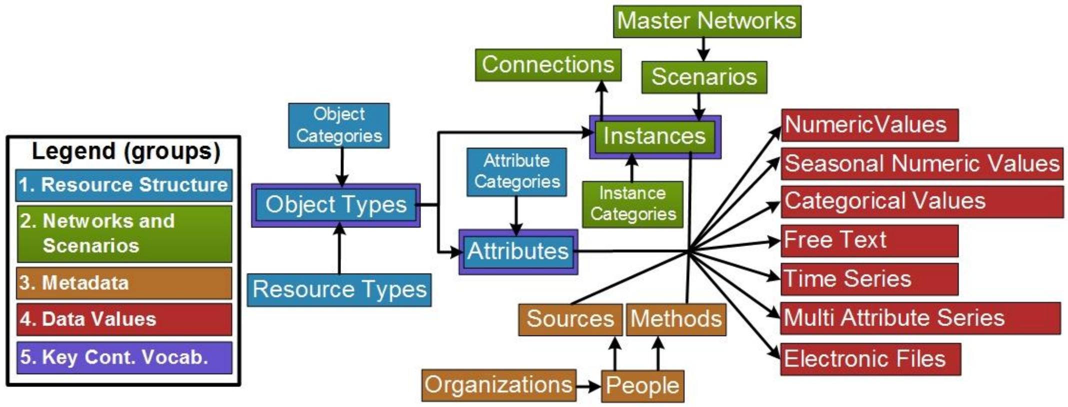


Figure 2.2: The conceptual diagram relating the first six design requirements for the water management data model. Key controlled vocabularies are introduced to the boxes outlined in purple.

2.3.2 WaMDaM Logical Data Model

The Logical Data Model schema shows the one-to-one, one-to-many, and many-to-one relationships among database entities (Figure 2.3). Blue, green, orange, red, and purple colors again indicate tables associated with the resource type, networks and scenarios, metadata, data values, and controlled vocabulary design requirements, respectively. A WaMDaM data value is described by fourteen required elements (Appendix A, Table A2). Here we describe six key requirements that are needed to interconnect schema components and specify the fourteen required elements and design requirements. We pluralize data model entities and list them in italics and capital letters.

First, *ResourceTypes* are datasets (like the U.S. Major Dams Inventory) or models (like WEAP) and have one or more system components called *ObjectTypes* (such as a reservoir, canal, water source, or demand site). *ObjectTypes* have typologies such as node or link and one or more quantitative or qualitative properties called *Attributes* (such as storage capacity, net evaporation, or delivery target). Here we use the broad term attribute, as a contextual property which also may include variables that are measured and might change with time (Sarle, 1995). Attributes could also describe model outputs. Each attribute has a unit, attribute data type, and by choice whether it is used as “Input” or “Output” in a water resources model.

Second, an object type such as a “Reservoir” can be specified (i.e., implemented) for zero or more locations as *Instances* (e.g., Hyrum Reservoir, Bear Lake, and Flaming Gorge Reservoir would be three separate reservoir instances). An instance inherits the *Attributes* of its object type and may be geo-referenced as a node in space with longitude and latitude coordinates. Instances can also be a link which has start and end nodes. The *Connections* entity specifies a start and end node for links and avoids a circular reference problem when connecting the *ObjectTypes* table directly to all the *Instances*, *Attributes*, and *ValuesMapper* tables. A circular reference in a database is problematic to database integrity as it may allow multiple transaction paths to insert or delete data. In the data systems modelers may represent the same water system component, such as reservoir, as a node or a link in a model. Thus, storing nodes and links in the *Instances* table and link connectivity info in the *Connections* table enables modelers to use the same query to access data for nodes or links and improves over prior approaches that require many different queries to access data for node or links (Yates et al., 2005; Abdallah and Rosenberg, 2014; Knox et al., 2014).

Third, one or more node and link *Instances* can be connected into *MasterNetworks* (e.g., water supply/demand, water distribution, or other network for a study area). Each master network contains one or many *Scenarios* in a study area (such as a base case, reduced inflow, or new infrastructure). *Scenarios* within the same master network may share the same exact network topology or versions of the network and its data. Each scenario also has a start and end date and time step to track the modeling time step and its extent.

Fourth, the *Mappings* bridge entity relates *Instances* to their *ObjectTypes*, *Attributes*, metadata *Sources* and *Methods*, *Scenarios*, and data values. This bridge entity is the central table in the WaMDaM database. This *Mappings* entity is needed because *ObjectTypes* can have i) many *Attributes* (e.g., reservoir object type can have evaporation depth, storage capacity, and volume-area attributes), ii) each *Instance* (e.g., Hyrum Reservoir, Bear Lake, or Flaming Gorge Reservoir) can have shared or instance-specific attribute values, and iii) *Instances* can also have shared or instance-specific *Sources* and *Methods* metadata values.

Fifth, data values are assigned to one of seven supported data types and connected through the *ValuesMapper* entity to the *Mappings* bridge entity. The seven supported data types (numeric, seasonal, categorical, free text, time series, multi-attribute series, electronic file) are commonly used in the models we reviewed (Appendix A, Table A3*)*. Similar to prior time-series data models such as ODM, the *TimeSeries* entity (e.g., flow versus time) captures key global metadata for the entire time series and can have one or many values, time stamps, aggregation statistics (e.g., average, cumulative, etc.), and year types to indicate water year or calendar year. The *MultiAttributeSeries* entity organizes paired data (e.g., area-elevation curve) by referencing multiple *Attributes*. Each paired attribute has one or many values and sequential order to preserve the order and pairing of values across many attributes within the same array. Additional attribute data types can be added and connected to the *ValuesMapper* entity without affecting any of the existing data model relations. The *ValuesMapper* entity helps to reuse and share attribute data across many *Instances* (Requirement #5). This WaMDaM approach of storing values once and sharing them is more efficient and allows the option to register the term one time with a controlled vocabulary. Sixth, the *ScenarioMappings* bridge entity further allows modelers to share similar *Instance*s, their *Attributes*, metadata, and values across *Scenarios* with no duplication. The WaMDaM Wizard, presented later in Section 2.4, also uses the *ScenarioMappings* bridge entity to query and compare how combinations of *Instances*, their *Attributes*, and data tables change between two *Scenarios* within the same master network. Seventh, *People, Organizations, Sources,* and *Methods* support four essential key metadata entities needed to interpret *Instances* and values. The *Sources* entity describes the origin or encompassing package of data such as a shapefile, web service, or a model for a study area which may have a citation and a webpage. The *Methods* entity describes how values were created, an instance is defined, data quality, and the resource type works (e.g., simulation or optimization method for a model program). Modelers may document uncertainty in the data and indicate the quality of data within the method that generated it. Each source or method is associated with a person (author) who set up the source or created the method. Each person belongs to an organization. If no person is associated with data, modelers can define a person as “unknown” and relate to the organization that created the source or method. We recognize that there is potential for a more complex and specific representation of metadata. We attempted to balance between the principles and practicality of metadata usage as recommended by Duval et al. (2002). Complex metadata requirements may discourage modelers to provide metadata while too little metadata might be insufficient to correctly interpret data. Modelers are required to provide the native unit name for each attribute and are encouraged to relate the unit with a list of controlled units. Using controlled unit vocabularies allows the user to convert values into other units. Eighth, controlled vocabularies have the following common fields of term, name, category, definition, and URL to a source. This approach is the same as the CVs defined for ODM2 (Horsburgh et al., 2016). The key CVs attach to *Object Types*, *Attributes*, and *Instances* to relate native terms and values across *Resource Types*. Each resource type (e.g., model) has its own native terms. Data of different models can be related using three controlled terms, object type (e.g., Reservoir), attribute name (e.g., Volume), and instance name (e.g., Hyrum) (Figure 2.4). Units can be converted using constant or linear multipliers. For example, a value of 1.000 liter has a 0.001 constant fraction in reference to a 1.0 cubic meter volume unit. We adopted the list of controlled units from Hydra Platform (Knox, 2018). Finally, software business rules (i.e., external code) are used to correctly enforce some of the complex relationships in the data model, especially when loading data into the database. For example, software business rules relate an object type and its typology with *Instances* through a dummy attribute and ensure that each link in the *Connections* entity has a start and end node. Another rule relates a resource type with a master network through the “NetworkAttributes” object type, the dummy attribute, and a dummy instance to allow modelers to query all the network implementations of a resource type. Correctly representing the many-to-many relationships among the entities within the first six design requirements while attempting to achieve parsimony and relatively simple querying consumed a significant portion of the iterative WaMDaM designs. We summarize the software business rules on GitHub (Abdallah, 2018).

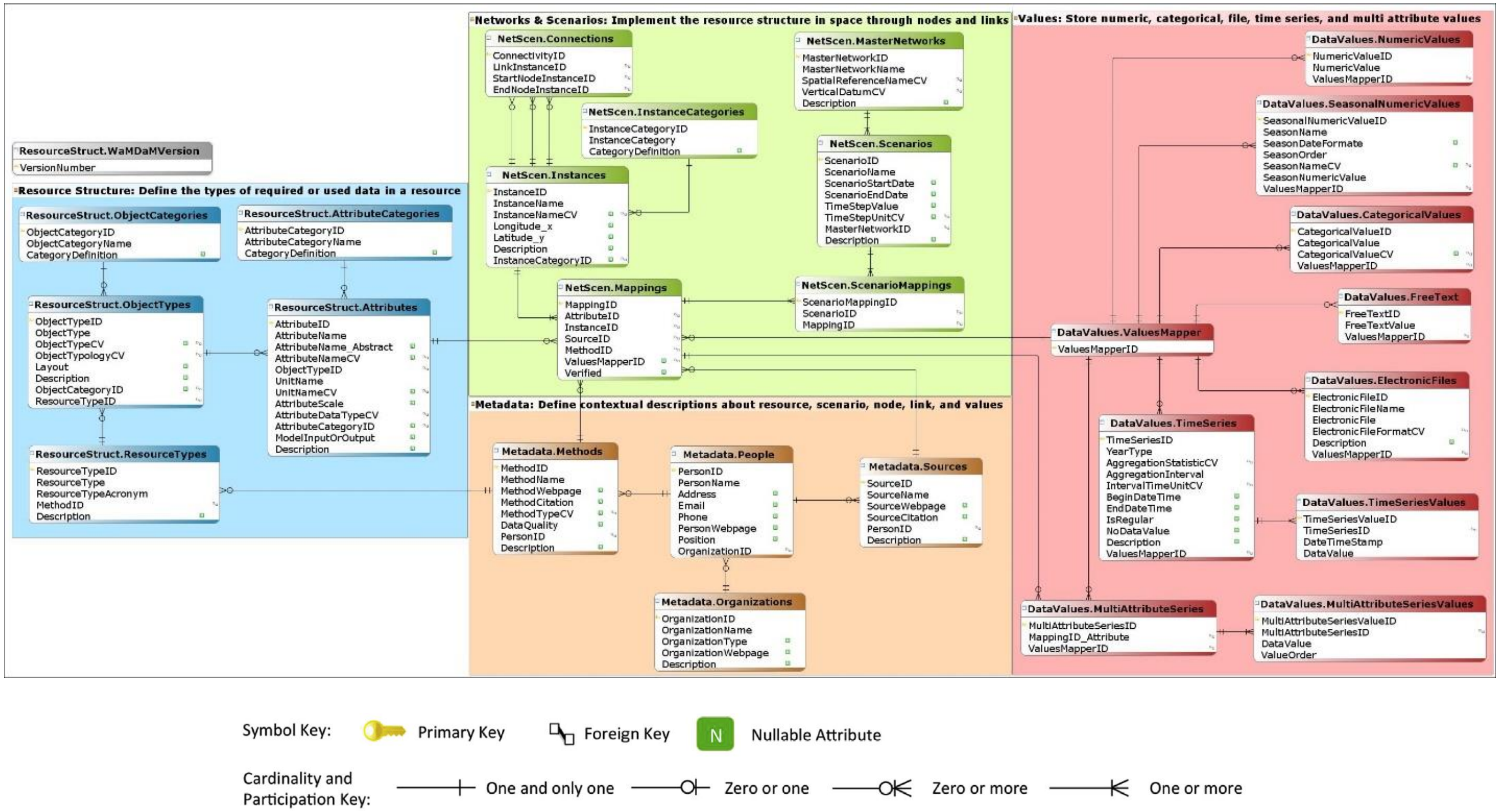


Figure 2.3: WaMDaM logical model tables grouped into the design requirements. Resource Type (Req.#1), Networks and Scenarios (Req.#2&3), Metadata (Req.#4), and Data Values (Req.#5). The diagram uses the crow's foot notation for relationship cardinality and participation. An interactive html copy is available at http://schema.wamdam.org/diagrams/01_WaMDaM.html (Abdallah, 2018). Controlled vocabulary tables (Req.#6) are not shown here for simplicity and can be viewed at http://schema.wamdam.org/diagrams/03_CVs.html. Each column name (field) that ends with "CV" indicates that the term is a controlled vocabulary.

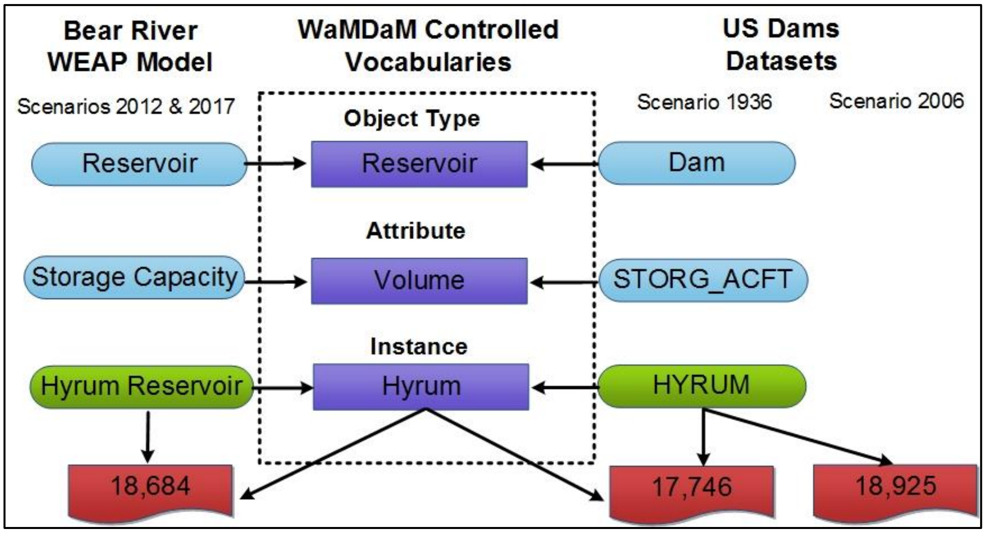


Figure 2.40: Relating native names with controlled vocabularies for object types, attributes, and instance names allows modelers to query and simultaneously access values across native terms. Identical storage is shared among scenarios of the Bear River WEAP Model while values in the US Dams Datasets are stored separately.

2.3.3 Physical Model Implementation

We implemented the logical data model schema within four physical Relational Database Management Systems (RDBMS), including PostgreSQL, MySQL, Microsoft SQL Server, and SQLite to demonstrate that WaMDaM is independent of the RDBMS (Abdallah, 2018b).

First, we selected a physical data type for each field in each logical model entity (e.g., integer, varchar) and we imposed physical constraints on each field (e.g., value cannot be null) by following the physical data types convention in the ODM2 (Horsburgh et al., 2016). Second, we adapted an existing Python 2.7 script developed by Horsburgh et al. (2016) to forward engineer the DBWrench schema file into a Data Definition Language (DDL) script containing a set of “create” statements from the for WaMDaM tables for each of the four RDBMS. Finally, we executed each of the DDL script within each RDBMS to create a physical blank WaMDaM database that modelers can load with data.

We chose to express the logical data model as a relational model to: i) support direct access to all data and metadata (Requirement #7), ii) be platform independent and implement as open-source on different operating systems for different relational database systems (Requirement #8), iii) support a standardized and stable Structured Query Language (SQL), and iv) follow common use and familiarity with the RDBMS within the water resources community (Horsburgh et al., 2008; Harou et al., 2010; Knox et al., 2014; Horsburgh et al., 2016).

The core contribution of WaMDaM is the description of a generalized design to help organize, compare, and analyze multiple water resources datasets and models. Our implementation in a relational database is just one way to solve the problem. Other methods, such as non-relational databases, also known as NoSQL, are increasingly used worldwide (Hoberman, 2014) and could likely satisfy the same use cases. NoSQL implementations may scale and adapt without being limited to a schema. Future work should test WaMDaM's ability to scale and adapt to much bigger and more diverse datasets and models.

2.3.4 Community Feedback on the Design

We iteratively revised this data model design in five key versions over the course of five years to satisfy the design requirements and use cases. The changes were in response to feedback from collaborators at the University of Manchester, University of California, Davis, and University of Massachusetts, Amherst on WaMDaM design and tools. We acknowledge the need for larger and more diverse community testing and feedback to serve a wider audience of users. We also incorporated feedback on an earlier design and its description (Abdallah and Rosenberg, 2014). The five key designs are available on GitHub (Abdallah, 2018b)

2.4 WaMDaM Related Software

We created software tools to demonstrate WaMDaM's functionality and allow users to more easily interact with its database.

2.4.1 WaMDaM Wizard

We developed a WaMDaM Wizard (hereafter the Wizard) in Python 2.7 for SQLite as a simplified demonstration to auto-read input data from an Excel Workbook template into a physical WaMDaM database implementation on the user's local machine (Abdallah, 2018c). The WaMDaM Wizard uses SQL Alchemy (https://www.sqlalchemy.org/) to load data into the database, and we use direct SQL scripts to query the database through a Python SQLite3 (https://www.sqlite.org) library. The Wizard provides key functionalities of the design and it is just one of many possible ways to import or export data of the database. We chose Microsoft Excel as a generic input data medium because modelers commonly use it. The Wizard validates entries to comply with the database schema, maps primary and foreign keys, and implements software business rules.

We elected to use SQLite (https://www.sqlite.org/index.html) because it is free, open-source, and server-less to satisfy open-source design (Requirement #8). We also used the DB Browser for SQLite (https://sqlitebrowser.org/) as an open-source user interface to view and execute queries against WaMDaM database tables.

The Wizard has tools to i) prepare and pivot a shapefile, time series, or seasonal data into the data structure of the workbook template, ii) import time series stream flow data from WaterOneFlow CUAHSI web-services, iii) import time-series WaterML files for reservoir inflow, release, storage, elevation from the U.S. Bureau of Reclamation (USBOR) Water Information System web service (https://water.usbr.gov/), iv) import network and data stored in WEAP using its Application Programming Interface (API) into the workbook template, v) use the provided controlled vocabularies in the workbook to register and relate native terms across sources as discussed in Section 4.2 , vi) adapt and use the example Jupyter Notebooks of Python scripts to execute data query, plots, and analysis across data sources, and serve data into the model, and vii) compare and verify differences in topology or input data values across modeling scenarios.

2.4.2 Controlled vocabulary registry

We deployed an online-hosted CVs system to physically implement the CVs design (Requirement # 6), allow multiple modelers to access, reuse, or suggest new consistent vocabularies across WaMDaM database instances and machines. We adapted the existing online CV registry system which is a Python/Django web application API developed by the ODM2 design team (Horsburgh et al., 2014; Horsburgh et al., 2016) to manage WaMDaM CVs (Abdallah, 2018a) (http://vocabulary.wamdam.org).

Because we adopted the CVs moderation system developed by the ODM2 team, modelers have the option to use WaMDaM CVs, submit suggestions to add new terms within the online registry, or use their own native terms without registering them with the WaMDaM controlled vocabulary. We populated the CVs system with example WaMDaM

CVs for the datasets we worked with and introduce in the next Section. Modelers can use the CVs system seamlessly in an Excel Workbook template and the WaMDaM Wizard. Within the Excel Workbook template, there is Visual Basic script button that downloads and updates look-up menus for all CVs. Excel sheets in the Workbook template contain a column for the native term and another as a controlled look-up term that register or relates them together. To get all the native terms registered to a controlled term, modelers can write a simple query against their local WaMDaM database.

2.5 Results

We present five use cases that demonstrate how WaMDaM and the software tools we developed can assist modelers to: i) identify specific input data to expand a model to a larger study area from previously-entered datasets in a WaMDaM database, ii) show the spatial configuration and network connectivity of natural and engineered system components, iii) compare retrieved data to help the user decide which data to use, and iv) compare changes in network topology, metadata, and data values among scenarios. These use cases also support a final common case to v) serve selected data to run an example WEAP model. These five use cases support common operations that water resources systems analysts and modelers perform to develop and use models.

The use cases apply one optimization and two priority-based simulation models for the Bear River study area: 1) the Watershed Area of Suitable Habitat (WASH) model that allocates water to maximize watershed habitat areas (Alafifi and Rosenberg, 2020), 2) the Bear River Systems Dynamic Model (BRSDM) (Sehlke and Jacobson, 2005), and 3)

WEAP model. These use cases expand modeling coverage for the Lower Bear River to more of the Watershed in Utah, Idaho, and Wyoming (light red to darker red in Figure 2.5**).**

The use cases assume a modeler used WaMDaM CVs, Excel templates, and the WaMDaM Data Wizard to load 13 diverse and overlapping U.S. national, regional, and local data sources and models (Table 2.2) into a WaMDaM SQLite database. The database file is 35 Megabytes with 73 *ObjectTypes*, 563 *Attributes*, 15,464 *Instances*, and 214,352 rows in the central *Mappings* table. Readers can use the instructions and Python 2.7 scripts in Jupyter Notebooks (Abdallah, 2020) to load data into the database and replicate queries and figures as well.

Table 2.2: Data sources used in WaMDaM use cases

#	Data Source	Instances (#)	File Format
1	Water Data Exchange (WaDE) Program of the Western States Water Council http://wade.westernstateswater.org/	2	Excel, (Web-service for time series is in progress)
2	WaterOneFlow Web Services (CUAHSI) http://his.cuahsi.org/wofws.html	1	Web-service, WaterML
3	U.S. Bureau of Reclamation Water Information system web service https://water.usbr.gov	2	Web-service, WaterML
4	US Hydropower Dataset (Samu et al., 2017)	2,398	Excel (.xlsx), Shapefile
5	US Major Dams Dataset (U.S. Geological Survey, 2013)	8,121	Shapefile, text files, HTML
6	Bear River Commission Flows (Personal Communications, 2016)	1	Excel (.xlsx, .xls), Quattro Pro (.QPW)
7	Utah Dams Dataset (Craig Miller-Personal Communications, 2016)	910	Shapefile, Excel (.xlsx)
8	Utah Flows Dataset (Craig Miller -Personal Communications, 2016)	893	Shapefile, text file
9	Idaho Flows Dataset (Liz Cresto-Personal Communications, 2016)	164	Shapefile, Excel
10	Watershed Area of Suitable Habitat model (WASH) (Alafifi and Rosenberg, 2020)	104	Excel (.xlsx), shapefile
11	Bear River systems Dynamics Model (BRSDM) (Sehlke and Jacobson, 2005)	237	Excel (.xls)
12	Bear River WEAP Model 2012 for Utah (Rosenberg, 2017)	375	CSV, Paradox Database, shapefile
13	Bear River WEAP Model 2017 for Utah and Idaho (Rosenberg, 2017)	150	CSV, Paradox Database, shapefile

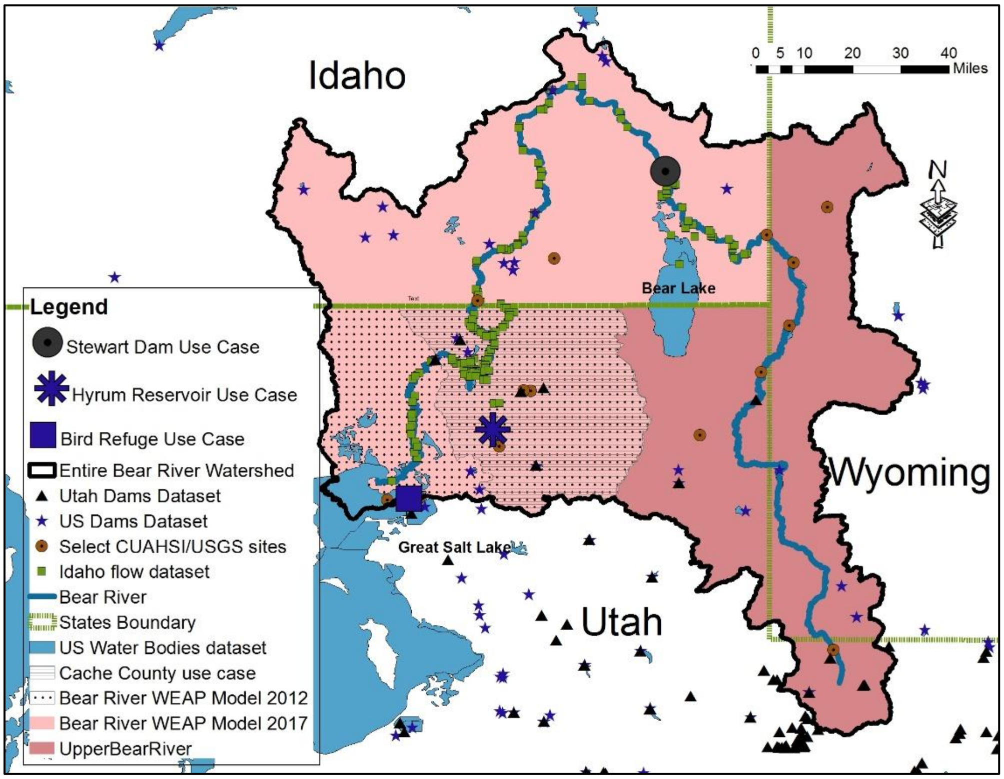


Figure 2.5: The Bear River Watershed in the western U.S. The dotted area shows the spatial domain of existing WEAP 2012 and WASH models for the Lower Bear River Watershed. Lighter red is area for the WEAP 2017 model and dark red is for the Upper Bear River Watershed. Symbols show examples available data.

Use Case 1: What data entered by others can be used to develop a WEAP water supply/demand model for the entire Bear River Watershed?

Using the populated instance of the WaMDaM database file, the user first specifies the resource type to search data (e.g., for the WEAP model) and min and max longitudes and latitudes of the Upper Bear River Watershed (dark red in Figure 2.5). Next, the user runs the SQL script to identify the available object types and attributes. WaMDaM uses CVs to match native WEAP terms with terms from the other 13 loaded data sources. The workflow is readily repeated for a second resource type like the WASH model. By excluding categories of water quality and cost attributes that are not used in the WEAP 2017 model, the WEAP model has 21 object types with 71 attributes, while the WASH model has six object types with 61 attributes.

WaMDaM found six data sources can provide data for the Upper Bear River Watershed for five WEAP object types and 15 of their attributes (out of 71 needed attributes; Table 2.3). Here, WaMDaM used the Reservoir CV term to mediate between the 13 datasets to return the local native terms "Dam" from the U.S. Dams Dataset and "Reservoir Node" from the BRSDM model. Similarly, the controlled attribute term Volume returns "STORG_ACFT" in the US Major Dam's Dataset, "Capacity" in the Utah Dams Dataset, and "Max Storage Capacity" in the BRSDM model for the WEAP attribute "Storage Capacity". To expand the Lower Bear WASH Model, WaMDaM finds six data sources can provide data for six attributes for demand site and reservoir object types. Data is still needed for 55 attributes. One reason for this mismatch is that the WASH model uses many ecologic parameters that do not have analogues in the other data sources.

This use case demonstrates that the same WaMDaM data search method can be applied to multiple models. Loading more diverse datasets into WaMDaM, such as water right priority to demand sites that are required by WEAP, would allow WaMDaM to identify more data for models.

Use Case 2: Which network connectivity should be used in a model?

After identifying types of data that describe water systems components, modelers must determine how water supply, demand, and other system components are connected to correctly represent modeled system components. Here, CVs, node connectivity, and links help modelers visualize network connectivity and select an appropriate network for a model scenario. We focus the use case on Hyrum Reservoir, which is located on the Little Bear River in Utah.

Table 2.3: Summary of the identified attributes and node and link instances in WaMDaM database to expand the Bear River WEAP Model 2017 to the entire Bear River Watershed.

Object Types	**WEAP Attributes with Data**	**Instances (#)**	**Resource Type**
Reservoir	Inflow, Initial Storage, Max. Turbine Flow, Net Evaporation, Observed, Volume, Storage Capacity, Top of Inactive, Volume Elevation Curve	SULPHUR CREEK, Woodruff Narrows Reservoir, Node 2.02, Node 6.01, Neponset Reservoir, …, Whitney Reservoir (34)	US Dams, Utah Dams, BRSDM
Demand site	Annual Activity Level, Annual Water Use Rate, Consumption, Monthly Demand	Node 1.02, Node 1.02, Bear River Watershed ag, Bear River Watershed I, Bear River Watershed M (4)	WaDE and BRSDM
Flow Requirement	Minimum Flow Requirement	Node 1.02 (1)	BRSDM
Gauge streamflow	Streamflow Data	BEAR RIVER AT BORDER, WY, BEAR RIVER NEAR UTAH-WYOMING STATE LINE (2)	Idaho Flows dataset, CUAHSI
Transmission link	Maximum Flow Volume	NUFFER, RIGBY, SORENSEN, WILLIAMSON (JENSEN) (4)	Idaho Flows dataset

We used SQL to query all links connected to Hyrum Reservoir in the WaMDaM database and then sort them by data source (i.e., model). Next, we used Microsoft Visio to draw query results which show Hyrum Reservoir supplies two demand sites in the Bear River WEAP Model 2012 (Figure 2.6-A) and three different demand sites in each of the Bear River WEAP Model 2017 and WASH models (Figure 2.6-B,C). The latter two models also return flow back to Hyrum Reservoir. The WASH Model has the same schematic as the Bear River WEAP Model 2017 model but uses different labels for its nodes and links (Figure 2.6-C). Using its source and methods metadata, the Bear River WEAP Model 2017 model in this area seems to be the most updated and detailed network, so we recommend using the Bear River WEAP Model 2017 model to expand coverage to the Upper Bear River (Figure 2.6-B).

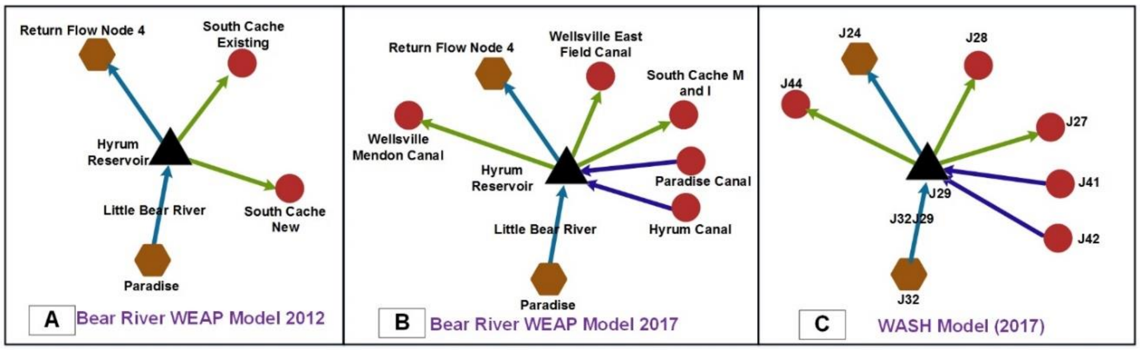


Figure 2.6: Node-link schematics for flows entering/leaving Hyrum Reservoir for three models in the Lower Bear River Watershed, Utah. Arrows indicate direction of flow. Nodes and links with the same color and shape belong to same controlled object type across models.

Use Case 3: How do data values differ across datasets and which value to choose for a model?

Once modelers have identified the types of data available for a modeling study and the model network, they must choose the data sources and values to use for network components. Here, WaMDaM's multiple attribute data types (e.g., time series, seasonal parameters), CVs, direct access, and metadata design requirements can help modelers compare datasets, put context to values, and select the appropriate value for a modeling application. We next illustrate this process using a subset of the data identified in the first use case for 1) time series and seasonal streamflow below Stewart Dam, Idaho, 2) water use in Cache Valley, Utah, and 3) storage elevation curves (i.e., bathymetry) for Hyrum Reservoir in Utah.

Use Case 3.1: What water supply flow values should a modeler choose at a site (e.g., below Steward Dam)?

Reusing the query for use case 1, controlled vocabulary for the instance and attribute names, and shifting the water year time reference, we identified four data sources with flow data for the site below Stewart Dam in Idaho. The datasets are the USGS, the Utah Division of Water Resources (UDWR), Idaho Department of Water Resources (IDWR), and the Bear River Commission (Figure 2.7-A). We used a second SQL query to aggregate and convert all the time series datasets into a comparable cumulative monthly flow in acre-feet per calendar year. The query used the time series metadata of attribute unit, year type, aggregation statistic, and aggregation interval to automate conversions. The four resulting traces span 92 years from 1923-2015 and show data values from the four sources are typically identical except for a few discrepancies in 1996 and 1999 (circles in Figure 2.7-B). The source and methods metadata show that the data originates from stream gage data collected by the PacifiCorp power company. PacifiCorp shares raw data (not available to the authors) with each state. The states interpolate missing data points. We recommend using the UDWR dataset which has the longest available record and documented metadata.

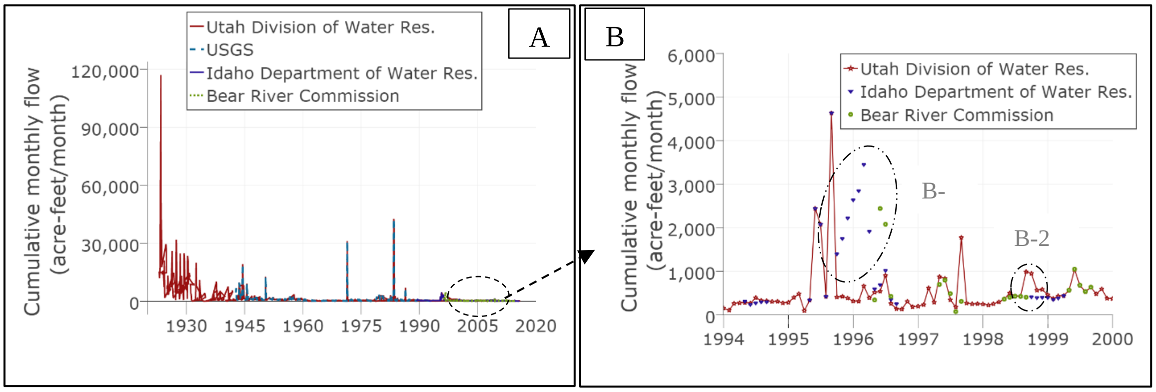


Figure 2.7: Compiled time series data of flow below Stewart Dam, Idaho reported by different agencies over time. (A) 1923 to 2015 and (B) a six-year window that highlights similarities and discrepancies (B-1 and B2) among sources after converting the water year into calendar year.

Water management models like WEAP also use seasonal (i.e., average monthly) flow data, and modelers need to choose appropriate datasets for them. The same query above also returned seasonal data from a fifth source, the BRSDM model, which has three scenarios for monthly flow (dry, normal, and wet) for the same Stewart Dam site (Figure 2.8-A). The BRSDM materials did not document how seasonal monthly values were derived. However, by comparing seasonal values to June high flow values (UDWR data from 1923 to 2015), we estimated the observed flow is lower 48% of the time than the dry June flow value of 666 acre-ft/month. We also found the observed flow is higher about 5% of the time than the wet June seasonal flow value of 17,187 acre-ft/month (Figure 2.8-B). These BRSDM model flow values do not capture dry and wet seasons evenly. Thus, we recommend that systems modelers in this study area use newly derived and more representative flow-frequencies from the UDWR dataset like the 5, 50, 95 percentiles which are 184, 702, and 24,900 acre-ft/month for dry, normal, and wet June months.

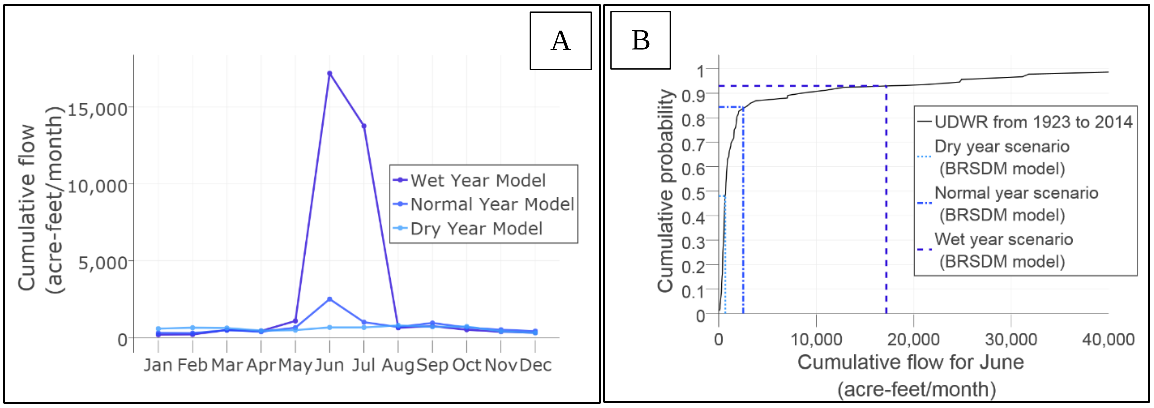


Figure 2.8: Relating dry, normal, and wet year scenario flows below Stewart Dam, Idaho in BRSDM model (A) to cumulative distribution defined by 91 years of UDWR flow records (B).

Use Case 3.2: What agriculture water use data should a modeler choose for a demand site?

Systems models often require data for agriculture, and other water uses, which might be derived or estimated. Here, we use CVs, metadata, and multiple attribute data types to query, aggregate, and compare multiple resource types (data sources) for agriculture water use in Cache County in the Lower Bear River, Utah and recommend data to use in a WEAP model. The query used the controlled term "diverted flow" and returned data from three datasets: WASH model scenarios, WEAP model scenarios, and the WaDE web-service source. The Bear River WEAP Model 2017 uses seasonal demand data for eight sites and annual demand for two sites. Besides the diverted flow-controlled term, using another controlled term, called "depleted flow", returned a fifth time series form the WaDE source which distinguishes the types of demand (dashed line in Figure 2.9).

We used the source and method descriptions for attributes, node instances, and scenarios to identify how the data sources represent water use in spatial and time extents. Data either represent i) the entire county area annually in one node as diverted or depleted water like the WaDE dataset (two curves), ii) the entire county seasonally and annually across eight demand sites (WEAP Model 2017), iii) part of the county monthly in one or seven sites as in the Bear River WEAP Model 2012 and WASH models, respectively. The reported annual water use data in WaDE is close to and validates the annual water demand values for the Cache Valley as used in the Bear River WEAP Model 2017. We recommend modelers to use the WaDE "Diversions" data which are annually reported by all water irrigation users in Cache County compared to using demand data that are constant across the years or covers part of Cache County. Here WEAP accepts input data with daily, monthly, seasonal, and annual spacing and aggregates or disaggregates them into the model's time step.

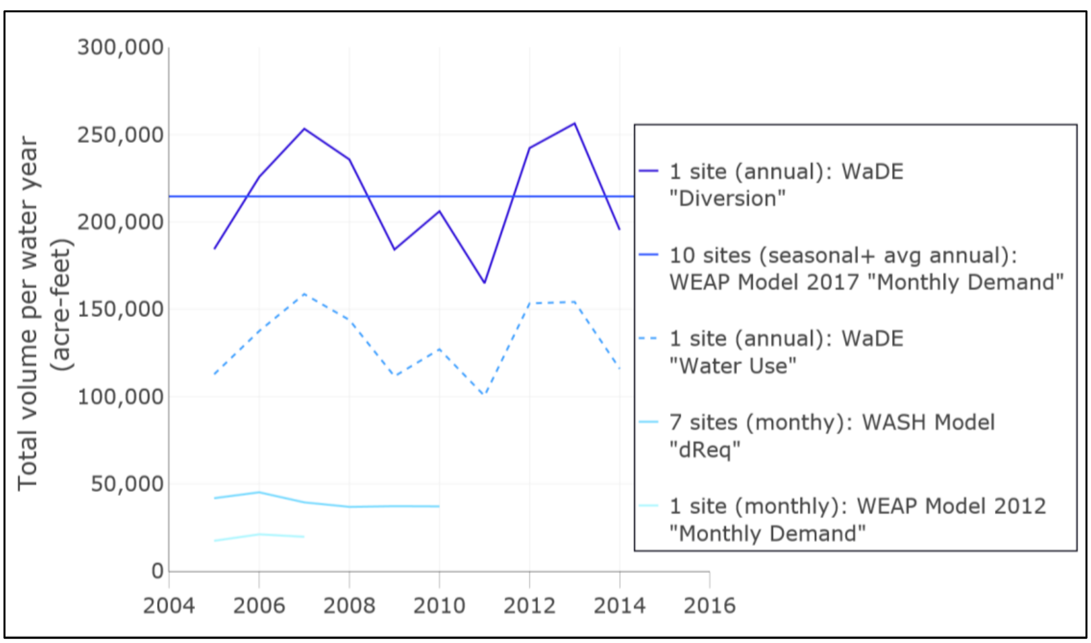


Figure 2.9: Water demand in Cache County, Utah by source with native attribute term in quotes.

Use Case 3.3: What reservoir volume-elevation curve should a modeler choose for a model?

Modelers also search for data describing multi-attribute series such as reservoir bathymetry (elevation versus storage) to represent the physical capacity of reservoirs in their models. Here, we use the controlled instance name of Hyrum Reservoir and controlled attribute names Volume and Elevation to identify four volume-elevation curves for Hyrum Reservoir from the USBOR, Utah Dams, and WEAP model datasets. The USBOR Water Info System dataset has two time series datasets for storage and elevation, which have the same daily time step from January 2010 to May 2017. We plotted both series (Figure 2.10) and used the WaMDaM CVs, metadata, and multiple data types to readily identify and compare multi-attribute bathymetry curves across data sources that had different semantics, measurement periods, and extrapolated versus measured methods. Metadata and semantics are valuable here as misrepresenting the total or live storage or using an old survey could over or underestimate water available to meet demand targets, especially in dry years.

Metadata indicate the four curves originate from two sources: the Utah Dams set and USBOR who owns the dam. The Bear River WEAP model used an older curve from the UDWR, while Utah Dams and USBOR datasets used USBOR source. Here we report the following three comparison insights, which are related to semantics, the range of data, and date of measurement. First, the top two red curves in Figure 2.10 indicate “live storage” which does not account for “dead storage,” while the lower two brown curves reflect “total storage.” The percentage of total storage that is dead storage is relatively high, about 17% in this small reservoir. Second, the slight differences between the two identical lower curves and the top curve are for two bathymetry surveys in 1935 and 2006, respectively. Between the two surveys, total storage decreased by 1,179 acre-feet which is 6% of the original storage due to a decrease in both the dead and live storage potential. Third, the lower brown curve has physical range that extend up to 70,000 acre-feet volume and 4,750 feet elevation (not shown) for a future scenario that raised the dam height. From the comparative analysis and metadata, we select the BOR 2006 curve which is for the recent bathymetry survey, used total storage as needed by WEAP, and stayed within the existing operational range of the reservoir.

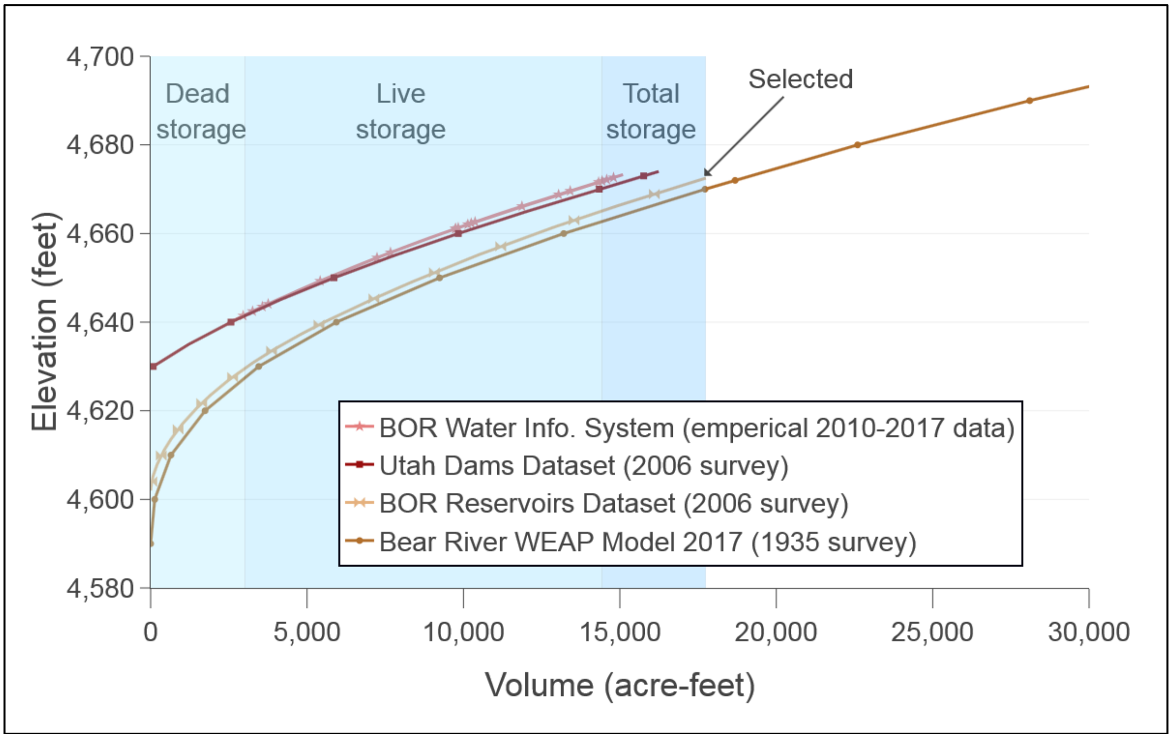


Figure 2.10: Four volume-elevation curves for Hyrum Reservoir, Utah. Lighter red and brown curves indicate larger volumes at the same elevation. Dead, Live, and Total storage zones are from the 2006 USBOR survey.

Use Case 4: What are the differences between two scenarios and which scenario should a modeler use?

Modelers use scenarios to evaluate how potential management alternatives can affect system performance. However, scenarios typically have numerous attributes and inputs and it is often difficult to determine the differences in nodes and links, data values, or data sources between multiple scenarios. Here we use the WaMDaM master network, scenario requirement, CVs, and the WaMDaM Wizard Data Loader comparison utility to help a modeler identify differences between existing scenarios in a model. The Wizard executes a script that queries the *ScenarioMappings* table and identifies the data that is shared among and unique to each scenario. Comparison results are exported to an Excel Workbook.

For example, the Bear River WEAP Model 2012 (Utah portion) and Bear River WEAP Model 2017 (Utah and Idaho portions) model scenarios share about 12% of the network node and link instances, 22% network metadata, 14% attribute metadata, and 14 % data (Table 2.4). Similarly, the BRSDM dry, normal, and wet scenarios have identical master network and metadata for the Wyoming portion of the Bear River Watershed and share about 93% of data like demand requirements with 3.5% unique values to each scenario, such as change in headflows (i.e., supply inflows into the system) (Appendix A, Table A4). The larger percentage of shared elements among the BRSDM model scenarios means a correspondingly larger savings in database storage than the WEAP model scenarios.

Because the Bear River WEAP Model 2017 model scenario has more node and link elements, metadata, attributes, and data values, we recommend using this model scenario as a starting point to expand coverage to the entire Watershed to include the Wyoming (dark red in Figure 2.5). The BRSDM model network covers the Upper Bear River in which can be used as a source to expand the WEAP Bear River WEAP Model 2017 to the entire Watershed.

Table 2.4: Unique and shared network nodes and links, metadata (source and method) and data between two WEAP Bear River Watershed model scenarios

Scenario comparison element	**Unique to "Bear River WEAP Model 2012" Scenario** Count of instances (%)	**Shared** Count of instances (%)	**Unique to "Bear River WEAP Model 2017" Scenario** Count of instances (%)
Network nodes and links	88 (23.5%)	45 (12%)	242 (64.5%)
Network metadata	88 (20.85%)	92 (21.81%)	242 (57.35%)
Attributes metadata	1,225 (26.5%)	654 (14.15%)	2,743 (59.35%)
Data	1,230 (26.61%)	696 (13.93 %)	2,748 (59.45%)

Use Case 5: How do annual water shortages at the Bear River Migratory Bird Refuge in the Bear River Watershed change when serving the Bear River WEAP Model 2017 model with new bathymetry, flow, and demand data selected in use cases 2 and 3?

We selected the Bear River Migratory Bird Refuge (hereafter, the Bird Refuge) at the mouth of Bear River as an environmental demand site to test the sensitivity of water shortages to changes in input of upstream supply, demand, and storage identified in use cases 2 and 3. The site has an annual 425,761 acre-feet water delivery target that is primarily required in the winter months. The WaMDaM CVs, consistent data storage, and query method enabled selecting the 1) dry seasonal headflow (i.e., supply inflows into the system) estimates for the Bear River at Stewart Dam that we derived from the UDWR dataset, 2) total maximum annual demand as reported by the WaDE dataset for the entire Cache County, and 3) bathymetry curve for Hyrum Reservoir from the USBOR dataset. We then used a Python 2.7 script in a local Jupyter Notebook and the WEAP API to export the selected data and populate data automatically in the Bear River WEAP Model 2017. This setup also allowed us to automate the process to create a WEAP scenario for each parameter change, execute the model, and report results for annual unmet demand (shortage) at the Bird Refuge. Each WEAP model run included the simulation period 1966 to 2006.

The modeled annual unmet demand ranged from 0% in wet years to up to 15% of total demand in dry years across the four scenarios (Figure 2.11). Updating Hyrum Reservoir with the new bathymetry (1,179 acre-feet less storage, 6% of capacity) had no observable effect on the annual unmet demand. The average annual unmet demand increased to 1.9% and 2.6% of total demand with higher upstream Cache County irrigation demand and updated headflows for dry years.

2.6 Discussion and Further Work

WaMDaM's eight design requirements of modular and extensible components, networks of nodes and links, scenarios, reusable contextual metadata, support for seven data types, extensible controlled vocabularies, direct access to data, and an open-source environment improve prior work that focused on managing water management data for a single model or dataset and select systems modeling data types. Here we discuss how modelers can use WaMDaM, list limitations of the work, present future work, and invite the community to get involved and provide feedback.

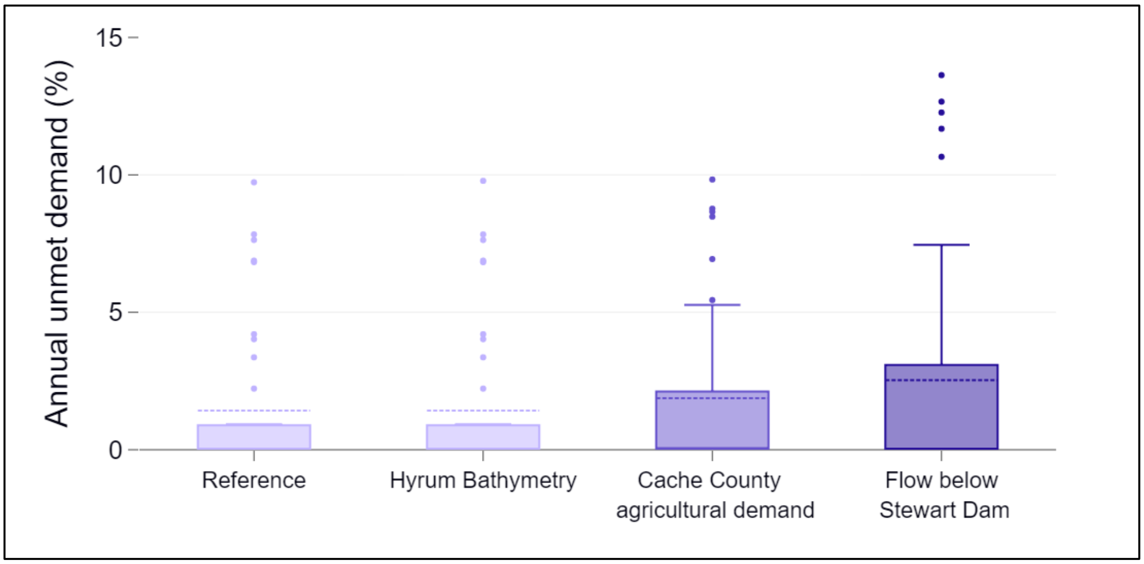


Figure 2.11: Sensitivity of annual unmet demand at the Bird Refuge, Utah over the simulation period 1966-2006 to changes in upstream storage capacity, demand, and supplies (mean values are in dash lines)

2.6.1 How can modelers use WaMDaM database and its software?

We show how researchers of five recently published systems modeling studies can use WaMDaM tools to organize, relate, and analyze input data, networks, and scenarios. For example, Ahmadaali et al. (2018) used WEAP to evaluate economic aspects of proposed water management strategies in Urmia Lake, Iran while Angarita et al. (2018) also used WEAP to examine 97 proposed hydropower facilities within a total of 1400 scenarios in the Magdalena River basin, Colombia. Both projects can use the WEAP importer in WaMDaM Wizard to manage the WEAP networks and compare input data for current and future scenarios.

Dogan et al. (2018) developed an open-source version of the California Value Integrated Network (CALVIN) model and separate the model from model data which is stored in a large number of CSV and JSON files in a structured GitHub repository. The researchers could use the WaMDaM Wizard to load input data into the WaMDaM database and compare the input data for different models runs such as for 10 and 40 years' time spans. Wheeler et al. (2018) developed a systems optimization model to identify cooperative management strategies for the large reservoirs on the Eastern Nile Basin. The researchers could use WaMDaM and its scenario comparison tool to track different projected climate change flows for the Nile Basin. Finally, Chini et al. (2018) created a network of virtual water flows for the US electric grid based on six years of empirical data on water use and electricity transfers. The authors could use WaMDaM to store the created network and its disparate water and energy datasets. WaMDaM can be especially useful to manage the data for the proposed analysis to assess regional interdependencies on a seasonal scale. For each of these studies, storing the modeling data in WaMDaM with its defined schema and publishing it online such as on GitHub will allow other researchers to query and reuse data in other studies. This reuse could further increase each study's impact.

2.6.2 Current limitations

WaMDaM supports numerical, seasonal, categorical, free text, time series, multi-attribute series, and electronic data formats. WaMDaM, however, does not support gridded data since gridded data are not common to the water resources models we reviewed. The WaMDaM design is implemented in a relational schema, which has limitations to adapt and scale compared to NoSQL databases. The WaMDaM tools help users interact with its

SQLite database installed on one machine with no distributed access compared to database servers with API. These software tools are prototypes that are tested using the study datasets on Windows machines. The WaMDaM Wizard is slow to load and validate large datasets.

2.6.3 Future Work

To improve access and security, future WaMDaM implementations should build web-server APIs with data query functions that distribute and manage the access to many users at the same time and protect the database integrity from unintended changes. Future software tools to load data to the database and export it to models should be time-efficient, more user-friendly, and compatible with Windows, Mac, and Linux. To support more use cases, future work should involve a larger number of diverse datasets, models, and research groups. Future work also should use WaMDaM and web-services to publish, discover, and visualize models and their data and allow multiple users to work with the same datasets. Additionally, future work could leverage scenario and attribute metadata to test use cases that convert data in one-time step to other time steps.

In response to earlier feedback, we are collaborating to build a software ecosystem to make WaMDaM interoperable with Hydra Platform web-services (Knox et al., 2014), OpenAgua (Rheinheimer et al., in review), and HydroShare (Tarboton et al., 2014). The ecosystem tools will allow WaMDaM users to import data stored in Hydra Platform as a new source of data. Users will also be able to export WaMDaM data into Hydra Platform and visualize networks and their data in OpenAgua. We are also integrating WaMDaM as a new HydroShare resource type to publish populated WaMDaM SQLite files and extract their metadata to enable search and discovery (Horsburgh et al., 2015). Lastly, we are developing workflows to automate the steps to prepare and export all the data needed to run multiple models. These workflows will more readily allow modelers to use the same datasets to run multiple comparison models for the same study domain (e.g., simulation versus optimization) or different spatial domains (e.g., Bear River versus Colorado River). These tasks are now difficult because the modeler must manually build two (or more) models from scratch.

2.6.4 Invitation to community involvement and feedback

Over the past five years, we sought and received feedback from colleagues and collaborators on the WaMDaM design and tools. There is still need for testing and feedback from a larger, more diverse community of users. In all these efforts, we seek community involvement to 1) add new datasets and models for new locations, 2) build new exporters to serve data to new models, and 3) further define the system of controlled vocabularies that can help relate native vocabulary of existing models and datasets. More involvement can benefit a variety of people who work with systems simulation and optimization data and models. WaMDaM can serve as a first step toward a standardized method to store, organize, and share water resources systems modeling data.

2.7 Conclusions

This paper addressed the problem of needing multiple methods to organize, store, query, and analyze water management data to identify input data to develop or extend a water management model. We contributed a new data model (WaMDaM) implemented in a relational database to organize water management data with contextual metadata and controlled vocabularies to generalize data analysis for multiple data sources, models, and study areas.

The design of WaMDaM integrated eight design requirements that were previously only partially supported by forty prior water resources data systems, models, and standards. The requirements include: 1) modular and extensible components, 2) networks of nodes and links, 3) scenarios and version control, 4) reusable contextual metadata, 5) support for multiple data types used by systems models, 6) extensible controlled vocabularies, 7) direct access to subsets of data and metadata, and 8) an open-source environment.

We demonstrated the WaMDaM design by using 13 datasets and models to answer five use case questions in the Bear River Watershed, United States. The use cases allowed modelers to: i) search for input data within a model study area, ii) identify flow directions and connections among natural and engineered system components, iii) identify and compare water supply, demand, and reservoir data across multiple datasets and models, iv) show data similarities and differences among modeling scenarios, and v) select data, serve the data to a model, and run multiple model scenarios.

Results showed how WaMDaM unifies data formats, structures, and controlled vocabulary identified data for 15 attributes (out of 71 needed) from six data sources to expand the spatial extent of a WEAP model. Results also showed discrepancies in river discharge data, demand, and reservoir area-elevation curves. Results helped select input data and develop multiple scenarios. Serving the data to run an existing WEAP model revealed and quantified that shortages at an environmental demand site were sensitive to changes in upstream agricultural water demand and river headflows but not reservoir capacity.

The WEAP API and SQL make it possible for users to use WaMDaM to set up scenarios, replicate, and extend the work. WaMDaM facilitates these data wrangling tasks by reconciling the disparate datasets into a homogenous structure and by using controlled vocabularies to relate the different native terms across datasets. Modelers can then spend more time on data analysis and synthesis than on time consuming and error-prone steps to manipulate data to set up and run a model.

In further work, we are collaborating on a software ecosystem to make WaMDaM interoperable with Hydra Platform and OpenAgua to visualize networks and their data. We are also developing workflows to automate the steps to serve the same input data already organized in WaMDaM to multiple comparison models for a study area. We also seek community involvement to load larger and more diverse data and model sets which will allow others to reuse data and build models in new areas. These expansions will require more robust methods to define, relate, specify, and expand controlled vocabularies for water management data. We invite the systems modeling and hydroinformatics communities to provide feedback to improve WaMDaM.

Acknowledgments

This work was supported by the National Science Foundation through the CI-Water Project grant 1135482, iUtah grant 1208732, and Utah Mineral Lease Funds. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The Intermountain Section of the American Public Works Association, the Utah Chapter of the American Public Works Association, and the Utah Water Users Association provided additional financial support. Jeffery Horsburgh provided extensive feedback on the final design and writings. David Tarboton, Stephen Knox, and David Rheinheimer provided thoughtful feedback and comments on earlier designs and drafts. Hadia Akbar and Jiada Li used the WaMDaM Wizard and Jupyter Notebooks to reproduce use case results. We thank the two anonymous reviewers for their constructive feedback.

References

Abdallah, A.M., 2018a. WaMDaM Controlled Vocabularies Source Code. GitHub.

Abdallah, A.M., 2018b. WaMDaM Information Model Source Code. GitHub.

Abdallah, A.M., 2018c. The WaMDaM Wizard Source Code. GitHub.

Abdallah, A.M., 2020. WaMDaM JupyterNotebooks.https://zenodo.org/record/1484581

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).

Ahmadaali, J., Barani, G.-A., Qaderi, K., Hessari, B., 2018. Analysis of the Effects of Water Management Strategies and Climate Change on the Environmental and Agricultural Sustainability of Urmia Lake Basin, Iran. Water 10(2) 160.

Alafifi, A.H., Rosenberg, D.E., 2020. Systems modeling to improve river, riparian, and wetland habitat quality and area. Environmental Modelling & Software 126 104643.

Angarita, H., Wickel, A.J., Sieber, J., Chavarro, J., Maldonado-Ocampo, J.A., Herrera-R, G.A., Delgado, J., Purkey, D., 2018. Basin-scale impacts of hydropower development on the Mompós Depression wetlands, Colombia. Hydrol. Earth Syst. Sci. 22(5) 2839-2865.

Chini, C.M., Djehdian, L.A., Lubega, W.N., Stillwell, A.S., 2018. Virtual water transfers of the US electric grid. Nature Energy.

Dogan, M.S., Fefer, M.A., Herman, J.D., Hart, Q.J., Merz, J.R., Medellín-Azuara, J., Lund, J.R., 2018. An open-source Python implementation of California's hydroeconomic optimization model. Environmental Modelling & Software 108 8-13.

Harou, J.J., Pinte, D., Tilmant, A., Rosenberg, D.E., Rheinheimer, D.E., Hansen, K., Reed, P.M., Reynaud, A., Medellin-Azuara, J., Pulido-Velazquez, M., Matrosov, E., Padula, S., Zhu, T., 2010. An open-source model platform for water management that links models to a generic user-interface and data-manager In: David A. Swayne, W.Y., A. A. Voinov, A. Rizzoli, T. Filatova (Ed.), International Congress on Environmental Modelling and Software, Modelling for Environment's Sake ed. International Environmental Modelling and Software Society (iEMSs) Ottawa, Ontario, Canada.

Hoberman, S., 2014. Data Modeling for MongoDB: Building Well-Designed and Supportable MongoDB Databases. Technics Publications.

Horsburgh, J.S., Aufdenkampe, A.K., Mayorga, E., Lehnert, K.A., Hsu, L., Song, L., Jones, A.S., Damiano, S.G., Tarboton, D.G., Valentine, D., Zaslavsky, I., Whitenack, T., 2016. Observations Data Model 2: A community information model for spatially discrete Earth observations. Environmental Modelling & Software 79 55-74.

Horsburgh, J.S., Morsy, M.M., Castronova, A.M., Goodall, J.L., Gan, T., Yi, H., Stealey, M.J., Tarboton, D.G., 2015. Hydroshare: Sharing diverse environmental data types and models as social objects with application to the hydrology domain. JAWRA Journal of the American Water Resources Association.

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.

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.

Rheinheimer, D.E., Medellin-Azuara, J., Ramirez, A., Brown, C., Park, D.K., Torres, E., Sood, A., Knox, S., Garza-Díaz, L., Abdallah, A.M., in review. OpenAgua: A web-based decision support platform for collaborative water systems modeling and analysis. Environmental Modelling and Software.

Rosenberg, D., 2017. Bear River WEAP Models (2012 and 2017).

Samu, N.M., Kao, S.-C., O'Connor, P.W., 2017. Existing Hydropower Assets [series] FY17Q4, In: National Hydropower Asset Assessment Program (Ed.): Oak Ridge National Laboratory, Oak Ridge, TN.

Sehlke, G., Jacobson, J., 2005. System Dynamics Modeling of Transboundary Systems: The Bear River Basin Model. Ground Water 43(5) 722-730.

Tarboton, D.G., Idaszak, R., Horsburgh, J.S., Heard, J., Ames, D., Goodall, J.L., Band, L., Merwade, V., Couch, A., Arrigo, J., 2014. Hydro share: Advancing collaboration through hydrologic data and model sharing, 7th International Congress on Environmental Modelling and Software, iEMSs 2014. International Environmental Modelling and Software Society.

U.S. Geological Survey, 2013. USGS Small-scale Dataset - Major Dams of the United States 200603 Shapefile: U.S. Geological Survey, 30-May-2013 ed. http://nationalatlas.gov: Reston, VA.

Wheeler, K.G., Hall, J.W., Abdo, G.M., Dadson, S.J., Kasprzyk, J.R., Smith, R., Zagona, E.A., 2018. Exploring Cooperative Transboundary River Management Strategies for the Eastern Nile Basin. Water Resources Research 0(ja).

CHAPTER III

OPEN SOURCE PYTHON SOFTWARE TO MANAGE, POPULATE, COMPARE, AND ANALYZE WEAP MODELS AND SCENARIOS

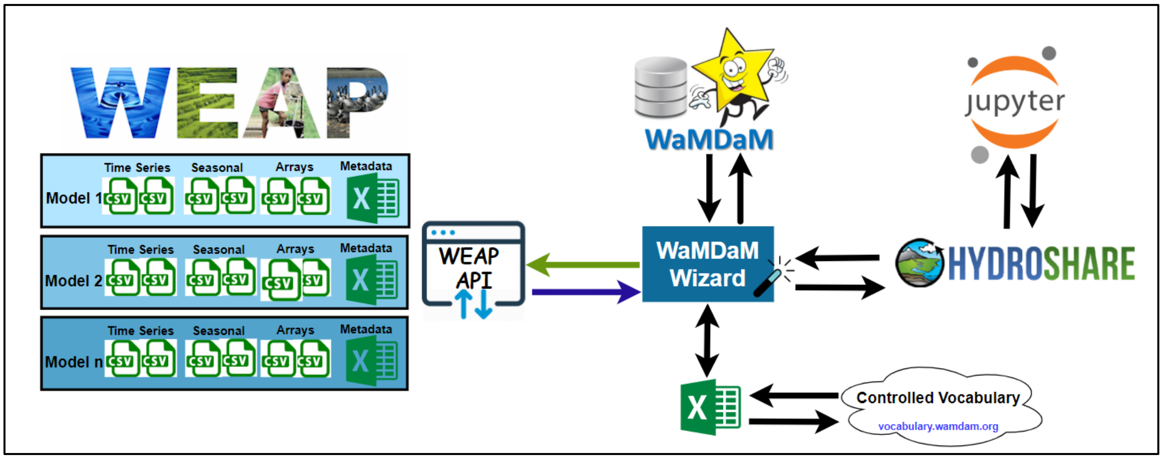
Abstract

The Water Evaluation and Planning system (WEAP) is a proprietary systems simulation software that is used globally for water management modeling studies. WEAP has a simple and powerful Application Programming Interface (API), however most WEAP modelers manually populate data into their WEAP area (model). Manual operations are error-prone and time-consuming. We contribute open-source Python software that automates and generalizes the processes for WEAP modelers to prepare and load data and run sensitivity analysis for multiple WEAP areas and their scenarios without writing code. The software also allows users to export and store model data then run independent analyses without a WEAP license. We demonstrate the software with existing WEAP areas for the 1) Bear River Basin in Idaho and Utah and 2) Weber River Basin in Utah. Results show estimated demand reliability for changes in reservoir capacity, demand, evaporation, and river headflows. . Reliability to meet demands in both the Bear and Weber Rivers models varied from 50% to 100%.

Keywords

Sensitivity analysis, reproducibility, systems modeling, water resources, decision support

Graphical Abstract



Highlights

- We present a software architecture and functionality to manage and automate data import to and export from multiple WEAP models and scenarios
- Modelers can automate the setup and run of many WEAP models and scenarios, then compare results across the models and scenarios

Software availability

<u>Name of software:</u> The WaMDaM Wizard

<u>Developer</u>: Adel M. Abdallah

<u>Contact:</u> Adel M. Abdallah; 8200 Old Main Hill, Logan, UT 84322, USA; Email amabdallah@aggiemail.usu.edu

<u>Year first available:</u> 2019

Required hardware and software: The WaMDaM Wizard executable is available for use with Microsoft Excel (2007 and later versions) and SQLite3 on Windows 64-bit computers.

Input data and directions: Documentation of all source code, datasets, use cases, and instructions to use the WaMDaM Wizard and WEAP and replicate results are available on (Abdallah, 2020b). Jiada Li used the WaMDaM Wizard, WEAP, and Jupyter Notebooks to replicate use case results.

3.1 Introduction

The Water Evaluation and Planning (WEAP) system is a proprietary software for integrated water resources planning. The user-friendly desktop software is used around the world for water-related policy analysis (Yates et al., 2005). WEAP supports integrated water resources planning with its built-in functions for aggregated rainfall runoff and infiltration, snowmelt, evapotranspiration, crop requirements and yields, surface water/groundwater interactions, and instream water quality. The software can read data from Excel and comma separated value (CSV) files and has functions to export both input and output data to Excel. WEAP has data visualization utilities that support plotting data for most model inputs and results. The “Results” dashboard in WEAP also allows users to compare select output parameters within and across scenarios. WEAP supports 14 different water system components (i.e., object types) such as Reservoir, Wastewater Treatment Plants, and Demand Site. Each object type has a list of attributes such as Monthly Demand, Capacity, and Capital Costs with a total of 220 attributes (Yates et al., 2005) (Figure 3.1). WEAP is similar to other proprietary water systems modeling software such as RiverWare

(Zagona et al., 2001), GoldSim (GoldSim Technology Group LLC, 2014) and eWater Source (Welsh et al., 2013).

WEAP's wide-range modeling capabilities require considerable input data to describe the physical system and water operations. WEAP modelers can benefit from a generalized, consistent, and reusable software that completely prepares all input data from one or more data sources and then populates data into WEAP with minimal user intervention. WEAP users can also benefit from a generalized open-source database that stores extracted WEAP data and allows users to query, compare, analyze, and plot all input and output data across multiple scenarios and models. Storing model data in an open-source database will also allow the broader community of researchers to discover, search, analyze, and publish WEAP modeling data online in public repositories such as HydroShare (Tarboton et al., 2014) without need for a WEAP license. Journal policies, funding agencies, and several recent studies have encouraged publishing modeling data along with code and directions to support reproducibility and data reuse (Stagge et al., 2019; Rosenberg et al., 2020).

To manually prepare WEAP input data, users must find the data, organize the data in the structure required by WEAP, reconcile the syntax that describes the data to WEAP's nomenclature, then enter data or the link to data in WEAP. For example, a monthly time series of flows for a reach high up in a watershed that is specified in an external file must be organized as a comma-separated values (CSV) file and related to the WEAP model input "headflow". The file must include three columns, the first for the year, the second for the month, and the third for the data value. Then, the user tells WEAP the file's path and name on their machine for the parameter. The user must manually repeat these steps for all other input data such as monthly demand, arrays of reservoir storage and elevation curves, numeric parameters for demand priorities, and expressions (i.e., equations) the describe complex interactions or rules between system components. Manually preparing large volumes of input data and populating data to a WEAP model for a study area is time-consuming and error prone.

WEAP has an Application Programming Interface (API) that supports outside programming languages such as Python, Visual Basic, and C to read and write data and execute commands. Most WEAP publications that we reviewed do not mention using the API (Sanvicente-Sánchez et al., 2009; Mourad and Alshihabi, 2016; Gao et al., 2017; Winter et al., 2017).

If modelers use the API for scenario and sensitivity analysis, they often use a mix of custom and study-specific Excel spreadsheets, CSV files, and Visual Basic scripts to load data into or extract it out of WEAP (Craven et al., 2017; Jamshid et al., 2017; Mehta et al., 2018). . For example, Jamshid et al. (2017) developed a study-specific Visual Basic scripts within Excel that automates changing decision variables of specific reservoirs' storage capacity and its filling priority to couple WEAP to a multi-objective optimization model. Craven et al. (2017) developed custom Visual Basic code within Excel to automate updating input data in WEAP from input cells in the Excel sheets. Mehta et al. (2018) used Visual Basic scripts within Excel to call the WEAP API and populate their WEAP model with 84 combinations of the seven identified strategies, two demand projections, three climate projections and two groundwater pumping curtailment projections. They then used WEAP to evaluate seven water management strategies as part of the groundwater sustainability plans for Yolo County in the Central Valley of California. These automation efforts can only be used for the specified model and scenarios. WEAP modelers who want to apply the methods to other models and scenarios must develop new custom spreadsheet files and Visual Basic code. Developing new files and code is also time consuming and error prone.

Here we develop generic, automated tools to 1) extract, query, compare, and analyze many WEAP models' data outside its proprietary database, and 2) quickly set up multiple WEAP models and scenarios and populate them with data stored in an open-source, relational database called the Water Management Data Model (WaMDaM) (Abdallah and Rosenberg, 2019). The automation tools can work across many WEAP models and scenarios because they draw on several generic WaMDaM database capabilities to:

- Organize systems modeling data with metadata which describe the locations, observed variables, sources, methods, units, and people and organizations involved in creating or reporting input data.
- Provide users with controlled vocabulary to relate native modeling terms across models. For example, relate a native "Transmission Link" term in WEAP to the controlled term "Canal" which can be further related to a different native term "Diversion Link" in another model.
- Run supporting Python-based software to validate inputs,
- Import data from multiple sources including (i) Generic Microsoft Excel workbook template, (ii) Stream discharge time series data from the Consortium of Universities for the Advancement of Hydrologic Science, Inc.

(CUAHSI) Hydrologic Information System (HIS) web services, (iii) Reservoir storage and releases time series data from the U.S. Bureau of Reclamation Information System, (iv) Hydra Platform web-services and the OpenAgua online application (Abdallah, 2019), and (v) resources published in HydroShare (Abdallah, 2019).

We demonstrate the capabilities to 1) extract, query, compare, and analyze many WEAP models' data and 2) quickly set up multiple WEAP models and scenarios and populate data in two use cases for separate existing WEAP models of the Bear and Weber River watersheds, USA. Both watersheds cover 9,913 square miles and terminate into the Great Salt Lake (GSL), Utah and on average contribute 1,450,000 acre-feet, about 40% of the GSL total annual inflow (SWCA Environmental Consultant, 2013). The use cases answer the questions: how are attributes, networks components, and data used among the two models similar and different? How is demand reliability in each river basin sensitive to response reservoir sedimentation, increased net evaporation and demand, and reduced river headflows?

The remainder of the paper is organized as follows: Section 3.2 describes the methods to automate the workflows to extract data out of WEAP into WaMDaM, conduct independent analyses, generate scenarios, and then populate scenario data back into WEAP. Sections 3.3 and 3.4 describe the Bear and Weber River watersheds and compare results across the two WEAP models. Section 3.5 discusses the results, presents limitations and recommendations, and invites the community to build similar connections with other systems modeling software. Section 3.6 concludes.

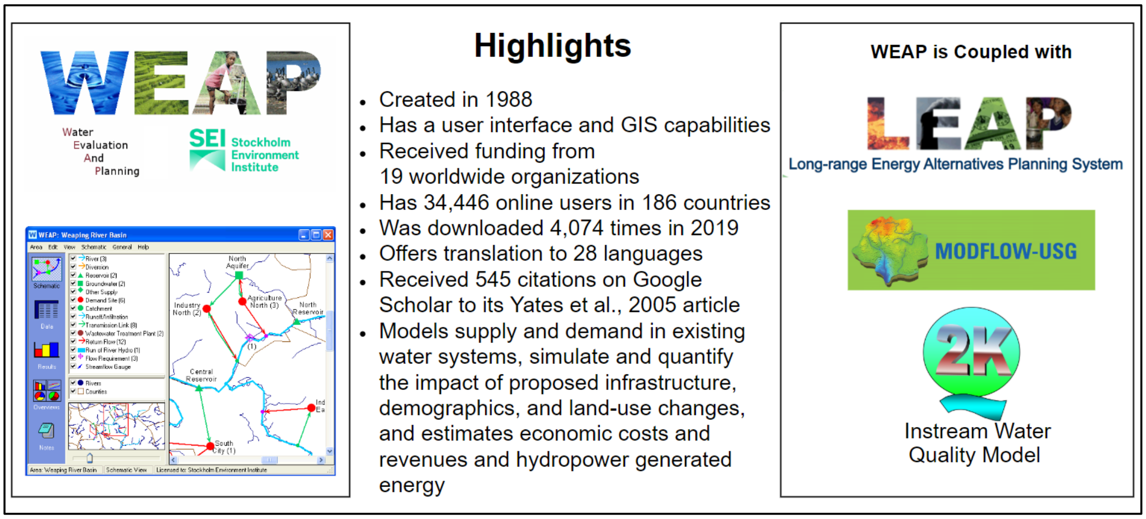


Figure 3.1: Water Evaluation and Planning (WEAP) software capabilities and use around the world

3.2 Methods

We first use the WEAP API to i) extract data from each separate WEAP model within its proprietary database and store and organize them into a single WaMDaM open source database. Then we compare the two models and import new data for external scenario analysis into each WEAP model and execute sensitivity analysis. In both import and export functions, we used Python, Structured Query Language (SQL) and the WEAP API to move data out of WEAP for analysis and comparisons or to prepare new WaMDaM data, and populate it into many WEAP models and their scenarios and thus complete the circle in moving data (Figure 3.2). We implemented both the extract and populate functions in Python scripts as part of the WaMDaM Wizard. To use the functions, users must have WEAP installed on their Windows machines with an active license. Users no longer need a WEAP license once data is exported from WEAP. The WEAP API is designed to act as a standard Component Object Model (COM) Automation Server which an object-oriented system for Windows that supports other programs in languages such as Python and Visual Basic. The API allows outside scripts to read and write input data and then execute WEAP. We used Python to work with WEAP through the "PyWin32" library which gives Python access to the Windows COM Automation Server API (Hammond, 2020). Using Python to connect to the WEAP API results in a data object that contains WEAP classes and their properties. We mapped those WEAP classes and properties into their equivalent metadata elements in WaMDaM (Table 3.1). The next two subsections describe how we specifically used these key mapped elements to transfer data between WEAP and WaMDaM.

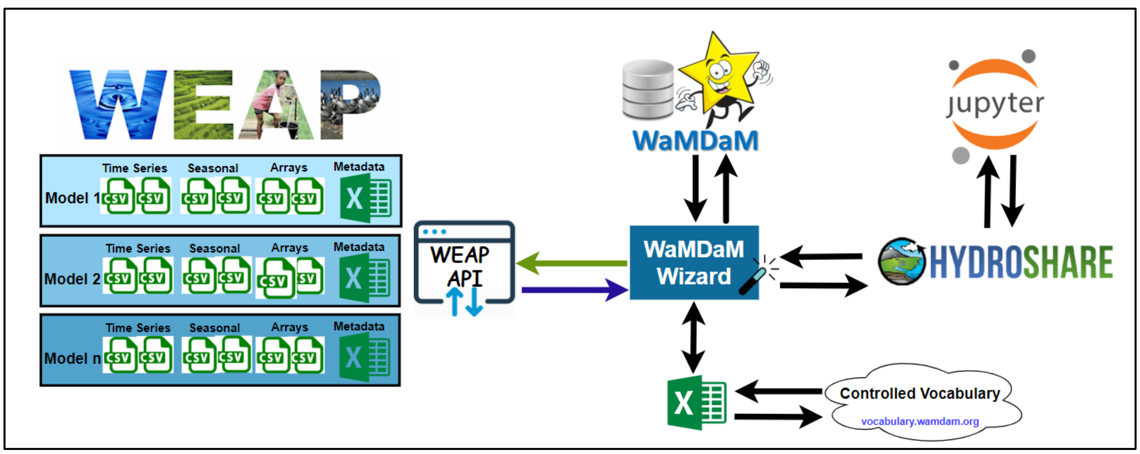


Figure 3.2: Workflow to 1) automate extracting WEAP model data into WaMDaM (green arrow)2 and 2) populate new scenario data from WaMDaM into WEAP models (green arrow). Outputs can be a published WaMDaM SQLite file in HydroShare for a WEAP model's data

3.2.1 Extract WEAP Areas into WaMDaM

There are five steps to extract a WEAP Area, including its structure (i.e., list of object types and attributes), network and scenarios, and data into a WaMDaM Excel workbook. The steps are: i) connect to the WEAP API and get its network, scenario, and directory of the model files on desktop; ii) get the list of object types and attributes and their units in WEAP and load them into WaMDaM database; iii) extract WEAP model network of nodes and links; iv) get the values of WEAP variables and transform them to WaMDaM data structures; and v) write the extracted WEAP data into WaMDaM workbook sheets in Excel according to WaMDaM template. We describe the logic of each step in Appendix B.

Table 3.1: Mapping the common key equivalent metadata elements between WaMDaM and WEAP

WaMDaM	**WEAP**	**Common description**
Master Network	WEAP Area	A collection of node and link interconnected water system components that serve a common purpose such as allocating water supplies for competing demand sites given the capacities of a natural and built system elements.
Scenario	Scenario	A specific socio-economic setup of the network that has changes in references to a baseline condition.
Object Type	BranchType	A water system component type (e.g., reservoir, river)
Node, Link, Network attributes	Node, Line, key assumptions	The typology of a water system component.
Attribute	Variable	A property of a water system component with values
Instance	Branch	A specific implementation of a water system component that may be referenced geospatially
Data Value	Expression	A quantitative or qualitative measure for an attribute of a system component instance

These steps were implemented as Python functions in the WaMDaM Wizard under the “Import Data into WaMDaM” tab and “From WEAP” button. We use the WaMDaM

Excel workbook template as an intermediate step in the extract process between WEAP and WaMDaM SQLite for three reasons: (i) to take advantage of the WaMDaM Wizard data loader that works with and validates the Excel workbook data, (ii) to allow users to optionally enhance the extracted WEAP data with metadata such as organizations, people, sources, and methods that are used in WEAP data, and (iii) to allow users to relate their model's nodes and links native names with controlled vocabulary terms. The WaMDaM workbook includes 14 spreadsheets that generically organize water management data, metadata, and provide lookup-controlled vocabularies to allow users to relate them with their native terms.

3.2.2 Populate WEAP Models from WaMDaM

The following are the steps a modeler follows to automate the populating of hundreds of attributes into multiple WEAP models and scenarios all at once using a generalized Excel workbook without needing to write custom specific scripts or spreadsheets. Within WEAP interface for a study area, first, draw a WEAP model node and link schematic and choose specific units (i.e., metric or English). Second, use the WaMDaM Wizard utility to extract (export) the blank WEAP data structure, which includes the network of nodes and links, and scenarios into a WaMDaM workbook. Third, provide data values and metadata into the workbook, such as all reservoirs storage or demand sites priority. Provide values in bulk (i.e., each value is in a row) rather than manual entry of every single value one-at-a-time using the WEAP interface. If needed, we the WaMDaM tools to define new scenarios in the workbook and provide data for them such as parameter values for sensitivity analysis. Here modelers need to prepare their data to fit into the WaMDaM workbook structure using the same units they chose while defining the WEAP Area. Fourth, load the Excel workbook populated with input data into the SQLite WaMDaM database. The WaMDaM Wizard checks and validates the provided input data and metadata and their correct association with the network components, scenarios, and WEAP data structure including data types such as time series, seasonal or numeric data. Fifth, from the WaMDaM Wizard "Export Data to Models tab" tab, select the "Serve to WEAP" button to serve the WEAP input data in the SQLite database into the WEAP model schematic defined in the first step.

This Wizard function queries the WaMDaM database for the selected model, network, and scenario. The function iterates over each object type and its instances (nodes or links) in WaMDaM and looks for their match in WEAP. For each object type, the function iterates through the object attributes in WaMDaM and looks for a match in the attribute name and its unit in WEAP and then queries them based on their data type (e.g., time series, seasonal). The function then transforms the structure of each value and prepares it as required by WEAP. Finally, each value is provided as an expression in WEAP for its location as defined by the unique triple metadata: object type, instance name, and attribute name (Appendix B, Table B1). The function only serves data to WEAP where there is a match in WEAP and WaMDaM object types, instances and attributes.

The Wizard creates a subfolder inside its WEAP Area folder for each populated scenario Within each folder, the Wizard creates subfolders for the CSV files for seasonal, time series, and multi-column data (Figure 3.3). Finally, the Wizard creates a metadata Excel file for the record that lists all the input loaded parameters, their source and method names, and the data values. WEAP users can share the CSV files for input parameters and they can read their metadata to understand where data originate from and how data were calculated in the metadata Excel file.

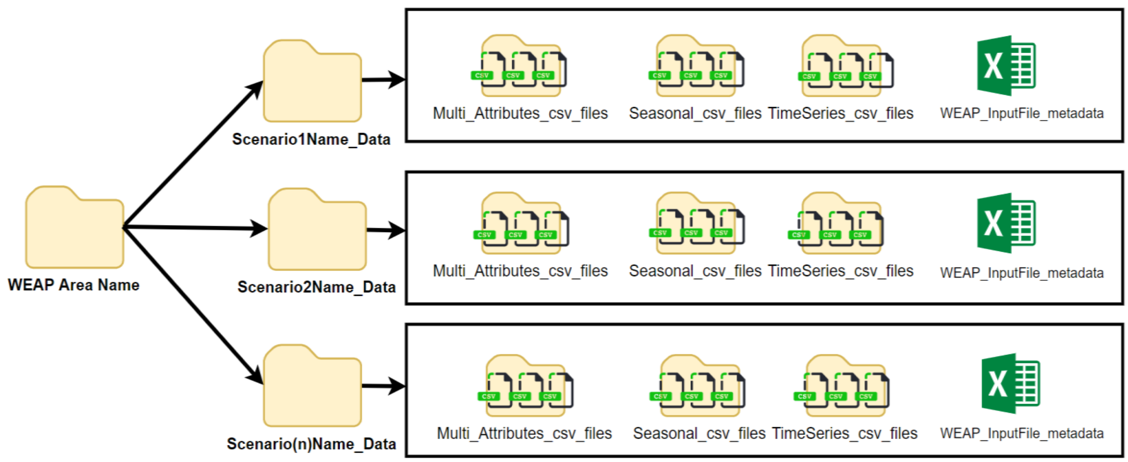


Figure 3.3: Files structure that is generated by the WaMDaM Wizard for each scenario inside a WEAP Area folder on the user's desktop machine.

3.3 Use Cases

Two use cases illustrate that automated export of data out of and importing of data into WEAP models and scenarios for the Bear and Weber Rivers in Utah, USA. The use cases draw on existing WEAP models for the Bear River (Utah and Idaho portions) (Rosenberg, 2017) and Weber River (Utah) (Tesfatsion and Rosenberg, 2013) watersheds USA (Figure 3.4). Both WEAP model instances allocate water to competing demand sites based on water right priority. The Weber model spans 1951-2006, and the Bear model spans 1966-2006. The models use a mix of data types: seasonal data for demand, time series for river headflows, arrays for reservoir storage and elevation curves, numeric parameter for demand priority, and expressions (i.e., equations) that represent a text value.

We used the same WaMDaM Wizard software to work with the two models and scenarios, which demonstrates that the software is independent of the study location, WEAP model instance, and scenario. The use cases assume that the user already has a license to the WEAP software and has downloaded both WEAP and the WaMDaM Wizard to a windows desktop machine.

Use case 1: <u>How are attributes, networks components, and data used in the Bear and Weber WEAP models similar and different?</u>

The first use case demonstrates how modelers can use the WaMDaM Wizard to query, summarize, and compare many WEAP models and their datasets. The use case represents modelers' needs to query, plot, and analyze modeling data across models outside the WEAP proprietary database. These steps can help modelers answer the practical question how do ratios of river flow to basin storage compare across basins? This ratio is sometimes called the flow regulation factor and is the percent of total built-storage divided by the average annual river discharge (Nilsson et al., 2005).

Here we used the WaMDaM Wizard to extract models for the Bear and Weber Rivers into a WaMDaM Excel workbook and load the data into a WaMDaM SQLite database. We then used a Python script and Structured Query Language (SQL) in a Jupyter Notebook to query and summarize the two models' input data. Then we estimate, and compare the average annual discharge (acre-feet) including river headflows and reach gains attributes in WEAP, total built reservoir capacity (acre-feet), and average annual demand (acre-feet) for both models. Finally, we used the WaMDaM Wizard to publish the SQLite database into HydroShare to enable its discovery as described in Abdallah and Rosenberg (2019).

Use case 2: Estimate sensitivity of demand reliability to changes in reservoir sedimentation, net evaporation, demand, and headflows?

This use case supports modelers' needs to prepare, populate, and run WEAP models with multiple scenarios and large input data. We defined four scenarios for the Bear and Weber Rivers that simulate changes in reservoir sedimentation, net evaporation, demand, an supplies and are similar to changes in input data and sensitivity analysis often carried by WEAP users (Craven et al., 2017; Jamshid et al., 2017; Mehta et al., 2018). Here we define water system reliability as the number of years with zero total annual shortage at a demand site divided by the number of simulation years (percentage) (Loucks et al., 2005). Reliability was calculated for five demand sites (out of 21) in the Bear River model and two out of 19 sites in the Weber River (Figure 3.3). Meeting demand depended on water availability, demand target, timing (i.e., month), and demand priority. The higher the demand priority (i.e., seniority in water right), the less the demand site is affected by a small reduction in water availability.

The first reservoir sedimentation scenario reduces current reservoirs' capacities by 10% at once due to sedimentation generally trapped in reservoirs over time (Graf et al., 2010; SWCA Environmental Consultant, 2013). Reservoir capacity in WEAP is defined in three input parameters, storage capacity, initial storage, and storage elevation curves. The second scenario simulates a 10% urban and agriculture demand conservation from current demand targets following national conservation trends (Dieter et al., 2018) and partial fulfillment of Utah regional conservation goals (Jones, 2019). The demand input parameter is defined as seasonal and represents stationary demand across the simulation years. The third scenario increases net evaporation in reservoirs by 10%, which represents warmer and drier climate projections (Hill et al., 2014). Net evaporation is often defined as a time series or seasonal with monthly time steps. The fourth scenario decreases the rivers' headflow by 10% (Kopytkovskiy et al., 2015). These headflows supply inflows to the system and represent climate projections of reduced precipitation. Headflows are often specified using time series data, but can also be input as seasonal data, especially for springs.

We defined the input data for each scenario using the generic WaMDaM Excel workbook. Then, we used the WaMDaM Wizard to load them into a WaMDaM SQLite database. Finally, we used the WaMDaM Wizard to connect to WEAP and completely populate it with input data for each scenario one-at-a-time from the SQLite database. We used Jupyter Notebooks and the WEAP API to analyze and plot system reliability to meet demand at each site. We then used the WaMDaM Wizard to publish both models' input data in a SQLite file in HydroShare (Abdallah, 2020a). We note that these changes do not fully capture the dynamics of geology (for sedimentation), temperature, storage, surface area, and evaporation (Mitchell et al., 2018).

We verified the integrity of the use case steps and workflows. First, we extracted the original WEAP model data for both models into the WaMDaM database. Second, we created copies of the original WEAP models in WEAP and set input parameters to zeros in one copy of each model. Third, we repopulated the WEAP models with input data from WaMDaM. Lastly, we ran the saved and repopulated models and compared results. We specifically compared the simulated result of unmet demand at all sites and the total supplied water at the most downstream sites. This verification helped us uncover issues (e.g., typos in code) and fix them. As a result, WEAP users can organize and store their WEAP modeling data into a WaMDaM database and use the WaMDaM Wizard to prepare and populate their models with data and allow modelers to perform sensitivity analysis using this generalized framework.

3.4 Results

Results for the first use case to compare attributes, network components, and data across the Bear and Weber models show both models have hundreds of nodes/links and input parameters that use diverse data types (Table 3.2). The Bear River model represents demand in seasonal format as stationary across the years while the Weber model represents it as historical time series that reflect demand changes across years. Both models represent river headflows as time series which reflect the natural hydrology cycle that includes wet and dry years. These comparisons show the large amount of input data that WEAP modelers must typically prepare manually.

Further analysis suggests that the Bear River flow is much more highly regulated than the Weber River flow (Table 3.3). The Bear Lake storage capacity alone of 1,516,633 acre-feet exceeds the Bear River annual demand and discharge. This suggests the lake's significance in the system and the importance of including it in strategic cooperation between Idaho and Utah to manage its storage, especially in droughts. The demand to storage ratio for both basins indicates that storage can satisfy a fraction of the annual demand. Finally, the discharge to demand ratio measures how much the river basin's flow is used or appropriated where the Bear River's demand is one and a half times its annual discharge and the Weber's demand is about half of its annual discharge.

Table 03.2: Summary for the number of attributes with input data that apply to different instances (nodes or links) across the two WEAP different model instances

	Bear River model		**Weber River model**	
Value type	# Attributes	# Nodes/Links	# Attributes	# Nodes/Links
Descriptive	89	60	32	30
Numeric	281	197	231	152
Seasonal	36	32	14	14
Time series	24	24	27	27
Array	11	11	8	8
Total	441	324	312	231

Table 3.3: Comparisons of total discharge (headflows and reach gains), storage, and demand and their ratios between the Bear and Weber River Basins, USA. The Bear River model used here only includes Utah and Idaho downstream portion and did not include the upstream Utah and Wyoming portions.

Attribute	**Bear River Model**	**Weber River Model**
Annual discharge (acre-feet)	2,301,804	975,502
Storage (acre-feet)	1,657,044	551,240
Annual demand (acre-feet)	1,068,352	473,385
Storage/discharge (%) (regulated flow%)	72	57
Demand/storage (ratio)	0.64	0.86
Demand /discharge (ratio)	0.46	0.49

For the second use case that developed and tested scenarios of reservoir sedimentation, net evaporation, water conservation, and river headflows, we found that reliability to meet demand targets varied from 50% to 100% in the Bear and Weber models (Figure 3.5). A few demand sites, such as "Logan Potable" for urban demand in the Bear River model and "Wanship to Echo" for agriculture demand in the Weber are insensitive to any scenario changes and have a 100% reliability.

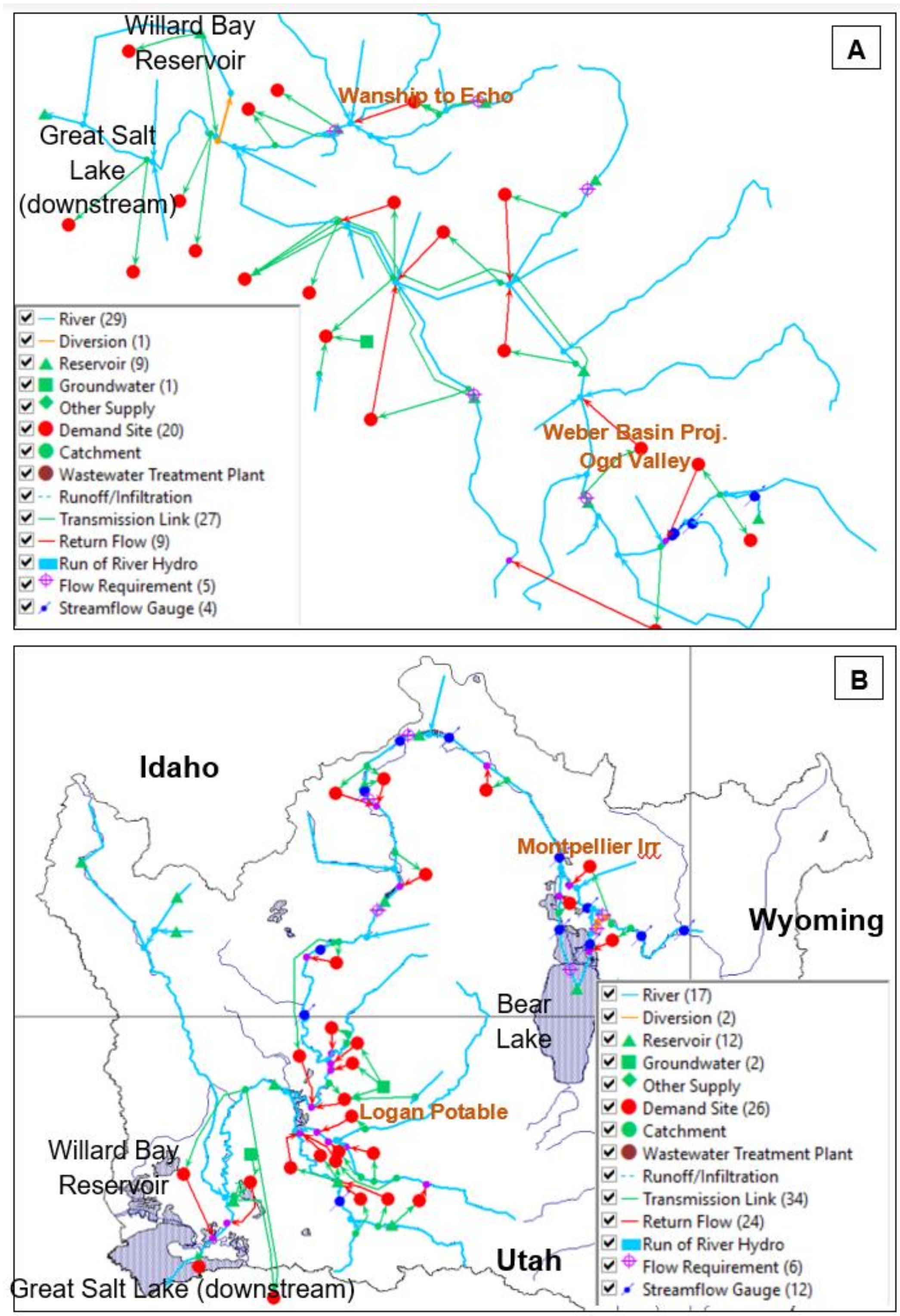


Figure 3.4: WEAP model schematics for the (a) Bear River Watershed (Utah and Idaho portions) and (b) Weber River Watershed (Utah). Both models end in the Great Salt Lake, Utah. Demand site names in orange are referenced in results.

Both sites have high water demand priority and thus their demand is met much earlier. Other demand sites such as Highline and Hyrum Canals are sensitive to changes especially to a reduction in river headflows, due to a mix of factors such as demand target volume, timing, and priority or demand fulfillment order in comparisons to other sites. Conserving or reducing demand seems to improve the demand reliability while reducing the capacity of the reservoirs or increasing their net evaporation changes demand reliability a small amount. The "Weber Basin Proj. Ogd Valley" site in the Weber model (Figure 3.5) is the only demand site out of 19 where reliability is sensitive to all scenarios. This site has one of the lowest demand priorities and is located upstream on the South Fork tributary of the Weber River.

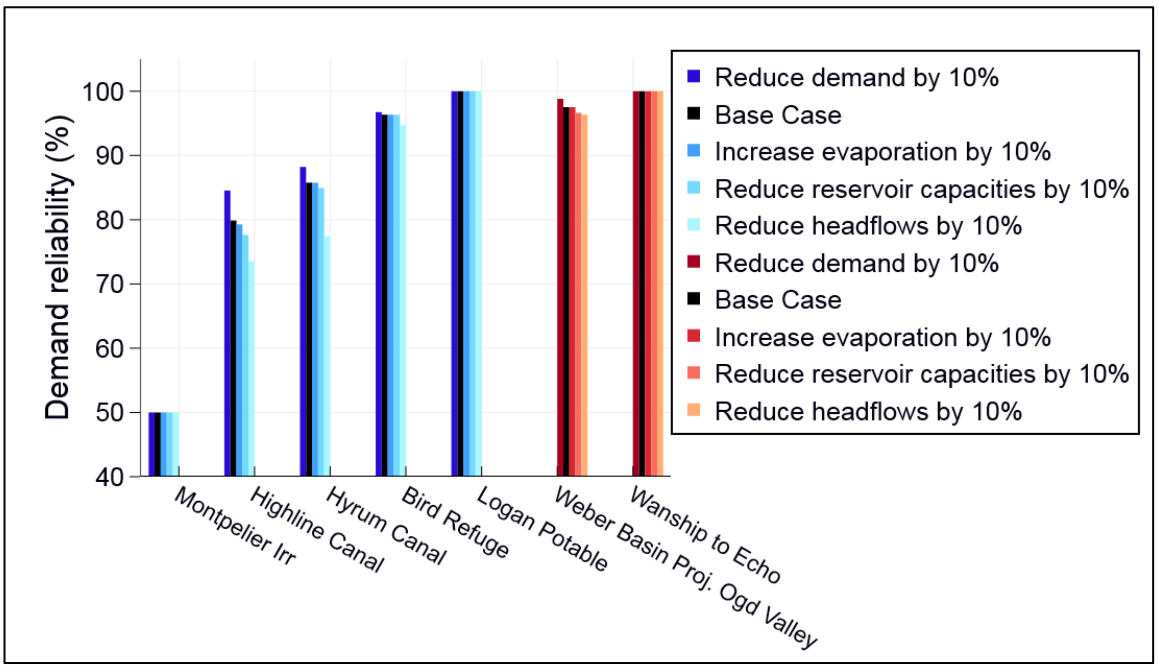


Figure 3.5: System reliability to meet demand targets across scenarios in the Bear (blue) and Weber (red) Rivers WEAP models.

3.5 Discussion

The presented open source Python-based software with its use of the WaMDaM database allows WEAP modelers automate keys steps to 1) extract data from the proprietary WEAP database and store and organize data in a single WaMDaM open source database, 2) query, compare, and analyze model data externally, 3) develop new scenarios, and 4) import data for the new scenarios, run multiple WEAP models and scenarios and compare results. The tools show the large volume of data needed to set up and run a systems model; without the tools modelers manually prepare most of this data. The tools also allow modelers to develop additional analysis such as estimating the total built storage in a basin and comparing that storage volume to annual discharge, estimate how regulated a river is, and compare the regulation factor with other rivers. For example, the Weber River and the Bear River Basins regulated-flow to storage ratios of 72% and 57% rank in the top 97% and 95% percentiles of ratios reported by Nilsson et al. (2005) for the largest 296 river basins in the world. Modelers can also automate and perform a much larger number of sensitivity analysis across models and scenarios.

The automated steps to export data from and input data to WEAP help improve the reproducibility of modeling workflows. The steps work across different numbers of models, scenarios, attributes, nodes, and links. The automated steps also cover all the types of data used by WEAP, including seasonal, time series, multi-column arrays for reservoir storage and elevation, numeric, and descriptive texts including equations. While WEAP provides version control of changes across a model, it is not easily possible for users to see what changed from one version to another. Thus, exporting the network and data of each WEAP version into WaMDaM allows a modeler to compare changes to each node, link, and their variables and metadata as demonstrated by (Abdallah and Rosenberg, 2019). WEAP users who use the software can also better manage their input data, track metadata among scenarios, reproduce WEAP model setup with consistent input data, prepare and populate the data with virtually no errors that otherwise can occur due to manual copy and paste. Automating these tasks allows modelers to focus more on analyzing and communicating results with their stakeholders.

By exporting WEAP model data to WaMDaM, modelers can make use of existing tools to auto-publish their water systems model data online, such as in HydroShare (Abdallah 2020). Publishing WEAP data into HydroShare allows others to discover the model data. Publishing also meets funding agency and journal requirements to manage data and allows for reuse of model data beyond one study. The model data for this study are available in HydroShare at Abdallah (2020).

3.6 Limitations and Future work

The presented software does not support customized additions to the default WEAP data structure. One custom example is a demand site with sub-groups for "Institutional" and "Manufacturing" sectors and within the Manufacturing sector further "Cooling" and "Process" categories. Such sub-groupings do not have a "node" typology in the WEAP API and thus an import/export script cannot access the group to extract data from WEAP or populate data into WEAP. The WEAP import and export functions also require the user to define the schematic and the general WEAP configuration parameters such as time step, water start year, units, simulation period, month type (calendar, or equal length) because these parameters cannot be automatically set through the WEAP API. The software is most useful when users enter data inputs as values and avoid mathematical operations. Even with these limitations, the software functions allow modelers to automate the import and export of large volumes of WEAP model data.

Future work should consider expanding the import and export functions to include user customized WEAP data branches and attributes outside the commonly used data structure. The WaMDaM Wizard also needs to be improved to be more user friendly, especially in handling potential errors and how users could solve them. Using the software on a larger inventory of the published WEAP models can develop further capabilities and expose limitations outside the Bear and Weber River model cases. Future work should consider extending the software to also include WEAP's sister model "Long-range Energy Alternatives Planning systems" (LEAP) which is widely used for energy-policy systems modeling (Heaps, 2012). LEAP and WEAP were both developed by Stockholm Environment Institute, and they have similar interface and APIs but for two different domains: water and energy. Both WEAP and LEAP can be coupled to transfer water modeling data into related energy simulations.

Another potential extension is to develop import and export functions for the widely used RiverWare water resources systems model. RiverWare does not have an API but it does have automated demand management interfaces to read input data from and output data to standardized Excel, CSV, and database files. Setting up another automated connection with WaMDaM requires time to code and test the import and export features for multiple models. The set up generally should start with a conceptual mapping between the new model and WaMDaM key metadata such as object types, attributes, nodes and links, scenarios, and data of different types.

3.7 Conclusions

This paper addressed the problem of developing general – rather than model-specific – data management systems and scripts to extract data out of the proprietary Water Evaluation and Planning system (WEAP) model, set up multiple models and scenarios, populate data back into WEAP, and run the large number of models and scenarios. We demonstrated the software on two different existing WEAP models and five scenarios for the Bear and Weber River Basins in Utah, USA.

Results show how the software facilitates comparison of input data across two different and originally separate WEAP models for the two basins. The comparison shows that the Bear River model has the larger network with 441 input data parameters. These results show size and complexity of the model and its granular coverage of systems components. The results also show the large effort needed to prepare and populate data into each model. The data automation presented here helps WEAP modelers prepare and load all parameters at once. Comparisons show how the Bear River has much less regulated flow than the Weber River. Such estimates can be applied on other WEAP models imported into WaMDaM to benchmark and compare rivers' flow-regulation around the world. These comparisons are reproducible and were possible because model data was extracted into the open-source WaMDaM database.

Results from running large number of model and scenario sensitivity analyses show that system reliability to meet demand is sensitive to changes in river headflows and demand. Reliability does not change much in response to increasing reservoir evaporation or reducing reservoir capacity by 10%. In the Bear River model, reliability to meet demand site delivery targets range from 50 to 100% across the scenarios while the Weber model has only one demand site that is sensitive to changes in reservoir sedimentation, water conservation, inflows, and evaporation.

The automated import and export tools allow WEAP users to quickly query model data for a river basin, present aggregate analysis for a basin such as storage/discharge and demand/ storage estimates, and help modelers compare and benchmark basin characteristics across river basins. Using this software allows WEAP modelers to spend more time on modeling and communicating results with stakeholders and less time to develop study-specific tools that cannot be reused by others.

Acknowledgments

This work was supported by the National Science Foundation through the CI-Water Project grant 1135482, iUtah grant 1208732, and Utah Mineral Lease Funds. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The Intermountain Section of the American Public Works Association, the Utah Chapter of the American Public Works Association, and the Utah Water Users Association provided additional financial support.

References

Abdallah, A., 2019. Chapter 4: An Interoperable Software Ecosystem To Store, Visualize Online, And Publish Water Resources Systems Modelling Data. Utah State University.

Abdallah, A., 2020a. Bear and Weber Rivers WEAP Models Input Data. https://www.hydroshare.org/resource/fc2855fee61d40629d34f4f7db5a1ba2/

Abdallah, A.M., 2020b. WaMDaM Jupyter Notebooks.http://doi.org/10.5281/zenodo.1484581

Abdallah, A.M., Rosenberg, D.E., 2019. A data model to manage data for water resources systems modeling. Environmental Modelling & Software 115 113-127.

Craven, J., Angarita, H., Corzo Perez, G.A., Vasquez, D., 2017. Development and testing of a river basin management simulation game for integrated management of the Magdalena-Cauca river basin. Environmental Modelling & Software 90 78-88.

Dieter, C., Maupin, M., Caldwell, R., Harris, M., Ivahnenko, T., Lovelace, J., Barber, N., Linsey, K., 2018. Estimated use of water in the United States in 2015: US Geological Survey Circular 1441, 65 p.

Gao, J., Christensen, P., Li, W., 2017. Application of the WEAP model in strategic environmental assessment: Experiences from a case study in an arid/semi-arid area in China. Journal of Environmental Management 198 363-371.

GoldSim Technology Group LLC, 2014. GoldSim User's Guide Version11.1.

Graf, W.L., Wohl, E., Sinha, T., Sabo, J.L., 2010. Sedimentation and sustainability of western American reservoirs. Water Resources Research 46(12).

Hammond, M., 2020. Python for Windows (pywin32) Extensions. https://github.com/mhammond/pywin32

Heaps, C.G., 2012. Long-range energy alternatives planning (LEAP) system. Somerville, MA, USA: Stockholm Environment Institute.

Hill, R.A., Hawkins, C.P., Jin, J., 2014. Predicting thermal vulnerability of stream and river ecosystems to climate change. Climatic change 125(3-4) 399-412.

Jamshid, M.S., Anzab, N.R., Asl-Rousta, B., Kim, J.H., 2017. Multi-Objective Optimization-Simulation for Reliability-Based Inter-Basin Water Allocation. Water Resources Management 31 3445.

Jones, S.C., 2019. Utah Regional Municipal and industrial Water Conservation Goals.

Kopytkovskiy, M., Geza, M., McCray, J.E., 2015. Climate-change impacts on water resources and hydropower potential in the Upper Colorado River Basin. Journal of Hydrology: Regional Studies 3 473-493.

Loucks, D.P., Van Beek, E., Stedinger, J.R., Dijkman, J.P., Villars, M.T., 2005. Water resources systems planning and management: an introduction to methods, models and applications. Paris: UNESCO.

Mehta, V.K., Young, C., Bresney, S.R., Spivak, D.S., Winter, J.M., 2018. How can we support the development of robust groundwater sustainability plans? California Agriculture 72(1) 54-64.

Mitchell, N., Kumarasamy, K., Cho, S.J., Belmont, P., Dalzell, B., Gran, K., 2018. Reducing High Flows and Sediment Loading through Increased Water Storage in an Agricultural Watershed of the Upper Midwest, USA. Water 10(8) 1053.

Mourad, K.A., Alshihabi, O., 2016. Assessment of future Syrian water resources supply and demand by the WEAP model. Hydrological Sciences Journal 61(2) 393-401.

Nilsson, C., Reidy, C.A., Dynesius, M., Revenga, C., 2005. Fragmentation and flow regulation of the world's large river systems. Science 308(5720) 405-408.

Rosenberg, D., 2017. Bear River WEAP Models (2012 and 2017).

Rosenberg, D.E., Fillion, Y., Teasley, R.L., Sandoval-Solis, S., Hecht, J.S., Zyl, J.E.v., McMahon, G.F., Jeffery S. Horsburgh, Kasprzyk, J.R., Tarboton, D.G., 2020. The Next Frontier: Making Research More Reproducible. Journal of Water Resources Planning and Management.

Sanvicente-Sánchez, H., González, E., Patiño, C., Villalobos, A., 2009. Surface water management model for the Colorado River Basin, Water resources management V. Proceedings of the Fifth International Conference on Sustainable Water Resources Management, Malta, 2009. WIT Press, pp. 35-46.

Stagge, J.H., Rosenberg, D.E., Abdallah, A.M., Akbar, H., Attallah, N.A., James, R., 2019. Assessing data availability and research reproducibility in hydrology and water resources. Scientific Data 6 190030.

SWCA Environmental Consultant, 2013. Final Great Salt Lake Comprehensive Management Plan and Record of Decision.

Tarboton, D.G., Idaszak, R., Horsburgh, J.S., Heard, J., Ames, D., Goodall, J.L., Band, L., Merwade, V., Couch, A., Arrigo, J., 2014. Hydro share: Advancing collaboration through hydrologic data and model sharing, 7th International Congress on Environmental Modelling and Software, iEMSs 2014. International Environmental Modelling and Software Society.

Tesfatsion, B.K., Rosenberg, D.E., 2013. Evaluating Storage Carryover in the Weber River Basin Using the Water Evaluation and Planning (WEAP) System, Water Resources Systems Analysis through Case Studies, pp. 102-113.

Welsh, W.D., Vaze, J., Dutta, D., Rassam, D., Rahman, J.M., Jolly, I.D., Wallbrink, P., Podger, G.M., Bethune, M., Hardy, M.J., Teng, J., Lerat, J., 2013. An integrated modelling framework for regulated river systems. Environmental Modelling & Software 39(0) 81-102.

Winter, J.M., Young, C.A., Mehta, V.K., Ruane, A.C., Azarderakhsh, M., Davitt, A., McDonald, K., Haden, V.R., Rosenzweig, C., 2017. Integrating water supply constraints into irrigated agricultural simulations of California. Environmental Modelling & Software 96 335-346.

Yates, D., Sieber, J., Purkey, D., Huber-Lee, A., 2005. WEAP21—A Demand-, Priority-, and Preference-Driven Water Planning Model. Water International 30(4) 487-500.

Zagona, E.A., Fulp, T.J., Shane, R., Magee, T., Goranflo, H.M., 2001. RiverWare: A Generalized Tool for Complex Reservoir System Modeling. JAWRA Journal of the American Water Resources Association 37(4) 913-929.

CHAPTER IV

AN INTEROPERABLE SOFTWARE ECOSYSTEM TO STORE, VISUALIZE, AND PUBLISH WATER RESOURCES SYSTEMS MODELLING DATA

Abstract

Water modelers often develop and use software tools to store, query, visualize, and share their data. Developing these tools is time-consuming, requires programming experience, and is model specific. This paper presents an interoperable software ecosystem of independently developed, state-of-the-art open-source data storage, web visualization, and repository tools to systematically set up scenarios, update input data, compare model networks and outputs, discover data inputs, and visualize and publish data and models online. Use for two models for the Bear River Watershed, United States and one model for Monterrey, Mexico show different spatial extents and depths for the modeling networks, differences in modeled urban and agricultural water demand patterns, and how the models respond to population growth and conservation. The software ecosystem makes it easier for researchers and stakeholders to discover, use, reproduce, extend, and build new water resources systems models. We welcome contributions of new open-source tools to expand the software system functionality.

Keywords

Systems analysis, OpenAgua, WaMDaM, Hydra Platform, open source, HydroShare

Highlights

- Modelers can use the software ecosystem to visualize, compare, edit, run, and publish data, models, and scenarios for multiple systems models.
- The software ecosystem encourages reuse of tools and reproducibility of analysis
- We welcome additional tools to expand software system functions

Software availability

Name of software: The WaMDaM Wizard

Developer: Adel M. Abdallah

Contact: Adel M. Abdallah; 8200 Old Main Hill, Logan, UT 84322, USA; Email amabdallah@aggiemail.usu.edu

Year first available: 2019

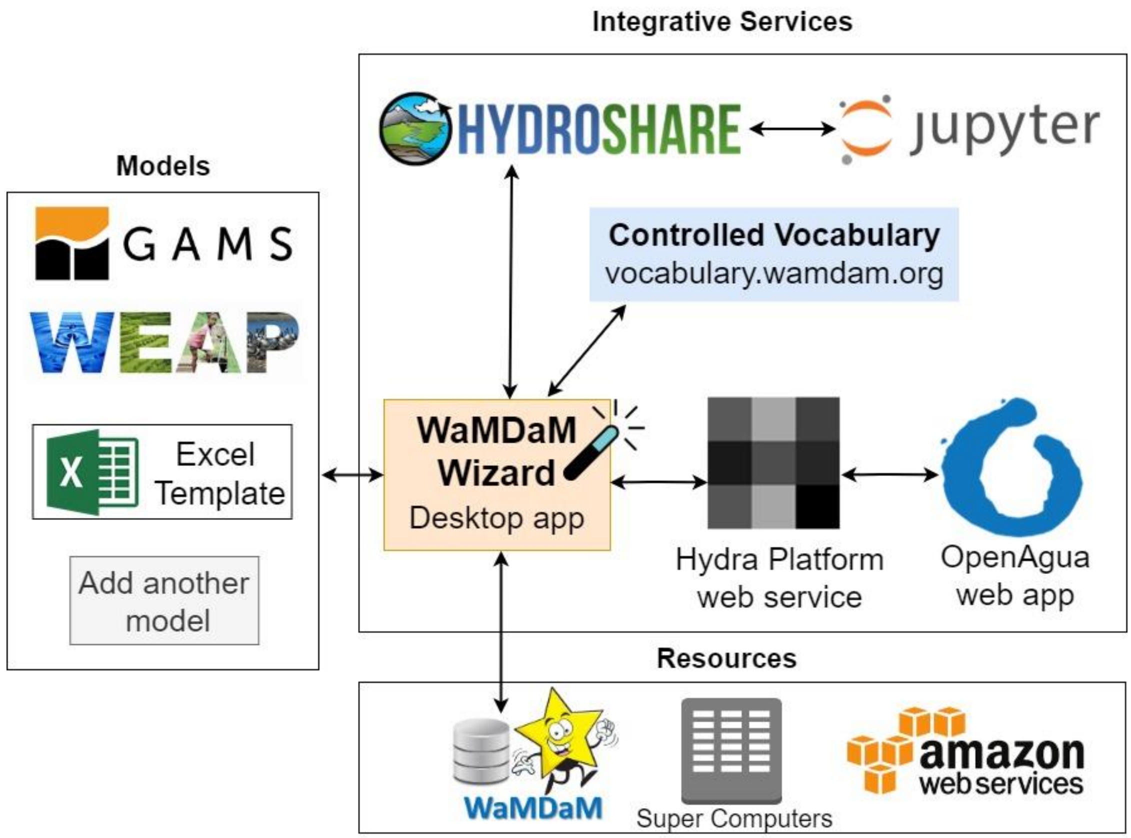
Required hardware and software: The WaMDaM Wizard executable is available for use with Microsoft Excel (2007 and later versions) and SQLite3 on Windows 64-bit computers. Hydra Platform web services are hosted by OpenAgua, which is available online on any browser. HydroShare is available online.

Input data and directions: Documentation of all source code, datasets, use cases, and instructions to use the ecosystem and replicate results are available on GitHub. Jupyter Notebooks can be executed on a local machine or run on the cloud using MyBinder service https://github.com/WamdamProject/WaMDaM_JupyterNotebooks/blob/master/2_VisualizePublish/00_WaMDaM_Directions_and_Use_Cases.ipynb

Programming languages: Python and Structured Query Language (SQL).

Cost and license: Free. Software and source-code are released under the New Berkeley Software Distribution (BSD) 3-Clause License, which allows for liberal reuse.

Graphical Abstract



4.1 Introduction

Over the last half century, the water resources systems analysis community has made significant advancements to improve the modeling of interrelated natural and built water resources infrastructure and inform decisions regarding system planning and management (Maass et al., 1962; Rosenberg and Madani, 2014; Brown et al., 2015). Systems models represent mass-balance interactions between supply and demand components and have been widely used to support water resource systems analysis. Despite modeling advances, modelers face technical challenges to develop and use these models. First, systems modelers must manage and store input and output data and track metadata. Second, they need to set up socio-economic and infrastructure management scenarios and track differences in input and output data (Abdallah and Rosenberg, 2019). Third, modelers need to visualize water system components and their connectivity as nodes and links. Fourth, modelers must plot input and output data to communicate model results and engage stakeholders with minimum technical difficulties (Brown et al., 2015). Fifth, modelers are increasingly required by funding agencies and journals to publish the final modeling data and results to support reproducible science (Rosenberg and Watkins, 2018; Stagge et al., 2019; Rosenberg et al., 2020). Currently, most systems modelers use or develop separate, model-specific tools for each of these tasks. Developing these tools is time-consuming and requires programming experience. Modelers would benefit from generalized tools that can store data, visualize and compare results, and publish data for many datasets and models. These tools should be reusable, independent of any specific software or model, require minimal programming, and be open-source should users want to modify or extend software functions.

Traditional water resources systems models such as the Water Evaluation and Planning system (WEAP) (Yates et al., 2005), RiverWare (Zagona et al., 2001), HEC-ResSim (HEC, 2007), GoldSim (GoldSim Technology Group LLC, 2014), eWater Source (Welsh et al., 2013), Aquatool (Andreu et al., 1996), EPANET (Rossman, 2000), REALM (Perera et al., 2005), and the Stormwater Management Model (SWMM) (Rossman, 2010) provide data storage, data visualization, and results computation features in a tightly-coupled software architecture. For example, WEAP and Riverware store data using proprietary database methods, as comma-separated-values (CSV), or as data management interface files. Modelers often use the software's graphical user interface (GUI) to manually enter and access data while a few models, like WEAP, offer an Application Programming Interface (API) that allows programmatic access to its data. Most models have their own model engine which is one or more simulation or optimization algorithms to execute using input data (Loucks et al., 2005; Knox et al., 2014; Meier et al., 2014; Knox et al., 2018). Often, traditional software tools are proprietary such as WEAP, RiverWare, HEC-ResSim, GoldSim, eWater Source, and Aquatool and may require paid licenses to use. Other systems modeling software such as EPANET (Rossman, 2000), REALM (Perera et al., 2005) and the Stormwater Management Model (SWMM) (Rossman, 2010) are open source, free to use, but have specific user interfaces or input file formats. No software system can publish standardized data and associated metadata to online repositories. While each software has a community of users and these communities will likely continue to flourish. This heterogeneity among models reveals why it is difficult to reuse any of their data storage, visualization, or computational components. Additionally, sharing, publishing, or transferring data to another model may require significant effort to first understand the proprietary data structures and then write data export and import functions for each model (Abdallah and Rosenberg, 2019). Further, traditional software often requires installation on local machines, which adds a barrier to engage stakeholders like water resources managers, who often look to inspect models input, network configuration, and visualize results of interest (Alminagorta et al., 2016a).

Systems modelers and researchers also need to develop novel models with capabilities beyond traditional models (Lund et al., 2013; Alminagorta et al., 2016a; Kok et al., 2018; Lee et al., 2019; Alafifi and Rosenberg, 2020). Researchers often spend the time to prepare input data, develop algorithms, and recreate other data storage, visualization, and analysis features within their modeling environment even though other models support similar features. These modelers often use simple methods to manage data like Excel and text files (Sehlke and Jacobson, 2005; Alminagorta et al., 2016b). Using Excel allows modelers to easily access data but often requires the author to help others query or interpret data values. One reason for such difficulties to interpret and reuse these model files is because the files have limited or no metadata, are intended to be read by a computer rather than a person, and are intended to be used as input for a specific model in a specific location. This specificity can make model coupling and reuse difficult.

In the broader field of hydrology, researchers have developed a loosely-coupled and interoperable software architectures such as OpenMI to couple hydrologic components such as snowmelt, runoff, and infiltration processes (Elag and Goodall, 2013). Each modeling component exchanges inputs and outputs defined across space and time with the other components using a standardized data coupling interface, shared vocabulary, and data exchange functions (Moore and Tindall, 2005). The HydroCouple interface extends

OpenMI to include geospatial data formats and support simulation on high-performance computers (Moore and Tindall, 2005; Buahin and Horsburgh, 2018). The Community Surface Dynamics Modeling System (CSDMS) (Peckham et al., 2013) provides an environment to couple earth surface models using a common programming language interpreter and shared vocabulary. Example coupling methods developed a web-service approach to couple components of the TopoFlow spatially distributed hydrologic model (Jiang et al. (2017). Zhang et al. (2019) introduced a service-oriented wrapper system for geo-analysis models for gridded modeling such as the Soil and Water Assessment Tool (SWAT) model and the Unstructured Grid Finite Volume Community Ocean Model (FVCOM). In such software or component coupling, the output data of a model is used as input data for another where each model still uses its own model-specific database. Therefore, such component-based coupling methods are mainly used to couple hydrologic models. These methods execute in sequence without archiving model data. While hydrology models use gridded data, systems models, represent reservoir, diversion, irrigation, municipal, hydropower, return flow, groundwater, river reach, and other components as nodes or links (Harou et al., 2010; Knox et al., 2014; Knox et al., 2019).

Here we build interoperability between four independently developed, active, existing open-source software tools for water resources systems modeling. These four software frameworks are:

1. The Water Management Data Model (WaMDaM) with its defined metadata and use of controlled vocabulary to enable data query and comparisons across models and datasets (Abdallah and Rosenberg, 2019),

2. Hydra Platform, which allows users to encode and communicate systems modeling data over the web using a web services approach (Knox et al., 2014; Knox et al., 2019),
3. OpenAgua, a web-based application that lets users collaboratively visualize and edit model networks (Rheinheimer et al., in review), and
4. HydroShare that supports researchers to publish and discover water-related datasets and modeling data (Tarboton et al., 2014).

We connect the tools into a software ecosystem (Jansen et al., 2009) by defining common data sharing functions and their equivalent vocabularies. The software ecosystem can assist modelers to perform the following three key tasks: i) organize and store water systems modeling data with metadata and controlled vocabularies, ii) visualize, edit, and compare networks, datasets, and scenarios in an online application, and iii) publish systems modeling data with contextual metadata to enable data discovery and analysis. These tasks allow modelers to engage stakeholders, reproduce analyses, and meet journal and funder data management requirements. The software ecosystem serves both existing proprietary and novel models. Below, we define three use cases that motivate and illustrate the software ecosystem. Three subsequent sections describe the software ecosystem components -- WaMDaM, Hydra Platform, and OpenAgua, and HydroShare -- their coupling, and application in the Bear River Watershed USA and the Monterrey metropolitan area, Mexico. The final sections present use case results, limitations, recommendations, and invite community involvement to grow the software ecosystem.

4.2 Use Cases

The software ecosystem components focus on supporting three steps that modelers commonly follow to develop models: i) store model data, ii) visualize networks and plot data, and iii) publish data to enable their discovery. Three use case questions guide the software ecosystem work:

1. How are networks and their data similar and different for different models in the same study area?
2. How do water management scenarios in two different models of the same study area compare?
3. How do the values for an input data parameter compare in two published modeling datasets?

The answers to these questions address modelers' needs to visualize their model networks, visualize system data, verify data input, and engage stakeholders. These questions also address modelers needs to change model input data, run models, and then visualize output data across scenarios and models. Presently, these steps are often manual and specific to the model's input data file structure.

4.3 Software ecosystem

4.3.1 Components

Here we describe four existing, generic, open-source software components that provide some of key modeling features (Table 4.1). First, WaMDaM is a well-defined data and metadata management framework with software tools to load, relate, and compare data for many systems modeling data (Abdallah and Rosenberg, 2019). The WaMDaM Wizard is an open-source Python and desktop-based software that helps users load and query data from a WaMDaM SQLite database (Abdallah and Rosenberg, 2019). Second, Hydra Platform framework is an early example of a generic open-source user-interface coupled with a data-manager for systems water management data. Hydra Platform provides a web service approach to encode and communicate data between the three software components of storage, user interface, and models using a generic data storage system for networks and their data while not requiring metadata (e.g., source and method of data) (Harou et al., 2010; Knox et al., 2014; Knox et al., 2019). The software components that communicate with Hydra Platform are referred to as “client applications”. Third, OpenAgua is a client web-based application for collaborative modeling and visualization of water resources planning and management that uses Hydra Platform as its data storage system (Rheinheimer et al., in review). OpenAgua generically manages data for models where users can optionally add metadata and use terminology that describe each model. Fourth, HydroShare is the Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) online collaboration environment with web-services for sharing and discovering data, models, and code (Tarboton et al., 2014). HydroShare requires metadata according to the Dublin Core Metadata Initiative which describe digital resources (i.e., files) such as title, owner, coverage in space and time. HydroShare also creates a Digital Object Identifier (DOI) for the published resources (e.g., modeling data) so resources can be easily cited in journal publications and other documents. Both OpenAgua and HydroShare allow their users to make modeling networks and data publicly available online.

Table 4.1: Four generic, active, free, and open source software tools to manage, serve, visualize, and publish water resources data and models.

Component	**Purpose and use**	**Key strengths and capabilities**
Hydra Platform (Knox et al., 2014; Knox et al., 2019)	Data storage and web services to manage water resources systems networks	● Consistent storage facility for network topology and associated datasets ● A server that exposes all functionality as a web service to which applications can connect to access data. ● Utilities to import and export data from/to Excel, CSV, WaterML, and GDX for GAMS
OpenAgua (Rheinheimer et al., in review)	A web-based application for collaborative water systems modeling	● Uses Hydra Platform for data storage ● Online collaboration/sharing for modeling and scenario analysis ● Users can view and edit network structure (nodes and links), scenarios, and input data; connect with and run model engines, and view results through interactive graphs.
WaMDaM (Abdallah and Rosenberg, 2019)	A relational data model for water resources systems and supporting software to load data, organize and describe systems water management data	● Reconciles semantic and syntax differences across datasets and models through controlled vocabulary and contextual metadata. ● Supports scenario comparisons in topology, metadata, and data values ● Enables direct access to subsets of data and metadata ● The WaMDaM Wizard interface to load and export data
HydroShare (Tarboton et al., 2014)	Online collaboration environment for sharing data, models, and code.	● Cloud-based API services to publish and discover code, models, and data ● Supports social activities among its users to collaborate and comment on published data and search authors and their products ● Supports permanent data and model publications through DOIs

4.3.2 Coupling Components

We show the coupling of software ecosystem components in Figure 4.1. Together, the coupled components allow users to reuse components to store, visualize, compare, publish, and discover modeling data across many different models. Key connections shown by black arrows in Figure 4.1: i) move data from a WEAP model into the WaMDaM database, ii) export data from the WaMDaM database and upload data to Hydra Platform, iii) exchange data between Hydra Platform and OpenAgua, and iv) import data from Hydra Platform into WaMDaM (the oppostive direction of step ii). Final steps include v) publish data from WaMDaM into HydroShare, and (vi) use Jupyter Notebooks to query and analyze published datasets. Below, we describe the coupling of each pair of components.

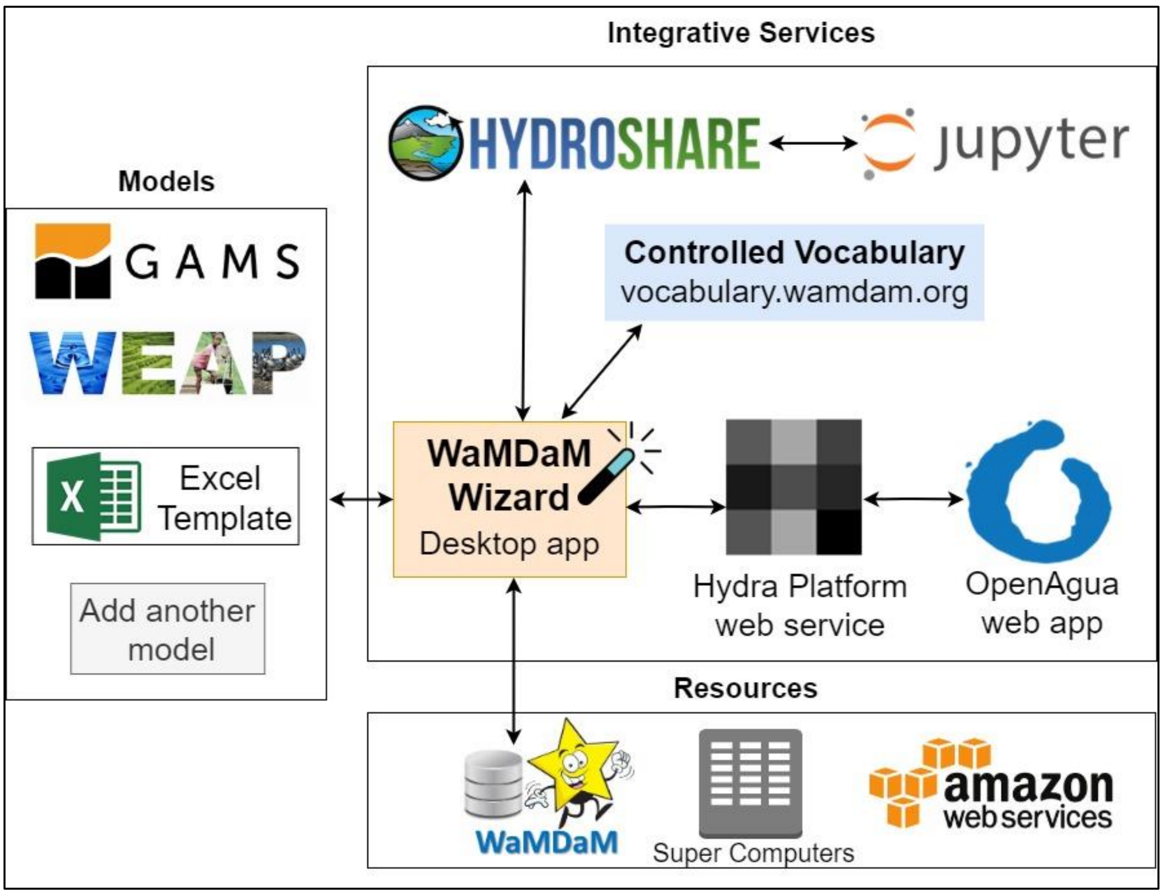


Figure 4.1: Coupling of independently developed components into a software ecosystem (black arrows). The software ecosystem allows users to extract or serve data from/to specific models, organize data with metadata and controlled vocabulary, visualize and edit networks online, download edited data, serve data to models, and publish model data online so it can be discovered.

4.3.3 Import model datasets to WaMDaM

Modelers can already organize and store their water management data in a WaMDaM SQLite database using the WaMDaM Wizard. The Wizard supports importing modeling datasets from a generic Microsoft Excel importer, WEAP models networks and their data using WEAP's API (Abdallah and Rosenberg, 2019), stream discharge time series data from the Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) hydrologic information systems web services or U.S. Bureau of Reclamation reservoir storage and releases time series data. Modelers can also export GAMS data for optimization models into CSV files and then use the generic Excel importer to import the data into a WaMDaM database (Abdallah and Rosenberg, 2019). Each model dataset that is connected to the WaMDaM database can in turn be connected to the other ecosystem components as described in the following sub-sections.

4.3.4 Export from WaMDaM to Hydra Platform

After importing modeling data into WaMDaM, modelers may need to move data to other tools such as Hydra Platform to take advantage of dependent online client applications such as OpenAgua. WaMDaM and Hydra Platform manage data for the same shared domain of water resources systems modeling. However, they have different motivating use cases and thus have different designs. WaMDaM organizes data and metadata and uses controlled vocabularies to relate terms across many models and datasets, whereas Hydra Platform provides generic storage and web-service approach that supports client applications (e.g., GUI) for systems models.

To couple WaMDaM and Hydra Platform, we first identified and mapped the equivalent common tables and contents between WaMDaM and Hydra Platform where each of them has different terms to describe the same metadata item (Table 4.2). Hydra Platform handles users' login information and has the concept of a project where users can collaborate on one or many networks for the same model. In contrast, WaMDaM uses controlled vocabularies to relate synonymous object types, attributes, and instances across models and supports the reuse of explicit metadata of sources, methods, organizations, and people that describe data. Hydra Platform only allows changing data values among scenarios for the same network while WaMDaM allows changing both the network and data values as part of scenarios. Thus, two scenarios in WaMDaM with differences in nodes and links for the same master network will be stored in Hydra Platform as scenarios in two separate networks.

Finally, we wrote a Python script to export WaMDaM data to Hydra Platform. The script is run from the WaMDaM Wizard and i) connects to a WaMDaM SQLite database that is previously populated with data for many systems models. Each model may have many networks, and each network may have many scenarios. Then ii) under the "Visualize and Publish" tab in the Wizard, the user clicks "Hydra Platform/OpenAgua" and fills out login credentials to their Hydra Platform account. Users then iii) upload WaMDaM data into an existing project in Hydra Platform or add a new project. Next, users iv) choose a resource type in the WaMDaM database (e.g., model or dataset name) to visualize its network and data. Then users v) choose a network for the model and vi) choose one or many scenarios inside the network. Finally, the user clicks "upload." The Python script calls the Hydra Platform web service to add a project, uses a SQL script to query the

WaMDaM database for the data to populate into each equivalent Hydra Platform table. The script then calls the “add_attribute”, “add_template”, and “add_network” methods to add new attributes, add a new resource type (i.e., a model, which includes all object types, their attributes), and add the network which includes all nodes and links and their scenarios

Table 4.2: Mapping common key equivalent metadata elements between WaMDaM and Hydra Platform

WaMDaM	Hydra Platform	Common description
Resource type	Template	A container or collection of many types of water system components that represent a model (e.g., WEAP)
Object Type	Template Type	A system component type (e.g., reservoir, demand site)
Attribute	Attribute	A property of a system component that takes data values
Instance	Resource	A specific implementation of the system component in space (e.g., Hyrum Reservoir)
Object Typology	Resource Type	The type of a system component instance as node or link
Mappings	Resource Attribute	A Bridge table that allows attributes of systems components types to be associated with many resources or instances
Scenario	Scenario	Contains and relates all data values within a network for a specific socio-economic, operation, physical, or other model set up
Scenario Mapping	Resource Scenario	A Bridge table that allows one scenario to be associated with many values for instances and their attributes
Master Network	Network	Contains many scenarios for a network in a study area
Attribute data type	Attribute data type	The structure of data as time series, array, numeric, text
Data Value	Dataset	A specific data entry with a relation with the above metadata

4.3.5 Export from Hydra Platform to OpenAgua

Once the modeling data are successfully uploaded into Hydra Platform, they automatically become available to Hydra Platform client applications such as OpenAgua that communicate with the Hydra Platform API to store, manage and retrieve data. The API calls include ‘login_user,’ ‘get_network,’ and ‘update_scenario.’ The OpenAgua web

API is written in Python, exposed to end users via a graphical user interface, and includes a wide range of functions that extend the core data management capabilities of Hydra Platform (Rheinheimer et al., in review). By design, any network, scenario, data value, metadata that is created in OpenAgua can also be accessed using Hydra Platform's web service functions. Hydra Platform manages users and login credentials; OpenAgua extends this user management capability with a public user whereby projects and networks can be publicly visible to any OpenAgua user.

4.3.6 Connect Hydra Platform to WaMDaM

Modelers may also want to export data hosted in Hydra Platform to WaMDaM. This export would allow further cross-model data queries and analysis that are not possible in Hydra Platform or OpenAgua and is the reverse of the connection described in Section 4.3.4. To export data from Hydra Platform to WaMDaM, open the WaMDaM Wizard and under the "Import Data To WaMDaM" tab, click the "Import From Hydra Platform" button. The user first provides their Hydra Platform account credentials to connect to the Hydra Platform server. Next, the user selects a project name, resource type (i.e., model name), a network, scenario, and a directory on the local machine to import the Hydra Platform data into a WaMDaM template Excel file. When the user clicks the "Import" button, the WaMDaM Wizard script calls in order the four main Hydra Platform web service functions: "get_template," "get_network," "get_scenarios", and "get_all_resource_data". These functions pull i) a list of all the object types and their attributes, ii) a list of nodes and links, iii) all the scenarios in the selected network and their metadata of start and end dates and time steps, and finally iv) a list of all attributes for the nodes and links and their data values of supported data types: time series, array, numeric, seasonal, and descriptors. Each call returns a JavaScript Object Notation (JSON) result which is parsed and mapped to the WaMDaM tables and their terminology. Modelers then can augment the Excel workbook template with additional metadata that are required by WaMDaM and may not be available in Hydra Platform such as source, method, people, and organizations for the nodes, links, and data values. If metadata is not known, modelers can define and reuse one generic metadata item (e.g., source) to all model data in WaMDaM. Modelers also can use controlled vocabularies in the workbook template to register native terms of datasets in Hydra Platform which allows the terms to be queried using controlled terms (Abdallah, Rosenberg 2019). Users then use the WaMDaM Wizard to load the imported Excel workbook into a WaMDaM SQLite database.

4.3.7 Connect WaMDaM and HydroShare

Modelers are increasingly required to publish their modeling data with contextual metadata that describe its content and coverage in space and time. Additionally, there is increasing need to provide programmatic access to read, query, and analyze published data. The WaMDaM database can contain modeling data from Hydra Platform, Hydra Platform-compliant applications like OpenAgua, model data sets such as for a WEAP model, or other data sources. To publish WaMDaM SQLite files into HydroShare as a "Composite Resource," we wrote a new Python script. The script is, accessed in the WaMDaM Wizard, harvests the Dublin Core Metadata from WaMDaM database, and uploads the SQLite file with the model data into HydroShare. More specifically, the script uses the "hs_restclient"

Python REST API library that allows programmatic access to publish and query HydroShare files and metadata (Black et al. 2019).

To use the script, a user should open the WaMDaM Wizard and connect to a WaMDaM SQLite database that contains the systems models, networks, and their data. The Wizard verifies that SQLite file complies with the WaMDaM schema including all tables and fields. Next, under the "Visualize and Publish" tab in the Wizard, the user clicks "HydroShare." The user then provides their login credentials to their HydroShare account, a title, abstract, and author name(s) for the new dataset publication. Finally, the user clicks "Publish". This button executes a Python script that auto queries the following generic and extended metadata from the SQLite file: i) temporal coverage from the modeling scenario(s) as the minimum start and maximum end dates, ii) spatial coverage box from the minimum and max latitude and longitude for nodes in the network(s), iii) list of resources type(s) (i.e., model names), unique object types, and attribute controlled vocabularies (if they exist), iv) network and scenario name(s), and v) list of sources, methods, people, and organizations metadata. The script adds three keywords to the created HydroShare resource: "WaMDaM, "systems models", and "water management." These keywords allow HydroShare users to discover the published dataset and other prior-published datasets. Next, the script calls HydroShare's "createResource" method to upload the WaMDaM SQLite file and all the above metadata into a private resource in HydroShare where users can edit metadata, share the resource with other HydroShare members, or make the resource public. Finally, HydroShare creates a DOI for permanent publication. Once the user makes the published resource public, modelers can use Jupyter Notebooks to programmatically access any of the published WaMDaM SQLite databases in HydroShare using WaMDaM's defined schema to query, analyze, and potentially reuse model data.

4.3.8 Additional coupling and testing

Coupling WaMDaM, Hydra Platform, and OpenAgua required each developer to make minor changes to their software. For example, we added a field named layout property in WaMDaM for object types to visualize a shape for each object type (e.g., reservoir icon as a triangle). We adopted OpenAgua's object type layout that encodes the icon shape and color in JSON format. We allowed the groupings of nodes and links in Hydra Platform to be optional--as opposed to required--to accommodate some WaMDaM data with no groupings. We added the source and method metadata fields to Hydra Platform and OpenAgua to match WaMDaM.

We required four scenario properties in WaMDaM so scenarios could be opened in OpenAgua: "ScenarioStartDate," "ScenarioEndDate," "TimeStep," and "TimeStepUnitCV." We also added two additional fields, ScenarioParentName and ScenarioType, to the Scenarios table in WaMDaM. "ScenarioParentName" explicitly maps scenario inheritance among scenarios as supported in Hydra Platform and OpenAgua. "ScenarioType" can take one of three potential values in OpenAgua: "Baseline," "Scenario," and "Results."

The "Baseline" type indicates the root (parent) scenario that can have many children scenarios. The "Scenario" type indicates a child scenario which can also be a parent to other scenarios. Each newly defined child scenario in Hydra Platform and OpenAgua references (i.e., reuses) the identical input data of its parent. Users can edit the new scenario's input data using the "Basic Data Input" editor in OpenAgua. Users can enter new values manually in the tabular format. The new values will be unique to the child scenario. The "Results" scenario type stores output values for a modeling scenario. This scenario type is used in OpenAgua to visualize and compare output datasets in the "Results Explorer." Finally, WaMDaM adopted the list of units used in Hydra Platform as a common controlled vocabulary. OpenAgua also adopted Hydra Platform units and added a unit conversion utility. These changes allow users to send WaMDaM data to Hydra Platform and OpenAgua to examine and edit scenarios online as well as send the data back to WaMDaM to run a model or publish the dataset. We anticipate that the coupling is the beginning of an update process where each software will continue to update to improve the user experience and accommodate more diverse use cases.

We validated the integrity of the import and export scripts to couple software ecosystem components by uploading the Bear River 2017 water allocation WEAP model (Abdallah, 2019) from WaMDaM to Hydra Platform and onto OpenAgua and then downloading the model dataset back into WaMDaM. We then used the WaMDaM Wizard scenario comparison tool to verify that both scenarios, the original in WaMDaM and the newly downloaded scenario form OpenAgua, were identical and no changes were unintentionally introduced in the upload or download mappings. Thus, modelers now can upload WaMDaM modeling data into Hydra Platform and use OpenAgua to visualize and edit data online. Users can also import models from OpenAgua into WaMDaM, run the model, and publish input or output results into HydroShare to enable data discovery, analysis, or to serve data to other models.

4.4 Application

We illustrate the numerous benefits of the software ecosystem with three use cases that include tasks to store, visualize, edit, publish, and compare modeling data for two models in the Bear River Watershed, USA and a third model for the Monterrey metropolitan area, Mexico. The first model is the Watershed Area of Suitable Habitat (WASH) optimization model that allocates water to maximize watershed habitat areas for the Lower Bear River Watershed (Utah portion) (Alafifi and Rosenberg, 2020). The WASH model uses the General Algebraic Modeling System (GAMS) engine which has no user interface. The second model is a WEAP simulation model that allocates water by water right priority within the Bear River Watershed (Utah and Idaho portions). WEAP has a proprietary database and does not support data publication. The third model is a water allocation model for the Monterrey metropolitan area, Mexico (Rheinheimer et al., in review). Both the WASH and WEAP models were developed from a predecessor 2010 Utah Division of Water Resources model for the lower Bear River basin that had a plain text input file and Fortran computational engine which was never run. The WASH model disaggregated irrigation demands within Cache Valley, Utah while the WEAP model extended the model domain upstream to Idaho and Bear Lake. The Monterrey model is stored in Hydra Platform within OpenAgua with no controlled vocabulary. The use cases assume a modeler has used the WaMDaM Wizard and loaded data for the three models into a WaMDaM SQLite database (Abdallah and Rosenberg, 2019) and that the user already has accounts for HydroShare and OpenAgua (free). Sharing the model input and output in public sites allows stakeholders to better access the modeling process and results.

The first use case exported the WEAP and WASH model data for the Bear River watershed from WaMDaM to HydraPlatform and onto OpenAgua. Then OpenAgua was used to visually compare the similarities and differences in the model networks in an online web browser. OpenAgua does not support running WEAP or WASH models online.

The second use case created new WEAP and WASH model scenarios in OpenAgua that increased and decreased annual urban water demand in Cache County, Utah by 25% from the base demand then exported the model data to Hydro Platform and on to WaMDaM to run the models. The model runs quantified the annual percent change in unmet demand at Cache County (WEAP model) and change in the suitable watershed area for aquatic, flood plain, and wetlands habitat for native Bonneville cutthroat trout fish, cottonwoods, and three indicator migratory bird species with differing needs for shallow, medium, and deep water habitat (WASH model). More specifically, scenario data were manually input and edited online in OpenAgua (Appendix C, Figure C1). The WaMDaM Wizard was used to download the new scenario data back to WaMDaM. Python scripts in Jupyter Notebooks (Abdallah and Rosenberg, 2019) were used to query the WaMDaM database for each model, serve the new scenario data into WEAP using its API write the .gms WASH input data file, execute both models, and read their results and store them in WaMDaM. Next, the WaMDaM Wizard was used to export scenario results for both models from WaMDaM to Hydra Platform and on to OpenAgua. Finally, OpenAqua "Results Explorer" utility was used to plot and compare the annual unmet demand across the baseline, conservation, and growth scenarios.

The third use case compares the magnitude and seasonality of agriculture water demand for the Monterrey metropolitan area, Mexico and the Bear River watershed in Utah then publish the datasets in HydroShare. For this use case, we uploaded each model dataset to HydroShare, then used a Python script in the Jupyter Notebooks to access and download the published SQLite files, query, and compare the controlled terms “Delivery target” for “Logan Irrigation” demand site in Utah and “Delivered flow” for agriculture demand for the “DR Bajo Rio San Juan” site in Mexico. Who choose sites in both models to compare the seasonality and magnitude of irrigated agriculture in two countries since they have comparable demand for irrigation.

4.5 Results

We present use case results to manage, visualize, edit, and publish water resources modeling data and results online.

Use Case 1: How are the networks of the WEAP and WASH models in the Bear River Watershed, USA similar and different?

Comparison of the two model networks (Table 4.3 and Figure 4.2) shows:

1. The WEAP model for the Bear River supports more water system components such as flow requirement, groundwater, and streamflow gage which are not explicitly supported in WASH. The common resource types between the models that use the same controlled vocabularies are “demand” and “dem”, “reservoir” and “v”, and Return Flow and “returnFlowExist” in WEAP and WASH respectively. WASH used the general node resource type “j” for any other network connection while WEAP is specific about the types such as “River Headflow” or “Diversion Outflow.” These results show the similarities and differences in the two models’ capabilities and a potential for input data

reuse or transfer between them in other watersheds (e.g., populate a new WASH model from an existing WEAP model),

2. The WEAP model has a larger number of object instances and covers a larger area upstream into the Idaho. For example, the WEAP model includes 20 demand sites within the same Lower basin area compared to 11 sites for WASH. More specifically, the WEAP model includes three urban demand sites for Cache County ("Logan Potable," "North Cache Potable," and "South Cache Potable") while WASH represents all of them in one node as "j3" that has a controlled term of "Cache County M&I." The reader can view these sites in OpenAgua (Figure 4.2).
3. The WEAP model also includes specific upstream supply and demand and storage especially Bear Lake (top half of the screenshot). In the WASH model, this part of the system is aggregated into a river headflow.
4. The Bear River WEAP model simulates demand reliability across 40 years of interannual monthly of dry, wet, and average water years compared to the WASH model which focuses on maximizing the watershed area for suitable habitat within a single year. Thus, the WEAP model could be useful to quantify cooperation scenarios between the Utah and Idaho states where downstream users in Utah could store water in Bear Lake in wet years and use it later in dry years.

Results for the use cases can be accessed as follows: Both the Bear River WEAP and WASH models are shared published in HydroShare (Abdallah, 2020b) (Appendix C-Figure C2).

Table 4.3: Key comparisons between WEAP and WASH models visualized in OpenAgua

Comparison element	**WEAP Model**	**WASH Model**
Water system component types	14 (see the list under “Resources” in Figure 4.2 -a)	6 (see the list under “Resources” in Figure 4.2 -b)
Geographic extent	Bear River (Utah, Idaho)	Lower Bear River (Utah)
Water system components	136	43
Time step and extent	monthly (1966 - 2006)	monthly (2003)

Use Case 2: What are the differences in WEAP and WASH model’s outputs in the face of water conservation and population growth scenarios in the Bear River Watershed?

Results follow the general expected trend that increased demand increases shortages while water conservation reduces shortages (Figure 4.3). There are four years, 1970, 1976, 1993, 1996, where water conservation completely eliminates shortages while shortages persist for the base case and increased demand scenario. In dryer years (e.g., 1987 to 1992 and 200 to 2004 where there is not enough water to meet site demand p), the conservation scenario reduces the magnitude of shortages compared to the baseline scenario. These results are also available online in OpenAgua for stakeholders to view and discuss without needing to install WEAP or purchase its license.

For the WASH model, the watershed area for suitable habitat for native vegetation, birds and fish in the baseline scenario (2003 hydrologic year) is estimated at 121,526 acres. Reducing Cache County urban demand by 25% would increase the WASH area by 144 acres while a 25% increase in the site’s demand would decrease the WASH area by 142 acres.

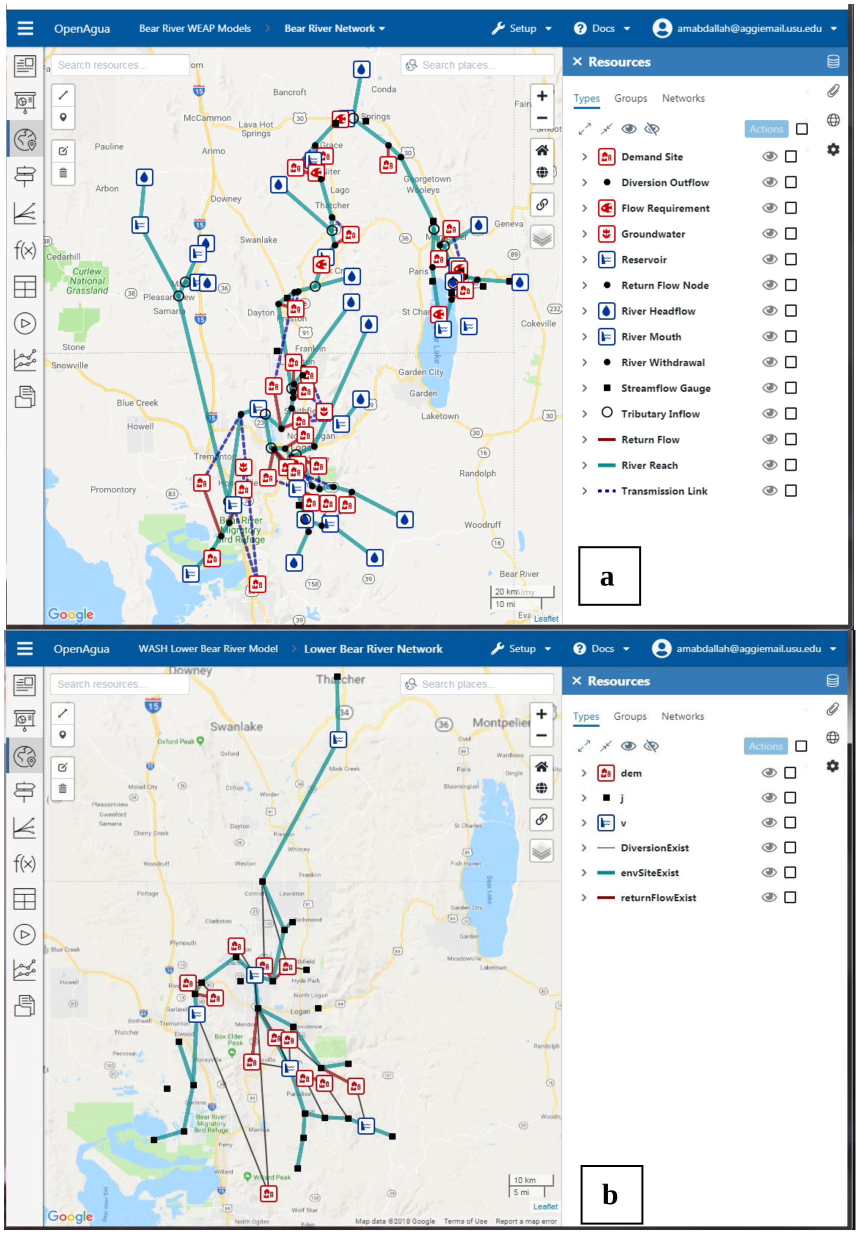


Figure 4.2: OpenAgua visualization of model networks for (a) Bear River WEAP simulation model 2017 (Utah and Idaho portion) and (b) Watershed Area for Habitat Suitability (WASH) optimization model (Utah portion). The models schematics and input data can be viewed and inspected online in OpenAgua under "Public Projects" (Rheinheimer, 2020)

This small increase or decrease in WASH area is because of the small influence of Cache County urban site which represents about 18% of the total annual agriculture and urban demand in this watershed of 415 million cubic meters (336,446 acre-feet). These results show the potential role of targeted urban water conservation and growth in improving or degrading suitable habitat areas in the watershed.

One very interesting comparison between WEAP and WASH models is that WEAP estimates 8%, 12%, 17% increase in demand shortage for Cache County urban site n 2003 (Figure 4.3, red box) while the WASH model meets completely satisfies demands (no shortages) for all three demand scenarios. If the WASH model could not meet the demands (constraints), the model would return an infeasible solution (Alafifi and Rosenberg, 2020). This discrepancy between the two models to meet demand at the Cache County urban site in 2003 is likely because of the two models different spatial extents and how they aggregate and disaggregate demand sites and upstream supplies (see Use case #1 results).

Use Case 3: How do the magnitude and seasonality of agriculture water demand in Monterrey metropolitan area, Mexico and Utah compare?

Results show on average that the monthly demand target for "Logan Irrigation" site in Utah is 1.8 cubic meter per second (cms) (black squares) compared to 0.15 cms demand (grey circles) for "DR Bajo Rio San Juan" agriculture demand site in Monterrey Mexico (Figure 4.4). Agriculture demand (i.e., crop growth) in Utah extends for six months and is much shorter than the 11 month irrigation season in Mexico. In Utah, agriculture demand begins in April, peaks in July, and ends in October. In Mexico, agriculture demand begins in December, peaks in April, and ends in October. It is unclear why the Mexico demand from June to October has two steps of increase and decrease. The two steps may represent switching to different crops or harvesting patterns. This comparison between two different models and counties was possible because of the software ecosystem interoperability and moving data between their systems. The Monterrey, Mexico water allocation model data can be accessed in HydroShare (Abdallah, 2020a).

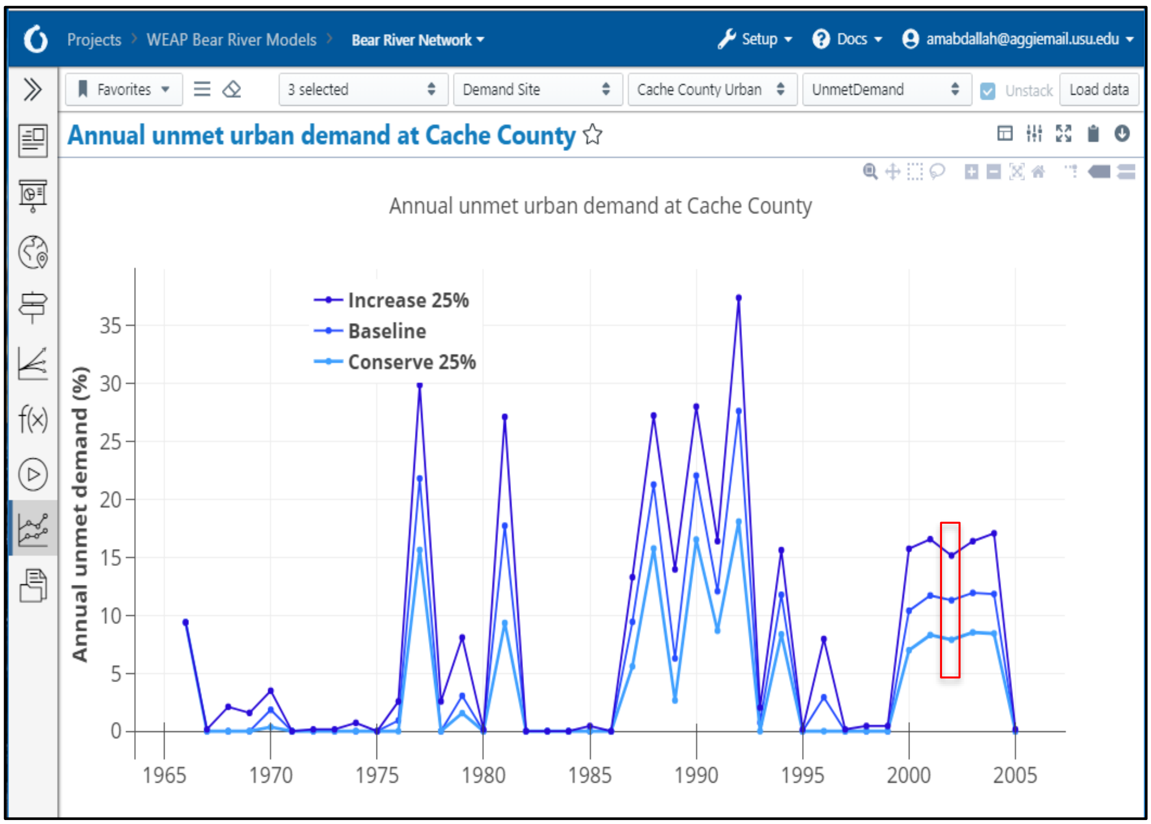


Figure 4.3: OpenAgua "Results Explorer" dashboard plot of annual shortage for Cache County, Utah as estimated in the WEAP model over the simulation period 1966-2006 for three demand scenarios. The red rectangle highlights unmet demands in 2003 that is the base year for the WEAP model.

All of the above results can be reproduced in Jupyter Notebooks (Abdallah, 2020c). While a WEAP software license is still required to run the Bear River model, stakeholders can use the OpenAgua tool (with a paid license) to examine the model input data and select results.

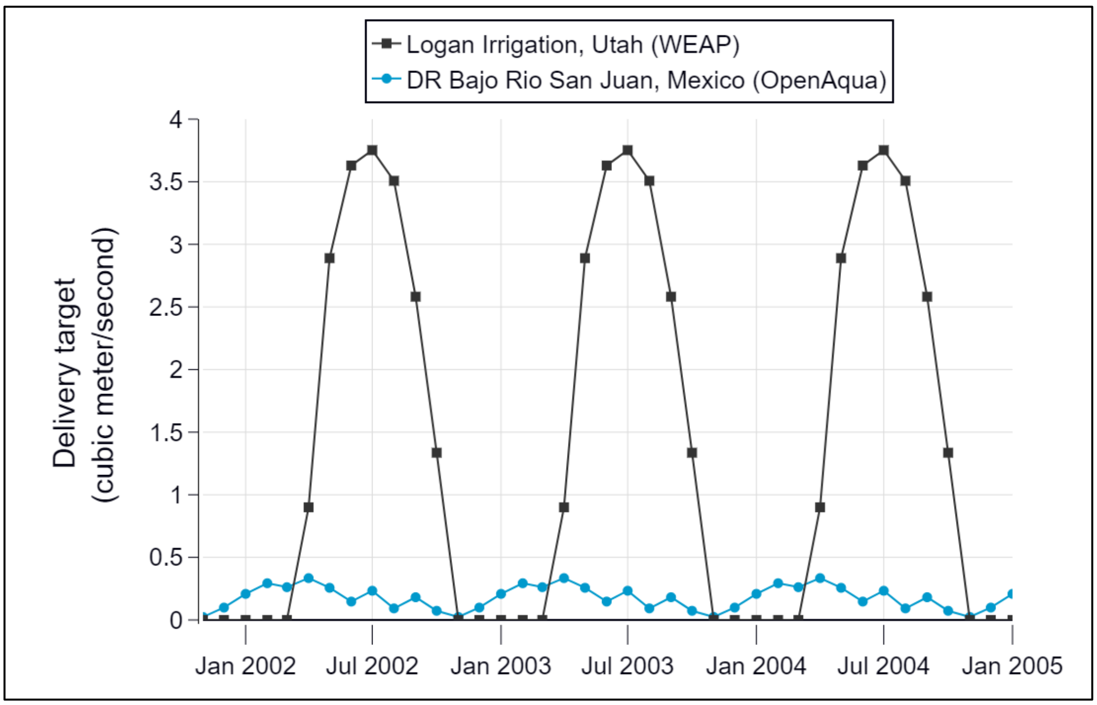


Figure 4.4: Comparing demand target for the Logan Irrigation, Utah (black) with "DR Bajo Rio San Juan" agriculture demand site in Monterrey, Mexico (grey).

4.6 Discussion

Connecting WaMDaM, Hydra Platform, OpenAgua, and HydroShare into a software ecosystem allows modelers to store, edit, run scenarios, visualize, and publish online water resources systems data. The software ecosystem facilitates the export of model data from one component to another to allow users to access data storage, analysis, visualization, and publishing features not supported by an individual component.

Together, these coupled features allow modelers to compare simulation and optimization models for the same modeling domain and different domains. For example, the WEAP Bear model supports more specific demand sites within the same area compared to WASH and thus WEAP offers more decision support analysis for each demand site. The WEAP Bear River model represents specific upstream supply and demand and storage compared to aggregated river headflows in WASH. Thus, the WEAP model could be useful to quantify cooperation scenarios between Utah and Idaho. The WEAP Bear River model includes 40 years of monthly supply data compared to a single year in this WASH model and thus the WEAP model would be more useful to simulate water allocations and potentially unmet demand under a spectrum of historic hydrologic years from dry to wet conditions. Reducing urban demand in Cache County in the Bear River WEAP model by 25% would reduce unmet demand relative to the base case including in dry years. The same software ecosystem tools and steps were used for a different model (WASH) to estimate the effect of decreases in Cache County urban demand by 25%. In the third use case, comparing Utah and Mexico agricultural demands from two models showed both agriculture demand sites from two different models for Utah and Mexico share high seasonal variability but have different growth seasons. The identified variability suggests the importance of water storage for both sites at different times when demand is low (winter) to use water later when demand is high such as Spring in Mexico and summer in Utah.

4.6.1 Advantages of using the software ecosystem tools

No model or software tool can do all data storage, scenario entry, visualization, comparison, stakeholder access, and publishing tasks well. The software ecosystem allows modelers to export their data to the software component that is best suited for the data or modeling task. Additionally, the software ecosystem allows users to construct workflows for tasks that cannot be done in any of the individual software system components. For example, the software ecosystem allows users to visualize and compare model data for many models and scenarios without being limited to one set of core object types and attributes as in WEAP and RiverWare. Users can also define model's scenarios online in OpenAgua then move data to WaMDaM to run the model and publish results. The ecosystem tools can help compare networks for the same basin in two different modeling software. These comparisons are facilitated by consistent data storage with metadata and controlled vocabularies in WaMDaM, Hydra Platform, and consistent visualization in OpenAgua.

The software ecosystem further allows each model and dataset to retain its native terms for object types and attributes. This feature allows users to view model data in OpenAgua and support broader stakeholder engagement. Stakeholders can inspect modeling networks and data using an internet browser without needing a paid license or to install software on local machines. This online setup provides users greater access to create new scenarios, edit and visualize input data using OpenAgua interface.

Researchers who develop novel models can use the software ecosystem components to manage their data, compare scenarios, and identify differences in networks, input, and output data without need to develop their own data management, online visualization, or publication features. Comparing datasets across novel and existing models will help researchers undertake benchmarking studies and distinguish similarities and differences in input data.

The automated publishing of water resources systems models and their data will make it easier for researchers and stakeholders to discover, use, reproduce, extend, and build new models. Sharing and publishing these models and datasets helps researchers fulfill data management requirements established by the National Science Foundation (https://www.nsf.gov/eng/general/dmp.jsp) and by journals (Rosenberg and Watkins, 2018; Stagge et al., 2019; Rosenberg et al., 2020). Sharing model datasets can increase the potential for their reuse, reduce the time to build models, and increase the value of water resources models within and outside the discipline. The use of the software ecosystem products by others can be measured by a simple discovery exercise in HydroShare: search resources for the keyword "wamdam". Currently, HydroShare returns six published WaMDaM datasets that are part of this work.

4.6.2 Limitations

There is a lot of work to do to improve the software ecosystem tools and coupling. There are still metadata and software-specific configurations and parameters to support. The WaMDaM Wizard is currently implemented on a local machine. Deploying WaMDaM in a cloud setting with web services coupling similar to OpenAgua, Hydra Platform, and HydroShare would make the WaMDaM Wizard less dependent on local computer configurations. Currently, OpenAgua users can only visually search for public models and networks, which will become difficult as the number of projects and networks grow with time. Currently, data and model discovery are limited to projects and networks in OpenAgua or SQLite databases and their metadata in HydroShare. HydroShare does not natively support data analysis on data within multiple SQLite files. It is also not easily possible to search with HydroShare for specific network, scenario, node, link, or attribute data that are contained within published SQLite files in HydroShare. Reproducing model results requires running the models. Running models may be difficult for models that must be installed on a desktop machine, require a paid license, and are operating system specific.

The current scenario parameters in WaMDaM and Hydra Platform use simple time steps such as day, month, or year. The software ecosystem might support complex time steps, such as leap years, number of days in the month, that are available in WEAP. Invariably, it will be difficult for software ecosystem developers to keep up with all the modifications and improvements that model developers make.

4.6.3 Future Work

Future versions of the software ecosystem should support geo-spatial search for individual water management infrastructure, its connectivity, and data. This feature can be added by building on the ability to search time-series data (e.g., HydroDesktop, Ames et al., 2012) and HydroClient (http://data.cuahsi.org/). WaMDaM support for controlled vocabularies would be particularly useful to search across different native terminology used in models and by users. This functionality could allow the CUAHSI web services to search for reservoir bathymetry curve, seasonal demand data, or network connectivity such demand sites supplied by a particular reservoir.

Both the Hydra Platform and OpenAgua development teams are currently exploring ways to integrate alternative database systems to accommodate "big data" that can result from many scenario analyses for large water networks. Instead of user login credentials to the Hydra Platform and HydroShare servers, future software implementations should consider using an API key-based approach to establish a reusable connection to the servers.

To remove user's needs to install software locally, future implementations of the WaMDaM Wizard should build read and write web services to a server-based database as an online application. Future work should provide tools to allow users to more easily provide metadata and register their native terms with existing controlled vocabularies besides the WaMDaM workbook Excel template. Future work should allow instantaneous interaction through an API between data storage in WaMDaM, visualization in OpenAgua and many simulation or optimization model engines. These needed improvements should also be paralleled with work to use the software ecosystem for more applications and models and connect additional tools to expand the ecosystem.

Finally, we note the need for continual alignment of development efforts, to help ensure ecosystem components remain inter-compatible over the long term. This continual alignment requires regular communications and code transparency between component projects even as each ecosystem component is developed independently. Addressing this challenge will require version control and strong documentation of respective tools.

4.6.4 Invitation for community involvement and feedback

We invite water modelers, analysts, students, faculty, professionals, managers, and other members of the water resources systems community to use the software ecosystem, provide feedback, and help develop new tools to expand the ecosystem. The current ecosystem is the product of feedback we received from collaborators, colleagues, workshop participants, and audiences since the inception of Hydra Platform, HydroShare, and WaMDaM in 2013 and OpenAgua in 2016. You can participate in multiple ways such as: 1) use the "wamdam" keyword to search and discover water systems datasets and models in HydroShare, 2) use the software ecosystem tools for your existing WEAP model, 3) use the WaMDaM wizard to link native vocabulary for your data sets(s) and model(s) to controlled vocabulary, 4) build exporters and importers to WaMDaM for your own custom model or dataset, and 5) build other interoperable software tools that will further your work and the work of others. For all of these steps, there will likely be bumps, hiccups, and surprises—if needed, contact us.

4.7 Conclusions

This paper addressed the problem of using many disconnected and often model-specific software tools to store, visualize, edit, run, analyze, and publish systems modeling data. We contributed a description, prototype, and demonstration of an interoperable set of open-source software tools (WaMDaM, Hydra Platform, OpenAgua, and HydroShare) that help modelers to i) store and organize data with metadata and controlled vocabularies in WaMDaM, ii) visualize, edit and compare model networks and their input and output data in an online application, and iii) publish systems modeling data and metadata to support data discovery and analysis.

Three use cases for two models in Utah and one model in Mexico 1) compared the networks of a WEAP simulation and WASH optimization models for the Bear River Watershed in Utah and Idaho, 2) identified differences in WEAP and WASH model's outputs between new water conservation and growth scenarios in the Bear River

Watershed, and 3) compared agriculture demand pattern for two sites in the WEAP model in Utah and the OpenAgua model in Mexico.

The coupled software tools enabled moving data from database to a model, to an online user interface, and to an online data repository. The coupled tools also allow comparisons between the three models in two different countries. The automated process to retrieve, query, compare, and visualize results for two different models in Utah and Mexico was possible because both of them are published publicly in HydroShare using the WaMDaM consistent schema.

This prototype of coupled software ecosystem aims to help modelers spend more time on modeling and less to develop specific tools for data storage, visualization, and publishing. We see the software ecosystem as a complement to existing models such as WEAP. WEAP has a unique useful capabilities and large user base and will flourish into the future. The software ecosystem offers a collaborative environment and additional tools to compare networks for the same basin in two different modeling software, set up and run multiple scenarios and models from an online portal, and automate the process to share and publish model data. Model datasets published in HydroShare can be discovered with the keyword “wamdam”, reproduced, and used in follow-on applications.

Future work should implement all components of the coupling software online to support use cases for instantaneous connection between WaMDaM and Hydra Platform. We invite water modelers, analysts, students, faculty, professionals, managers, and other members of the water resources systems community to use the software ecosystem, provide feedback, and help develop new tools to expand the ecosystem.

Acknowledgments

This work was supported by the National Science Foundation through the CI-Water Project grant 1135482, iUtah grant 1208732, and Utah Mineral Lease Funds. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The Intermountain Section of the American Public Works Association, the Utah Chapter of the American Public Works Association, and the Utah Water Users Association provided additional financial support. David Tarboton provided insights into connecting WaMDaM with HydroShare. We thank Hadia Akbar for using the WaMDaM Wizard, OpenAgua, HydroShare, and Jupyter Notebooks to reproduce results for each use case.

References

Abdallah, A., 2019. Chapter 3: Open Source Python Software to Manage, Populate, and Compare Weap Models and Scenarios. Utah State University.

Abdallah, A., 2020a. Monterrey Mexico model From OpenAgua. https://www.hydroshare.org/resource/e29c9283305045338be24a495c781ec9/

Abdallah, A., 2020b. WEAP and WASH Bear River Systems Models. https://www.hydroshare.org/resource/af71ef99a95e47a89101983f5ec6ad8b/

Abdallah, A.M., 2020c. WaMDaM JupyterNotebooks.https://zenodo.org/record/1484581

Abdallah, A.M., Rosenberg, D.E., 2019. A data model to manage data for water resources systems modeling. Environmental Modelling & Software 115 113-127.

Alafifi, A.H., Rosenberg, D.E., 2020. Systems modeling to improve river, riparian, and wetland habitat quality and area. Environmental Modelling & Software 126 104643.

Alminagorta, O., Rosenberg, D.E., Kettenring, K.M., 2016a. Systems modeling to improve the hydro-ecological performance of diked wetlands. Water Resources Research 52(9) 7070-7085.

Alminagorta, O., Rosenberg, D.E., Kettenring, K.M., 2016b. Systems modeling to improve the hydroecological performance of diked wetlands. Water Resources Research n/a-n/a.

Andreu, J., Capilla, J., Sanchís, E., 1996. AQUATOOL, a generalized decision-support system for water-resources planning and operational management. Journal of Hydrology 177(3) 269-291.

Brown, C.M., Lund, J.R., Cai, X., Reed, P.M., Zagona, E.A., Ostfeld, A., Hall, J., Characklis, G.W., Yu, W., Brekke, L., 2015. The future of water resources systems analysis: Toward a scientific framework for sustainable water management. Water Resources Research.

Buahin, C.A., Horsburgh, J.S., 2018. Advancing the Open Modeling Interface (OpenMI) for integrated water resources modeling. Environmental Modelling & Software 108 133-153.

Elag, M., Goodall, J.L., 2013. An ontology for component-based models of water resource systems. Water Resources Research 49(8) 5077-5091.

GoldSim Technology Group LLC, 2014. GoldSim User's Guide Version11.1.

Harou, J.J., Pinte, D., Tilmant, A., Rosenberg, D.E., Rheinheimer, D.E., Hansen, K., Reed, P.M., Reynaud, A., Medellin-Azuara, J., Pulido-Velazquez, M., Matrosov, E., Padula, S., Zhu, T., 2010. An open-source model platform for water management that links models to a generic user-interface and data-manager In: David A. Swayne, W.Y., A. A. Voinov, A. Rizzoli, T. Filatova (Ed.), International Congress on Environmental Modelling and Software, Modelling for Environment's Sake ed. International Environmental Modelling and Software Society (iEMSs) Ottawa, Ontario, Canada.

Hydrologic Engineering Center (HEC) of the US Army Corps of Engineers, 2007. Reservoir System Simulation (HEC-ResSim), 3.0a ed.

Jansen, S., Finkelstein, A., Brinkkemper, S., 2009. A sense of community: A research agenda for software ecosystems, 2009 31st International Conference on Software Engineering-Companion Volume. IEEE, pp. 187-190.

Jiang, P., Elag, M., Kumar, P., Peckham, S.D., Marini, L., Rui, L., 2017. A service-oriented architecture for coupling web service models using the Basic Model Interface (BMI). Environmental Modelling & Software 92 107-118.

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.

Knox, S., Meier, P., Yoon, J., Harou, J.J., 2018. A python framework for multi-agent simulation of networked resource systems. Environmental Modelling & Software 103 16-28.

Knox, S., Tomlinson, J., Harou, J.J., Meier, P., Rosenberg, D.E., Lund, J.R., Rheinheimer, D.E., 2019. An open-source data manager for network models. Environmental Modelling & Software 122 104538.

Kok, D.J.D., Pande, S., van Lier, J.B., Ortigara, A.R.C., Savenije, H., Uhlenbrook, S., 2018. Global phosphorus recovery from wastewater for agricultural reuse. Hydrol. Earth Syst. Sci. 22(11) 5781-5799.

Lee, S.H., Mohtar, R.H., Yoo, S.H., 2019. Assessment of food trade impacts on water, food, and land security in the MENA region. Hydrol. Earth Syst. Sci. 23(1) 557-572.

Loucks, D.P., Van Beek, E., Stedinger, J.R., Dijkman, J.P., Villars, M.T., 2005. Water resources systems planning and management: an introduction to methods, models and applications. Paris: UNESCO.

Lund, J., Josué Medellín, Samuel Sandoval, Wei Chu, Alvar Escriva, Ashley Vincent, Erik Porse, Prudentia, Zhinkalala, Timothy Nelson, Hui, R., 2013. Building Models from the Data Up: From Calvin to Hobbes.

Maass, A., Hufschmidt, M.M., Dorfman, R., Thomas, H.A., Marglin, S.A., Fair, G.M., Bower, B.T., Reedy, W.W., Manzer, D.F., Barnett, M.P., 1962. Design of water-resource systems.

Meier, P., Knoxa, S., Haroua, J.J., 2014. Linking water resource network models to an open data management platform, 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).

Moore, R.V., Tindall, C.I., 2005. An overview of the open modelling interface and environment (the OpenMI). Environmental Science & Policy 8(3) 279-286.

Peckham, S.D., Hutton, E.W.H., Norris, B., 2013. A component-based approach to integrated modeling in the geosciences: The design of CSDMS. Computers & Geosciences 53(0) 3-12.

Perera, B.J.C., James, B., Kularathna, M.D.U., 2005. Computer software tool REALM for sustainable water allocation and management. Journal of Environmental Management 77(4) 291-300.

Rheinheimer, D., 2020. OpenAgua: Collaborative water system modeling for a new generation. https://www.openagua.org/

Rheinheimer, D.E., Medellin-Azuara, J., Ramirez, A., Brown, C., Park, D.K., Torres, E., Sood, A., Knox, S., Garza-Díaz, L., Abdallah, A.M., in review. OpenAgua: A web-based decision support platform for collaborative water systems modeling and analysis. Environmental Modelling and Software.

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.

Rosenberg, D.E., Fillion, Y., Teasley, R.L., Sandoval-Solis, S., Hecht, J.S., Zyl, J.E.v., McMahon, G.F., Jeffery S. Horsburgh, Kasprzyk, J.R., Tarboton, D.G., 2020. The Next Frontier: Making Research More Reproducible. Journal of Water Resources Planning and Management.

Rosenberg, D.E., Watkins, D.W., 2018. New Policy to Specify Availability of Data, Models, and Code. Journal of Water Resources Planning and Management.

Rossman, L.A., 2000. EPANET 2: Users Manual. US Environmental Protection Agency: Cincinnati, Ohio.

Rossman, L.A., 2010. Storm water management model user's manual, version 5.0. National Risk Management Research Laboratory, Office of Research and Development, US Environmental Protection Agency.

Sehlke, G., Jacobson, J., 2005. System Dynamics Modeling of Transboundary Systems: The Bear River Basin Model. Ground Water 43(5) 722-730.

Stagge, J.H., Rosenberg, D.E., Abdallah, A.M., Akbar, H., Attallah, N.A., James, R., 2019. Assessing data availability and research reproducibility in hydrology and water resources. Scientific Data 6 190030.

Tarboton, D.G., Idaszak, R., Horsburgh, J.S., Heard, J., Ames, D., Goodall, J.L., Band, L., Merwade, V., Couch, A., Arrigo, J., 2014. Hydro share: Advancing collaboration through hydrologic data and model sharing, 7th International Congress on Environmental Modelling and Software, iEMSs 2014. International Environmental Modelling and Software Society.

Welsh, W.D., Vaze, J., Dutta, D., Rassam, D., Rahman, J.M., Jolly, I.D., Wallbrink, P., Podger, G.M., Bethune, M., Hardy, M.J., Teng, J., Lerat, J., 2013. An integrated modelling framework for regulated river systems. Environmental Modelling & Software 39(0) 81-102.

Yates, D., Sieber, J., Purkey, D., Huber-Lee, A., 2005. WEAP21—A Demand-, Priority-, and Preference-Driven Water Planning Model. Water International 30(4) 487-500.

Zagona, E.A., Fulp, T.J., Shane, R., Magee, T., Goranflo, H.M., 2001. RiverWare: A Generalized Tool for Complex Reservoir System Modeling. JAWRA Journal of the American Water Resources Association 37(4) 913-929.

Zhang, F., Chen, M., Ames, D.P., Shen, C., Yue, S., Wen, Y., Lü, G., 2019. Design and Development of a Service-oriented Wrapper System for Sharing and Reusing Distributed Geoanalysis Models on the Web. 111(Environ. Model. Softw. 21 7 2006) 498-509.

CHAPTER V

SUMMARY, CONCLUSIONS, AND FUTURE WORK

5.1 Summary and Conclusions

This dissertation contributed a novel framework and software tools to generalize data management for systems modeling to enable systematic data and modeling comparisons across many models and datasets. The contributed framework and software tools address the problem in existing tools used to store, query, prepare data to models, visualize, and publish them online that are model, location, and dataset specific. The need for specific tools arises because data are stored in different formats, described with different vocabularies, and require manual, model-specific, and time-intensive manipulations to identify, organize, compare, and then populate to models. The design of software tools is guided and demonstrated by use cases that represent common tasks done by modelers and water managers. The use cases show a fundamental significance to the science of water management by enabling readily comparisons that generate insight across datasets and models within or across study locations. Ready comparisons are useful to water managers to help them benchmark their water systems and learn from others. The use cases use over a dozen of different water resources datasets and four models in three watersheds, USA and Mexico. This dissertation presented three tools to: (1) identify, organize, analyze, and compare data to use in models, ii) prepare and populate data to many WEAP models, and iii) visualize networks, plot, and compare input and output for different management scenarios and models.

The first chapter presented the Water Management Data Model (WaMDaM) implemented in a relational database. WaMDaM uses contextual metadata, controlled vocabularies, and supporting software tools to organize, store, and compare water management data from multiple sources and models and allow users to more easily interact with its database. Five use cases use thirteen datasets and models focused in the Bear River Watershed, United States to show how a user can identify, compare, and choose from multiple types of data, networks, and scenario elements then serve data to models. The database design is flexible to accommodate new datasets, models, and associated components, attributes, scenarios, and metadata.

The second chapter presented an open-source Python-based software that generalizes and automates the process to prepare and load large input data into the world-wide used Water Evaluation and Planning system (WEAP) model or extract its network and data for many already existing WEAP models and scenarios. The software uses the WEAP Application Programing Interface (API) and the generalized Water Management Data Model (WaMDaM) to store and organize WEAP data and metadata. The software is demonstrated in two use cases using two different existing WEAP models. The first use case queries and compares networks and data of the three WEAP models extracted into a WaMDaM database. The second use case compares water systems reliability to meet demand across four new created scenarios in two models. The scenarios represent changes in reservoir capacity, demand, evaporation, and river headflows and how they affect demand reliability. The presented framework enables modelers to reuse the software tool to quickly setup WEAP models and create comparative scenarios and sensitivity analysis using a single source of data, reduce potential errors in loading data, and allow others to reproduce the set up and results, all without coding.

The third chapter presents an interoperable software ecosystem that integrates WaMDaM with other independently-developed, state-of-the-art, generalized tools to store water resources systems modeling data with metadata and controlled vocabularies, use web-based tools to visualize, compare, edit, publish, discover, and analyze model networks' input and output data. The software tools are Hydra Platform web service, OpenAgua online visualization platform, and HydroShare for data publication. Three use cases show how modelers can systematically reuse software ecosystem tools and web services to visualize and compare three different models in the Bear River Watershed, United States and Monterrey, Mexico, set up scenarios, update input data, and compare model outputs. The ecosystem is a collaborative environment that allows users of existing desktop-based systems models to visualize networks and their data and publish them online. The software ecosystem with its online visualization, editing capabilities, and data publication supports stakeholder engagement and reproducible data analysis.

All the presented software tools offer novel approaches to improve data management, analysis, and comparisons across many datasets and models compared to current approaches that are dataset, model, and location specific. The tools were iteratively revised over the course of five years to satisfy the design requirements, use cases, and feedback. The changes were in response to feedback from collaborators at the University of Manchester, University of California, Davis, University of Massachusetts, Amherst, and Utah State University, Logan. We acknowledge the need for larger and more diverse community testing and feedback to serve a wider audience of users.

5.2 Future Work

This dissertation presented novel software tools that advanced water resources systems modeling cyberinfrastructure to enable systematic data analysis, modeling and comparisons across models, datasets, and study locations. There are several opportunities to further improve these tools and sustain them. Future work includes:

First, to improve access and security, future WaMDaM implementations should build web-server APIs with data query functions that distribute and manage the access to many users at the same time and protect the database integrity from unintended changes. Future software tools to load data to the database and export it to models should be time-efficient, more user-friendly, and compatible with Windows, Mac, and Linux. Future work also should use WaMDaM and web-services to publish, discover, and visualize models and their data and allow multiple users to work with the same datasets.

Second, future work should support geo-spatial search for individual water management infrastructure, its connectivity, and data by extending on the successes for searching time-series data using HydroDesktop (Ames et al., 2012) and HydroClient (http://data.cuahsi.org/) in a desktop or online application. WaMDaM support for controlled vocabularies would be particularly useful to search across different native terminology used in models and users. Example data discovery searches that are not currently supported in the CUAHSI web services are to search for i) a reservoir bathymetry curve or a seasonal demand data at a site, ii) network connectivity such as the links that supply water to demand sites from a particular reservoir.

Future work should extend the WaMDaM and its software ecosystem by building additional tools to import and export data from other datasets and models. We identified three important steps that can help sustain the software tools presented in this dissertation. We already worked on the first two and we aspire to achieve the third in future work. The first step was in using an open-source license and publishing all source code in a GitHub repository under https://github.com/WamdamProject. The second step was in collaborating with colleagues that are actively working on the complementary software tools: Hydra Platform at the University of Manchester, UK, OpenAgua at the University of Massachusetts Amherst, and HydroShare at Utah State University. We expect that coupling WaMDaM software tools with the above projects increases the value of all of them to be useful as a set of interoperable tools. Other researchers can also learn from and follow this software ecosystem approach is couple other software tools. We suggest that the third step requires both human and financial resources that can continue to support these tools and improve them within one or many organizations that believe in the role of hydroinformatics in improving real-world water management. This software ecosystem approach that couples both open-source and proprietary software should learn from and build on the success of both WEAP and RiverWare among others.

APPENDICES

Appendix A: Supplemental tables for the paper: “A Data Model to Manage Data for Water Resources Systems Modeling”

Table A 1: Summary of reviewed water resources data management systems and models

#	Data management system	Name/description
1	Arc Water Utilities Data Model (Grise et al., 2001)	Maintain comprehensive water distribution, sewer, and stormwater records; coordinate and plan capital projects; and improve the operation of utility networks.
2	Arc Hydro (Maidment, 2002)	Delineates watersheds, groundwater and subsurface geo-processing tools, analyzes hydro geometric networks, manage time series data, and configure and export data to numerical models
3	ODM1 (Horsburgh et al., 2008)	A relational model for environmental and water resources data
4	NFCP (Optimal Solutions Ltd, 2009)	Natural Flow Computation Program
5	HEC-DSS (USACE, 2009)	database system designed to efficiently store and retrieve scientific data that is typically sequential
6	Arc Irrigation Data Model (Armstrong, 2010)	Provide a generic template data model to the Irrigation District clients
7	WISKI (Gál, 2010)	Enterprise Data Management application for environmental monitoring data
8	Hydro-Platform (Harou et al., 2010)	Linking water resource network models to an open data management platform
9	RiverML (Jackson, 2014)	Standardizing the Communication of River Model Data
10	Hydra (Knox et al., 2014)	An open-source software platform for water, energy and/or logistics system data management, visualization, model building and model sharing
11	WaDE 0.2 (Larsen and Young, 2014)	Sharing water planning and use data
12	Arc River (Kim et al., 2015)	A GIS-based relational data model for multi-dimensional representation of river hydrodynamics and morphodynamics
13	ODM2 (Horsburgh et al., 2016)	information model and supporting software ecosystem for feature-based earth observations
	Modeling software	**Name/description**
14	MODSIM 8.3.2 (Labadie, 1995)	River Basin Management Decision Support System
15	AQUATOOL (Andreu et al., 1996)	AQUATOOL, a generalized decision-support system for water-resources planning and operational management
16	EPANET 2.00.12 (Rossman, 2000)	Hydraulic and Water Quality Behavior of Water Distribution Piping Systems
17	RiverWare 6.5.2 (Zagona et al., 2001)	A Generalized Tool for Complex Reservoir System Modeling
18	Water-Strategy-Man (Manoli et al., 2001)	Water demand and supply analysis using a spatial decision support system
19	WAS 4.0 (Fisher et al., 2002)	Water Allocation System. The Middle East Water Project
20	CALVIN (Jenkins et al., 2004)	California Value Integrated Network
21	TOPNET (Bandaragoda et al., 2004)	Networked version of TOPMODEL
22	WEAP 2016.01 (Yates et al., 2005)	Water Evaluation And Planning system
23	GSSHA 6.1 (Downer and Ogden, 2006)	Gridded Surface/ Subsurface Hydrologic Analysis
24	ResSim 3.1 (USACE,2007)	Analyze and improve reservoir operations
25	OASIS (HydroLogics, 2009)	Generalized program for modeling the operations of water resources systems
26	SWMM 5.1.007 (Rossman, 2010)	Storm Water Management Model
27	IRAS (Matrosov et al., 2011)	Interactive River-Aquifer Simulation Program
28	HOBBES (Lund et al., 2013)	Bottom up approach to improve and organize the data for water modeling efforts in California
29	ArcSWAT 2012.10.19 (Winchell et al., 2007)	Predict the effect of management decisions on water, sediment, nutrient and pesticide yields with reasonable accuracy on large, ungaged river basins.
30	Source IMS (Welsh et al., 2013)	Source- Integrated Modelling System (IMS)
31	AdHydro (Lai et al., 2013)	Physics-based, high-resolution, distributed water resources model for simulating large watersheds
32	GoldSim 11.1 (GoldSim Technology Group LLC, 2014)	Monte Carlo Simulation Software for Decision and Risk Analysis
33	Basins (US EPA, 2015)	Better Assessment Science Integrating point & Non-point Sources
	OpenAgua (Rheinheimer, 2020)	open source, web-based decision support system for water planning
34	**Data standards and initiatives**	**Name/description**
35	OpenMI (Gregersen et al., 2007)	Open Modelling Interface
36	HY-Features (OGC, 2012)	Common Hydrologic Feature Model
37	DCMI (DCMI, 2013)	Dublin Core Metadata Initiative
38	CSDMS (Peckham et al., 2013)	Community Surface Dynamics Modeling System
39	WRC (Elag and Goodall, 2013)	Water Resources Component
40	Prov-dm (Moreau and Missier, 2013)	World Wide Web Consortium Provenance Working Group
41	HydroShare (Morsy et al., 2017)	HydroShare metadata framework for environmental models

Table A 2**:** Fourteen common required metadata elements for data values in WaMDaM. Other data types like time series, and multi-column attributes have additional specific metadata.

#	**Element**	**Definition**	**Example**
1	Resource type	The name of a collection of object types for a specific model	WEAP
2	Master Network	The name of a collection of scenarios in a specific area with a spatial reference	Lower Bear River Network
3	Connections	The relations of how data values are connected through their instances with others across system components of a water management system	Blacksmith Fork diversion supplies Hyrum reservoir
4	Scenario	The name of a specific configuration of instances, their metadata, and data values that represent management decisions across system components	Base Case Lower Bear River
5	Object Typology	Node, link, network	Node
6	Object Type	A generic type of water system component that can be replicated as instances with specific local data	Reservoir
7	Instance	A system component that represents a node or link instance "where"	Hyrum
8	Organization	The institution where the person who provided or generated the attribute's data value is affiliated with. "who"	Utah Water Research Lab
9	Person (people)	The individual who provided or generated the attribute's data value. "who"	David Rosenberg
10	Source	The origin of the attribute's data value	Lower Bear WEAP Model
11	Method	The procedure used to generate attribute data values. "how"	WEAP Manual
12	Unit	The unit of measurement of attribute data values	Acre
13	Attribute	The qualitative descriptive characteristic of a data value "what"	Surface area
14	Attribute Data Type	One of the seven means to store data value(s): time series, multi-column arrays, numeric or descriptive parameters, seasonal parameters, electronic files	Numeric value
15	Data Value	The numeric or categorical value(s)	480

Table A 3: Supported attribute data types, their definitions, and examples in water resources systems models (Requirement #5)

Data type	Definition	Example and use
Numeric	numeric values	Dam elevation is 450 feet.
Seasonal	parameter values over specified time periods	Water right parameter can have 20 acre-feet in winter and 5 acre-feet in summer or a water demand can take 10 cfs at day and 5 cfs at night. Modelers may optionally register the season name with a controlled term. For each record of season name and value, there is a season order field to preserve the seasons and values order as they are entered which can also be used to sort the season values.
Categorical	Categorical values	Reservoir purpose of "irrigation," "hydropower generation," or "flood control". Or True or false values that indicate dual system operational status e.g., "open", "closed"
Free text	any text values	Dam release rule stored as block of text, a script, or a description of a system
Time series	numerical values for specified times/dates	Stream discharge, evaporation, inflow, demand, supply
Multi-attribute series	paired numeric values for two or more attributes (i.e., columns)	Reservoir volume and surface area that change with elevation. Water cost that changes with demand month of the year.
Electronic file	physical file to attach to the database	Images, PDF documents, NetCDF and shape-files. They are stored as Binary Large OBject (BLOB) in the database.

Table A 4: Unique and shared network nodes and links, metadata (source and method) and data between two the Normal and Dry scenarios in the BRSDM Model in the Upper Bear River Watershed

Scenario comparison element	**Unique to "Bear Normal Year Model" scenario** Count of instances (%)	**Shared** Count of instances (%)	**Unique to "Bear Dry Year Model" scenario** Count of instances (%)
Network nodes and links	0	79 (100%)	0
Network metadata	0	240 (100%)	0
Attributes metadata	0	584 (100)	0
Data	21 (3.6%)	543 (93.0%)	20 (3.4%)

Appendix B: Supplemental tables for the chapter: "Open Source Python Software to Manage, Populate, and Compare WEAP Models and Scenarios"

Description of the five steps mention in section 3.2.1: Extract WEAP Areas into WaMDaM

i. Connect to WEAP API

To extract a WEAP model into WaMDaM, in the WaMDaM Wizard window, the user chooses i) a "WEAP Area" name among the many models on their machine and ii) scenario within the Area. If the WEAP Area is georeferenced, the user iii) has the option to keep the coordinate system projection as-is or provide the current WEAP map projection using the European Petroleum Survey Group (EPSG) identifier. The EPSG identifier is used to project the local system coordinates into the World Geodetic System 1984 (WGS84). Having the coordinates in the WGS84 would be useful for visualization purposes based on Google maps as in OpenAgua (Abdallah, 2019). Next the user iv) provides the directory on the user's machine to save the extracted data as a WaMDaM workbook template. When the user clicks "Extract", a Python script uses the WEAP API to activate the selected WEAP Area and Scenario and gets the Area's specific directory on the user's machine, which will be used to read and write the CSV or text files for time series data. Then the script will execute the next functions in order to extract the WEAP data structure (attributes and their units), the network (nodes and links), and data values, prepare the output, and save it to WaMDaM workbook.

ii. Get WEAP Data Structure

The WEAP data structure contains its object types, their attributes, units and data types. Although WaMDaM is generic to organize WEAP structures, WEAP allows users to change the unit and data type of an attribute across object instances. For example, “Monthly demand” for one demand site may be a constant value measured in acre-feet while another site may have “Monthly demand” specified by a time series in cubic feet per second. WEAP internally interpolates, extrapolates, and converts units to a common time step and unit and this feature gives modelers flexibility. However, WaMDaM requires an object attribute to have the same data type and units of measurement across all object instances (Abdallah and Rosenberg, 2019) This requirement allows WaMDaM to work for many models including WEAP.

There are many potential ways to resolve the mismatched requirement WEAP and WaMDaM have for attribute units and data types. We choose to create new attributes in WaMDaM that represent a unique combination of the data type and units (Table 3.2). For example, a monthly demand attribute with numeric data type and units of cubic feet per second took the new attribute name “Monthly Demand_Nu_C”. Similarly, a monthly demand attribute with seasonal data type and units of acre-feet per seasonal period took the new attribute name “Monthly Demand_Se_A.” Each attribute name always references its data type abbreviation (i.e., Nu) while the unit abbreviation is only added if the attribute has multiple different units. In all cases, the “AttributeName_Abstract” field stores the root name of the WEAP variable. This field allows WaMDaM to relate all the derivative attributes as well as register them once with controlled vocabulary (“Monthly Demand” in Table B-1). This approach to resolve the mismatched requirements for data type and units represents a tradeoff between consistency and flexibility and future work may improve the method.

Table B 1: An example of how WaMDaM handles WEAP attributes that have different data types and units across object instances

Instanced Name	AttributeName_Abstract	Data Type	Unit	AttributeName
Providence Irr	Monthly Demand	Numeric	Cubic Feet	Monthly Demand_Nu_C
Highline Canal	Monthly Demand	Numeric	AF	Monthly Demand_Nu_A
Cub River Irr	Monthly Demand	Seasonal	AF	Monthly Demand_Se
Mendon Canal	Monthly Demand	Time series	AF	Monthly Demand_Ts
Wasatch Front	Monthly Demand	FreeText	AF	Monthly Demand_Fr

iii. Extract WEAP network

This Python function uses the WEAP API to iterate over the selected WEAP Area's branches (WEAP.Branches) to get the nodes and links with their start and end nodes, their object types, and coordinates. Extracting the WEAP network and matching it with WaMDaM structure was the most challenging task due to a few unique ways that WEAP provides access to its nodes and links.

Here we explain the logic we developed in accessing nodes and links through the different WEAP API data calls. The WEAP API offers "Branch.IsNode" to get the object types (Branch.TypeName) and their instances (Branch.Name) with a node typology. The coordinates are available as (Branch.x) and (Branch.y). WEAP has two types of nodes, the first is nodes that take input data and the second is nodes for connectivity purposes which also are points with a calculated output. The main nodes with input data are accessible directly in the WEAP API: Catchment, Demand site, Reservoir, Flow Requirement, Stream

Gauge, Groundwater, Run of River Hydro, and Wastewater Treatment Plant. Other topologic connection WEAP nodes are accessed as part of the links as described next.

We used two ways to access links and their start and end nodes ("Branch.IsLine") based on how their start and end nodes are available in the WEAP API. The first one is straightforward for the links: Transmission Link, Return Flow, Runoff Infiltration, and Diversion. The links connectivity of start and end nodes are accessible in the WEAP API indirectly through the start and end to links "Branch.NodeAbove.TypeName" and "Branch.NodBelow.TypeName".

The second way to access nodes is for the River and River Reach links, which is complex. WEAP supports Rivers that are automatically segmented to River Reaches based on any connectivity nodes that are placed on the river such as diversion outflow, tributary confluence, streamflow gauge, and return flow. The River object type behaves as a start node at the upstream River segment with headflows and water quality input.

WEAP does not explicitly define a start and end nodes for object type "River" but WaMDaM requires them. Thus, we programmatically created a river start node that has the same River Name and a suffix Headflow (i.e., RiverName + 'Headflow'). The river start node longitude and latitude coordinate is obtained from the WEAP API call "Branch.NodeAbove.x" or "y". We also give this node an Object Type called "River Headflow". Similarity for the river mouth, the function programmatically creates a river end node that has the same River Name and a suffix Mouth (i.e., RiverName + 'Mouth') with an object type called "River Mouth". The river end node longitude and latitude coordinate are obtained from the WEAP API call "Branch.NodeBelow.x" or "y". Finally, we use the "Pyproj" Python library (Whitaker et al, 2019) to programmatically transform the local geographic latitude and longitude coordinates into the WGS84 system which has the code of “EPSG:4326”.

iv. Get WEAP Parameter Values

This Python function iterates over all the object types, their attributes, and extracted nodes and links to get their values. Each variable has a property called “expression,” which encodes the value(s) under the API property “Variable.Expression.” WEAP internally interprets each type of values and incorporates them into its calculations and plots. Alternatively, we programmatically in Python used the following patterns to interpret and map WEAP values into WaMDaM attribute data types (Table-B2)

Table B 2: Encoded data type interpretation approach from WEAP into WaMDaM

Data Type	**Interpretation conditions from WEAP into WaMDaM**
Time series	If value starts with "ReadFromFile". WEAP stores and reads times series data (daily, monthly, or annual) into external files (.csv or .txt). Each time-series file is referenced with its local path (location and name on a desktop) and structured such as the first column is for the year, the second is for the month, the third column is for the value. We excluded any metadata rows at the top of the file that start with "#" or "$"
Seasonal	If value starts with "MonthlyValues", which is an average monthly estimate across the modeling time step (e.g., years). WEAP default setting stores the seasonal data as a comma separated string where the first value is for the three letters of the month name and the second part is the numeric seasonal value
Multi-column	If value starts with "VolumeElevation", which is used in WEAP for the reservoir volume elevation curve. Similar to the "MonthlyValues", WEAP stores the "VolumeElevation" data as a comma separated string where the first part is for the volume value and the second part is elevation value
Numeric	If value is float or integer
FreeText	If none of the above criteria is met, which is a text value that largely includes the "functions" in WEAP where the modeler could create a function that calculates an attribute value based on other attributes values in other nodes or links.

WEAP uses a special branch name that contains "Key Assumptions," which are equivalent to "Network Attributes" in WaMDaM. Those attributes can be reused to apply to some or all the WEAP nodes and links. The script parses each branch full name that contains "Key" and defines it as a Network Attribute in WaMDaM. The expression values for each assumption are interpreted similar to the regular variables described above.

Finally, the GetWEAPValues Python function parses the values of each data type and manipulates them to be ready as input for the WaMDaM workbook. For example, the code opens each referenced time series file and combines the year and month columns into a date. It also associates each expression (data value) with its object type, attribute name, node or link (Table B 3).

v. Save Extracted Data to Excel

After extracting the data structure, network nodes and links, and values, each of them is organized in a Python Pandas dataframe to match the required column names and orders in each spreadsheet in the WaMDaM workbook. The function that extracts WEAP to WaMDaM creates a workbook and writes the output to the following sheets: Attributes, ScenarioNetwork, Nodes, Links, Numeric, FreeText, Seasonal, TimeSeries, TimeSeriesValues, and MultiAttributeSeries. Then, the user is required to enter the source and method names used to generate the input data in the extracted WEAP model data. Users can also register the WEAP model node and link instances with controlled vocabularies to allow linking them with other synonymous terms and enable their search. Finally, users load the workbook content into a WaMDaM SQLite database where they can query, compare, and plot data across models and scenarios outside WEAP's proprietary database.

Table B 3: Example data values for each types generated by the WaMDaM wizard as input ready for WEAP models

WEAP pattern to populate values

BranchTypeName/BranchName/Variable=Value

Seasonal

Reservoir/Willard Res/Net Evaporation=MonthlyValues(Oct, 42150, Nov, 3406, Dec, 0, Jan, 0, Feb, 4258, Mar, 60884, Apr, 59181, May, 61309, Jun, 46834, Jul, 50240, Aug, 43002, Sep, 54497)

Time series

Demand Site/Lost Creek/Monthly Demand=ReadFromFile(C:\Users\Adel\Documents\WEAP Areas\Bear_River_WEAP_Model_2017_scenarios\Headflow_ScenarioData\TimeSeries_csv_files\Monthly_Demand_Lost_Creek.csv)

Multi column array

VolumeElevation(0.0, 5450.0, 0.153, 5460.0, 0.894, 5470.0, 3.06, 5480.0, 6.73, 5490.0, 11.83, 5500.0, 18.48 ,5510.s0, 26.62, 5520.0, 36.1, 5530.0 ,47.2, 5540.0 ,59.88 ,5550.0 , 73.94, 5560.0)

Numeric

River Reach/Below Tributary to Weber 3 Headflow/ River Flooding Fraction=100

TextFree

Reservoir/Great Salt Lake/Top of Buffer=Top of Inactive[Thousand AF]

Appendix C: Supplemental tables for the chapter: "A Software Ecosystem to Store, Visualize, and Publish Modelling Data for Water Resources Systems"

Table C 1: Example systems water management models architecture (Generalized models)

#	Software/Model	Purpose	[# of users/members worldwide by Jan 2018]	Example study	A location applied	Data management	Model algorithms (engine) / configuration	User Interface
	Proprietary and fee-based							
1	RiverWare (Zagona et al., 2001)	Generalized river and reservoir object-oriented modeling tool, providing a construction kit for developing and running detailed site-specific models for planning and operating river systems	113 Organizations as of Sep 2015	Development and Implementation of an Optimization Model for Hydropower and Total Dissolved Gas in the Mid-Columbia River System (Witt et al., 2017)	Tennessee Authority, Colorado, Columbia River System, USA	Proprietary bundled + csv files		
3	WEAP (Yates et al., 2005)	WEAP ("Water Evaluation And Planning" system) is a user-friendly software tool that takes an integrated approach to water resources planning.	27,345	Integrating water supply constraints into irrigated agricultural simulations of California (Winter et al., 2017)	California, USA	Proprietary bundled (paradox database + csv files)		
3	eWater Source (Welsh et al., 2013) ecological influences	Australia's National Hydrological Modelling Platform (NHMP) – is designed to simulate all aspects of water resource systems to support integrated planning, operations and governance from urban, catchment to river basin scales including human and	3,517	An integrated modelling framework for building a daily river system model for the Murray–Darling Basin, Australia (Yang et al., 2017)	Australia	Proprietary bundled		
4	GoldSim (GoldSim Technology Group LLC, 2014)	GoldSim is the premier Monte Carlo simulation software solution for dynamically modeling complex systems in engineering, science and business. GoldSim supports decision-making and risk analysis by simulating future performance while quantitatively representing the uncertainty and risks inherent in all complex systems.	100s of commercial users and 200 academic institutions	Vulnerability Assessment to Support Integrated Water Resources Management of Metropolitan Water Supply Systems (Goharian et al., 2017)	Utah, USA	Proprietary bundled + csv files		
5	HEC-ResSim	model reservoir operations at one or more reservoirs for a variety of operational goals and constraints. The software simulates reservoir operations for flood management, low flow augmentation and water supply for planning studies, detailed reservoir regulation plan investigations, and real-time decision support	(NA) U.S. Army Corps of Engineers and around the world	Simulating system-wide effects of reducing irrigation withdrawals in a disputed river basin	Alabama, Georgia, Florida	Proprietary bundled		
	Open Source and free							
6	SWMM (Rossman, 2010)	EPA's Storm Water Management Model (SWMM) is used throughout the world for planning, analysis, and design related to stormwater runoff, combined and sanitary sewers, and other drainage systems in urban areas.	(#citation to the manual Google Scholar 1,175)	Layout effects and optimization of runoff storage and filtration facilities based on SWMM simulation in a demonstration area (Xing et al., 2016)	China	ASCII text files	Open source bundled and separate packages as well	
7	EPANET (Rossman, 2000)	EPANET is software that models drinking water distribution piping systems.	#citation to the manual Google Scholar 2,502	Management of a water distribution network by coupling GIS and hydraulic modeling: a case study of Chetouane in Algeria (Abdelbaki et al., 2017)	Algeria	ASCII text files	Open source bundled and separate packages as well	
8	REALM (Perera et al., 2005)	The REsource ALlocation Model is a computer program that can simulate the operation of water supply systems during droughts as well as during periods of normal and high streamflows.	123 Google Scholar citations	An integrated hydro-economic modelling framework to evaluate water allocation strategies I: Model development (George et al., 2011)	India, Australia	ASCII text files	Open source bundled	
9	HydroPlatform (Harou et al., 2010)	Generic open-source software interface and web repository for water management models	Google Scholar 14 citations	A computationally efficient open-source water resource system simulator – Application to London and the Thames Basin (Matrosov et al., 2011)	London, UK USA, Utah	SQLite generic database	Importers and exporters	Python 2.7 open source

Table C 2: Example systems water management models architecture (site specific models)

#	Software/Model	Purpose	Availability	Places applied	Data management	Model algorithms (engine) / configuration	User Interface
1	Bear River Systems Dynamics Model	System Dynamics Modeling of Transboundary Systems: The Bear River Basin Model (Sehlke and Jacobson, 2005)	author request	Wyoming, Idaho, USA	Excel Workbooks	VBA	Not available
2	Statewide Economic-Engineering Water Model (CALVEN)	CALVIN is a hydro-economic optimization model of California's intertied water system. It is the only model representing the extensive statewide system in terms of supplies, demands, and physical and economic adaptability.	Free online	California, USA	https://github.com/ucd-cws/calvin-network-data text files	HEC-PRM	Third party (HOBBES) Node.js® is a JavaScript **View** https://hobbes.ucdavis.edu/cwn **source code** https://github.com/ucd-cws/calvin-network-app
3	Systems model in Wetlands to Allocate water and Manage Plant Spread (SWAMPS)	Systems modeling to improve the hydro-ecological performance of diked wetlands (Alminagorta et al., 2016)	Free (GiHub) https://github.com/alminagorta/Systems-model-in-Wetlands-to-Allocate-water-and-Manage-Plant-Spread	Utah, USA	GDX, Excel	GAMS	Third party (Matlab)
4	Watershed Area of Suitable Habitat model (WASH) (Alafifi and Rosenberg, 2020)		https://github.com/ayman510/WASH		GDX, Excel	GAMS	Source code based https://github.com/ayman510/WASH/tree/master/WebMap View ArcGIS http://washmap.usu.edu/
5	Interactive River-Aquifer Simulation (IRAS)	Interactive River-Aquifer Simulation A computationally efficient open-source water resource system simulator – Application to London and the Thames Basin (Matrosov et al., 2011)	https://sourceforge.net/projects/iras/	London and the Thames Basin	GDX, Excel	GAMS	HydroPlatform

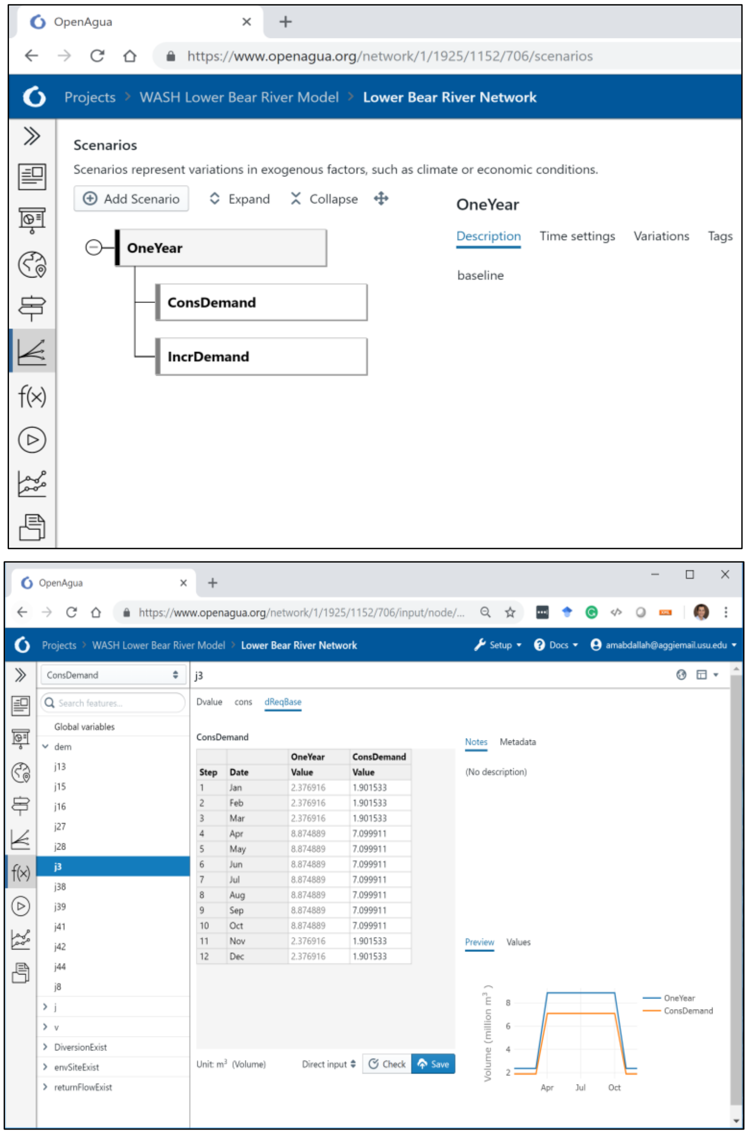


Figure C1: Open Agua interface (a-top) to add scenarios, (b-bottom) edit demand requirement input data for node “j3” which is Cache County urban demand for the Conservation Scenario

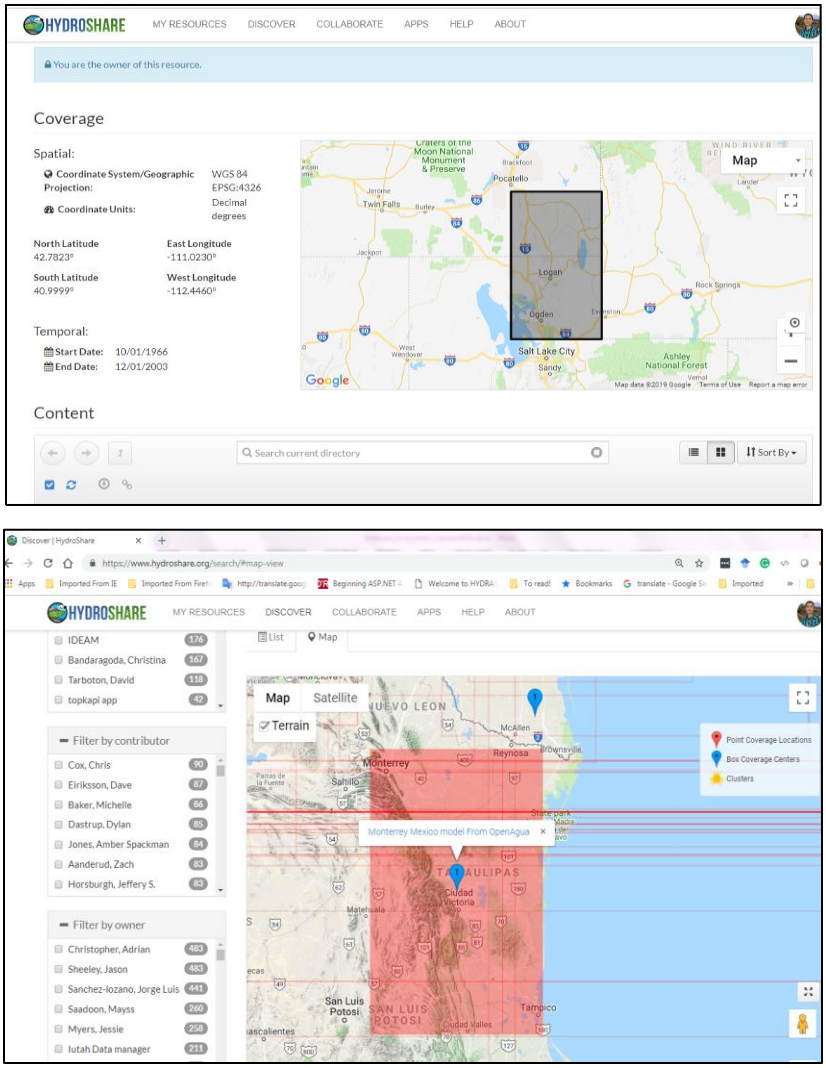


Figure C2: HydroShare **(**a-top): spatial coverage of the Bear River Models, USA (b-bottom) search box for Monterrey, Mexico model

References

Andreu, J., Capilla, J., Sanchís, E., 1996. AQUATOOL, a generalized decision-support system for water-resources planning and operational management. Journal of Hydrology 177(3) 269-291.

Armstrong, L., 2010. ArcGIS Irrigation Data Model. ESRI Water/Wastewater and Water Resources Team, Redlands, Ca.

Bandaragoda, C., Tarboton, D.G., Woods, R., 2004. Application of TOPNET in the distributed model intercomparison project. Journal of Hydrology 298(1–4) 178-201.

Downer, C.W., Ogden, F.L., 2006. Gridded Surface Subsurface Hydrologic Analysis (GSSHA) User's Manual; Version 1.43 for Watershed Modeling System 6.1. DTIC Document.

Dublin Core Metadata Initiative (DCMI), 2013. Dublin Core Metadata Element Set, Version 1.1 ANSI/NISO Z39.85-2012

National Information Standards Organization (NISO): Baltimore, MD.

Elag, M., Goodall, J.L., 2013. An ontology for component-based models of water resource systems. Water Resources Research 49(8) 5077-5091.

Fisher, F.M., Arlosoroff, S., Eckstein, Z., Haddadin, M., Hamati, S.G., Huber-Lee, A., Jarrar, A., Jayyousi, A., Shamir, U., Wesseling, H., 2002. Optimal water management and conflict resolution: The Middle East Water Project. Water Resources Research 38(11) 1243.

Gál, R., 2010. WISKI–A World Wide Used Environmental Monitoring System In: Pavel Bella, P.G. (Ed.), International Show Caves Association 6th Congress: Slovakia, Liptovský Mikuláš.

GoldSim Technology Group LLC, 2014. GoldSim User's Guide Version11.1.

Gregersen, J.B., Gijsbers, P.J.A., Westen, S.J.P., 2007. OpenMI: Open modelling interface. JOURNAL OF HYDROINFORMATICS 9(3) 175.

Grise, S., Idolyantes, E., Brinton, E., Booth, B., Zeiler, M., 2001. ArcGIS Data Models: Water Utilities. ESRI, Redlands, Ca.

Harou, J.J., Pinte, D., Tilmant, A., Rosenberg, D.E., Rheinheimer, D.E., Hansen, K., Reed, P.M., Reynaud, A., Medellin-Azuara, J., Pulido-Velazquez, M., Matrosov, E., Padula, S., Zhu, T., 2010. An open-source model platform for water management that links models to a generic user-interface and data-manager In: David A. Swayne, W.Y., A. A. Voinov, A. Rizzoli, T. Filatova (Ed.), International Congress on Environmental Modelling and Software, Modelling for Environment's Sake ed. International Environmental Modelling and Software Society (iEMSs) Ottawa, Ontario, Canada.

Horsburgh, J.S., Aufdenkampe, A.K., Mayorga, E., Lehnert, K.A., Hsu, L., Song, L., Jones, A.S., Damiano, S.G., Tarboton, D.G., Valentine, D., Zaslavsky, I., Whitenack, T., 2016. Observations Data Model 2: A community information model for spatially discrete Earth observations. Environmental Modelling & Software 79 55-74.

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.

Hydrologic Engineering Center (HEC) of the US Army Corps of Engineers, 2007. Reservoir System Simulation (HEC-ResSim), 3.0a ed.

HydroLogics, 2009. User Manual for OASIS WITH OCL.

Jackson, S.R., 2014. RiverML: A Harmonized Transfer Language for River Hydraulic Models, Civil, Architectural, and Environmental Engineering. University of Texas at Austin.

Jenkins, M., Lund, J., Howitt, R., Draper, A., Msangi, S., Tanaka, S., Ritzema, R., Marques, G., 2004. Optimization of California's Water Supply System: Results and Insights. Journal of Water Resources Planning and Management 130(4) 271-280.

Kim, D., Muste, M., Merwade, V., 2015. A GIS-based relational data model for multi-dimensional representation of river hydrodynamics and morphodynamics. Environmental Modelling & Software 65 79-93.

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.

Labadie, J., 1995. River basin network model for water rights planning, MODSIM: Technical manual. Department of Civil Engineering, Colorado State University, Fort Collins, CO.

Lai, W., Steinke, R.C., Ogden, F.L., Douglas, C., Miller, S.N., Zhang, Y., 2013. ADHydro: A Large-scale High-resolution Multi-physics Distributed Water Resources Model for Water Resources Simulations in a Parallel Computing Environment, AGU Fall Meetings: San Francisco, USA.

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.

Lund, J., Josué Medellín, Samuel Sandoval, Wei Chu, Alvar Escriva, Ashley Vincent, Erik Porse, Prudentia, Zhinkalala, Timothy Nelson, Hui, R., 2013. Building Models from the Data Up: From Calvin to Hobbes.

Maidment, D.R., 2002. Arc hydro : GIS for water resources. ESRI Press, Redlands, Calif.

Manoli, E., Arampatzis, G., Pissias, E., Xenos, D., Assimacopoulos, D., 2001. Water demand and supply analysis using a spatial decision support system. Global NEST: the international Journal 3(3) 199-209.

Matrosov, E.S., Harou, J.J., Loucks, D.P., 2011. A computationally efficient open-source water resource system simulator – Application to London and the Thames Basin. Environmental Modelling & Software 26(12) 1599-1610.

Moreau, L., Missier, P., 2013. Prov-dm: The prov data model.

Morsy, M.M., Goodall, J.L., Castronova, A.M., Dash, P., Merwade, V., Sadler, J.M., Rajib, M.A., Horsburgh, J.S., Tarboton, D.G., 2017. Design of a metadata framework for environmental models with an example hydrologic application in HydroShare. Environmental Modelling & Software 93(Supplement C) 13-28.

Open Geospatial Consortium (OGC), 2012. HY_Features: a Common Hydrologic Feature Model Discussion Paper, In: Rob Atkinson, Dornblut, I. (Eds.), OGC 07-041r1. Open Geospatial Consortium Inc.

Optimal Solutions Ltd, 2009. Natural Flow Computation Program User's Manual: Calgary, Alberta.

Peckham, S.D., Hutton, E.W.H., Norris, B., 2013. A component-based approach to integrated modeling in the geosciences: The design of CSDMS. Computers & Geosciences 53(0) 3-12.

Rheinheimer, D., 2020. OpenAgua: Collaborative water system modeling for a new generation. https://openagua.wordpress.com/

Rossman, L.A., 2000. EPANET 2: Users Manual. US Environmental Protection Agency: Cincinnati, Ohio.

Rossman, L.A., 2010. Storm water management model user's manual, version 5.0. National Risk Management Research Laboratory, Office of Research and Development, US Environmental Protection Agency.

US Army Corps of Engineers Hydrologic Information Center (HEC), 2009. HEC Data Storage System, 2.0 ed. US Army Corps of Engineers Institute for Water Resources Hydrologic Engineering Center (HEC) Davis, CA.

US EPA, 2015. BASINS 4.1 (Better Assessment Science Integrating point & Non-point Sources) Modeling Framework. National Exposure Research Laboratory: RTP, North Carolina.

Welsh, W.D., Vaze, J., Dutta, D., Rassam, D., Rahman, J.M., Jolly, I.D., Wallbrink, P., Podger, G.M., Bethune, M., Hardy, M.J., Teng, J., Lerat, J., 2013. An integrated modelling framework for regulated river systems. Environmental Modelling & Software 39(0) 81-102.

Winchell, M., Srinivasan, R., Di Luzio, M., Arnold, J., 2007. Arc-SWAT interface for SWAT2005-User's guide. USDA Agricultural Research Service and Texas A&M Blackland Research Center, Temple, Texas.

Yates, D., Sieber, J., Purkey, D., Huber-Lee, A., 2005. WEAP21—A Demand-, Priority-, and Preference-Driven Water Planning Model. Water International 30(4) 487-500.

Zagona, E.A., Fulp, T.J., Shane, R., Magee, T., Goranflo, H.M., 2001. RiverWare: A Generalized Tool for Complex Reservoir System Modeling. JAWRA Journal of the American Water Resources Association 37(4) 913-929.

Curriculum Vitae

Adel M. Abdallah
Western States Water Council
682 East Vine Street, Suite 7
Murray, UT 84107
Emails: amabdallah@aggiemail.usu.edu; adelabdallah@wswc.utah.gov

HIGHLIGHTS

- Co-authored five published peer-reviewed journal articles and preparing two manuscripts
- Manages the Water Data Exchange (WaDE) program at the Western States Water Council and lead the design of the WaDE 2.0 Data Model and architecture
- Designed the Water Management Data Model (WaMDaM): A database standard and software tools to manage water resources data for modeling http://wamdam.org/
- Interned at the Western States Water Council and the USAID Water for Food Program
- Recipient of the Best Research-Oriented Paper of the Year. “Heterogeneous Residential Water and Energy Linkages and Implications for Conservation and Management.” Environmental & Water Resources Institute (EWRI) of the American Society of Civil Engineers (ASCE), 2015
- Recipient of the Blue Goes Green Grand: “Measuring Water and Energy Conservation of High-Efficient Automatic Faucets.” $3,710. Utah State University Student Sustainability Office. 2013-2015
- Served as the College of Engineering Graduate Students Senator, Graduate Student Senate, Utah State University. 2014-2015
- Peer-reviewed 13 articles for four top tier water resources journals

EDUCATION

PhD Water Resources Engineering and Hydrology, Utah State University, Logan, UT **May 2020**

Dissertation “Advancing Water Resources Systems Modeling Cyberinfrastructure to Enable Systematic Data Analysis, Modeling, and Comparisons”

M.Sc Civil and Environmental Engineering, Utah State University, Logan, UT
July 2012

Thesis: "Heterogeneous Water and Energy End-Uses and Implications for Residential Water and Energy Conservation and Management"

<u>B.Sc</u> Civil Engineering, An-Najah National University, Nablus, Palestine **May 2008**

Capstone Project: "Modeling Fate and Transport of Chlorine in Drinking Water Distribution Network, Nablus City, Palestine"

PUBLICATIONS

Peer-Reviewed Publications

1. **Adel M. Abdallah** and David Rosenberg, (2019). "A data Model to Manage Data for Water Resources Systems Modeling." Environmental Modeling and Software.

2. James H. Stagge, David E. Rosenberg, **Adel M. Abdallah**, Hadia Akbar, Nour A. Attallah, Ryan James (2019). Assessing data availability and research reproducibility in hydrology and water resources. Nature Scientific Data

3. Andrea Cominola, Matteo Giuliania, Andrea Castellettia, David Rosenberg, **Adel M. Abdallah** (2018). "Implications of Data Sampling Resolution on Water use Simulation, End-Use Disaggregation, and Demand Management." Environmental Modeling and Software.

4. Jeffery S. Horsburgh, Miguel E. Leonardo, **Adel M. Abdallah**, David E. Rosenberg (2017). "Measuring Water Use, Conservation, and Differences by Gender Using an Inexpensive, High-Frequency Metering System." Environmental Modeling and Software.

5. **Adel M**. **Abdallah** and David E. Rosenberg (2014). "Heterogeneous Residential Water and Energy Linkages and Implications for Conservation and Management." ASCE-Journal of Water Resources Planning and Management.

Peer-Reviewed Conference papers

1. Andrea Cominola, Matteo Giuliani, Andrea Castelletti, **Adel M. Abdallah**, and David E. Rosenberg (2016). "Developing a Stochastic Simulation Model for the Generation of Residential Water End-Use Demand Time Series." Proc., 8th International Congress on Environmental Modelling and Software, International Environmental Modelling and Software Society (iEMSs), Toulouse, France.

2. **Adel M. Abdallah** and David E. Rosenberg (2014). "WaM-DaM: A Data Model to Organize and Synthesize Water Management Data." Proc., 7th International Congress on Environmental Modelling and Software, International Environmental Modelling and Software Society (iEMSs), San Diego, California, USA.

Conference Proceeding

- Marwan Haddad, Numan Mizyed, **Adel M. Abdallah** (2009) “Performance of Hydroponic System as Decentralized Wastewater Treatment and Reuse for Rural Communities.” 2nd International Conference on the Palestinian Environment, An-Najah National University, October 2009.

Manuscripts in review or preparation

1. Rheinheimer, D.E., Medellin-Azuara, J., Ramirez, A., Brown, C., Park, D.K., Torres, E., Sood, A., Knox, S., Garza-Díaz, L., **Abdallah, A.M**., in review. OpenAgua: A web-based decision support platform for collaborative water systems modeling and analysis. Environmental Modelling and Software.

2. **Adel M. Abdallah**, David Rheinheimer, David Rosenberg, Steve Knox, Julien Harou” An Interoperable Software Ecosystem to Store, Visualize, and Publish Water Resources Systems Modelling Data”

3. **Adel M. Abdallah** and David Rosenberg. “Open Source Python Framework to manage many WEAP models and scenarios”

RESEARCH, TRAINING, and WORK EXPERIENCE

1. July 2019 - present. **Water Data Exchange Program Manager**. Western States Water Council, Salt Lake City, Utah.
2. Jan-June 2019. **Senior Hydroinformatics Specialist**. Western States Water Council, Salt Lake City, Utah.
3. June 2017– Dec 2018. **Intern,** Western States Water Council, Salt Lake City, Utah.

4. January 2010 – Dec 2018. **Graduate Research Assistant**, Utah Water Research Laboratory, Logan, UT.

5. Fall 2017. **Teaching Assistant**, Engineering Economics, Utah State University

6. July 2014 – Sept 2018. **Visiting Graduate Student**, University of Utah, Salt Lake City, UT

7. 2014 – 2016. **E-intern,** the United States Agency for International Development (USAID), Program: Securing Water for Food Grand Challenge (three months each year).

8. June 2009 – December 2009. **Staff Engineer**, Universal Group for Engineering and Consulting, Nablus-Palestine.

9. June 2008 – June 2009. **Research Assistant**, An-Najah National University, Nablus-Palestine. Project: Decentralized Wastewater Treatment in Arid Regions in Palestine, a research pilot study.

AWARDS / HONORS

1. 2019 Outstanding Reviewer. Journal of Water Resources Planning and Management of the American Society of Civil Engineers.

2. Amazon Web Services Cloud Credits for Research (2019). Hosting a Moderated Registry of Controlled Vocabularies for Water Management Data ($800)

3. Diversity Graduate Scholarship (2018). The American Public Works Association, Intermountain Section

4. Graduate Student Travel Award (2017). Utah State University Office of Research and Graduate Studies

5. Utah Chapter Graduate Scholarship (2016). American Public Works Association

6. WaterSmart Innovations Scholarship (2015). University Council on Water Resources (UCOWR). Nominated by the Director of the Utah Water Research Laboratory to receive the scholarship to cover the full expenses of attending the conference, September 7, 2015

7. Best Research-Oriented Paper (2015). "Heterogeneous residential water and energy linkages and implications for conservation and management." Environmental & Water Resources Institute (EWRI) of the American Society of Civil Engineers, May 2015

8. Paper Competition Scholarship (2015) J. Paul Riley AWRA Utah Section: Student Water Conference & Paper Competition

9. President's Award and Scholarship (2015). Utah State University Student Association, April 2015

10. Second Place Best Graduate Poster Award (2015). 6th Annual Intermountain Sustainability Summit, Weber State University, Ogden UT, March 2015

11. Graduate Student Travel Award (2014). Utah State University Office of Research and Graduate Studies

12. Utah Water Users Association Scholarship (2013)

13. Graduate Enhancement Award (2013). Graduate Student Senate at Utah State University

14. Graduate Student Travel Award (2013). World Environmental and Water Resources Congress, Cincinnati, OH. May 2013

15. PhD Research Assistantship Scholarship (2012-2018). Utah Water Research Laboratory

16. Paper Competition Scholarship (2012). J. Paul Riley AWRA Utah Section: Student Water Conference & Paper Competition

17. Eva Nieminski Honorary Graduate Category Scholarship (2011). The Intermountain Section of the American Water Works Association (AWWA)

18. Utah Water Conservation Forum Scholarship (2011)

19. Great Basin Chapter of Air and Waste Management Association Scholarship (2011)

20. Ivanhoe Fellow, Ivanhoe Foundation Fellowship (2010 and 2011)

21. MSc. Research Assistantship Scholarship (2010-2012). Utah Water Research Laboratory,

FUNDED PROJECTS

- **Adel M. Abdallah**. "Measuring Water and Energy Conservation of High-Efficient Automatic Faucets." (2013) Utah State University. Utah State University Student Sustainability Office. Blue Goes Green Grant. **$3,710**. March 2013-October 2015

TEACHING ACTIVITIES

- Teaching Assistant. Engineering Economics. Undergraduate class Fall 2017 at Utah State University, Professor: Dr. David Rosenberg

- Code Camp Facilitator (2013 and 2014). Facilitated the implementation of a one-day code camp for high school students at USU. Toured USU's high-performance computing and data storage center, assisted the students in debugging their Python code for a reservoir release functions for Pineview reservoir, Utah that generated hydropower, delivered water for irrigation, and protected the city of Ogden from floods.

- Guest lecture. Water Resources Systems Analysis, Graduate class Fall, 2013 at Utah State University. Professors: Dr. Bruce Bishop and Dr. David Rosenberg. Topic: Monte Carlo Methods - Water Conservation. Nov. 6, 2013

- Guest lecture. Hydroinformatics, Graduate class Fall, 2013 at Utah State University. Professors: Jeff Horsburgh and Dr. Dan Ames, and Dr. Steven Burian. Topic: Water and Energy Conservation though High-Efficiency Automatic Faucets. Nov. 26, 2013

LEADERSHIP and PROFESSIONAL ACTIVITIES

- Peer-reviewer for four journals: ASCE-Journal of Water Resources Planning and Management, Journal of Hydrology and Earth Science Systems (HESS), Journal of Sustainable Cities and Society, Journal of Environmental Modelling and Software. Peer-reviewed nine articles (*copies of reviews are available upon request).* 2012-present.

- College of Engineering Graduate Student Senator. Graduate Student Senate, Utah State University. 2014 – 2015.

- Students Representative Chair and Competition Judge. Spring Runoff Conference, Utah State University, Logan, Utah, 2013 – 2016.

- Member of the Utah State University Interfaith Initiative Committee, 2014 – 2016.

ONLINE OPEN-SCIENCE REPOSITORIES

- Created code and documentation of the Water Management Data Model (WaMDaM) on GitHub https://github.com/WamdamProject

- Wrote code to streamline disparate water flow data files into a central database for the Utah Division of Water Resources https://github.com/amabdallah/UDWR_FlowStorageData

- Contributed to code and documentation of Water Data Exchange Program (WaDE), the Western States Water Council https://github.com/WSWCWaterDataExchange

GRADUATE COURSEWORK

- **PhD:** Hydroinformatics, Microeconomics, Water Law and Policy, Database Implementation, GIS in Water Resources, Advanced Web-based Management Information Systems Development, Research Integrity, and College Teaching Seminar, the Role of Cognition in Engineering Education (audited).

- **M.Sc:** Integrated River Basin/Watershed Planning and Management, Surface Water Quality Modeling, Water Resources Systems Analysis, GIS for Civil Engineers, Groundwater Engineering, Data Analysis and Experimentation in Environmental Science and Engineering, and Physical Hydrology

CERTIFICATIONS

- Certified Public Manager (CPM) training by the National Certified Public Manager® Consortium, 2020.

- Utah Division of Risk Management Defensive Driver Training, May 2013-July 2018
- Utah State University Research Scholars Certificate, 2015
- Institutional Review Board (IRB) Training Certificate. Social & Behavioral Research Investigators and Key Personnel. Collaborative Institutional Training Initiative (CITI), July 2010 - December 2016
- Structured Query Language (SQL) Certificate. ExpertRating.com, April 2013

PROFESSIONAL DEVELOPMENT

- "Getting Started as a Successful Proposal Writer and Academician," One-day intensive Seminar, USU sponsored: Grant Writers' Seminars & Workshops LLC, April 2012 and 2016
- "Software Carpentry Boot Camp." Software Carpentry, two-day intensive workshop, Logan Utah, March 23-24, 2013.
- "Using Python for Weather and Climate Applications" By Johnny Lin, Salt Lake City, UT, Mar 8, 2013
- "Integrated Modeling Workshop" Utah Water Research Laboratory, Logan, UT, August 9, 2012.
- "Great Work Great Career Seminar": Eight weeks seminar by the Stephen R. Covey group which partnered with the Huntsman School of Business, Utah State University, June-July 2011
- "Water Chemistry in Reverse Osmosis and Nanofiltration," Four-day intensive course, Middle East Desalination Research Center, Amman, Jordan, April 2009
- "Public Relations Skills," Continuing Learning Center An-Najah National University, August-2008, Nablus, Palestine

KEY COMPUTER SKILLS

Python, Matlab, ArcGIS, General Algebraic Modeling System (GAMS), Structured Query Language (SQL), WEAP, HEC-ResSim, GitHub, Tableau, Relational Database Modeling

LANGUAGES

- English: Fluent
- Arabic: Native speaker

PROFESSIONAL AFFILIATIONS AND ACTIVITIES

- The International Environmental Modelling & Software Society, member (2014-present)
- American Geophysical Union (AGU), Member (2011-present)
- American Water Works Association (AWWA), Member (2011-present)
- American Society of Civil Engineers (ASCE), Member (2012-present)
- American Water Resources Association (AWRA), Member (2011-present)
- Engineers Association - Jerusalem Center, Member (2008-present)

IN THE MEDIA and OUTREACH

- Co-presented in a short educational movie: "What is a Model?" Educational movie for grades 8 and up. October 2013. This movie was part of the outreach program of the Cyberinfrastructure (CI)-Water project. The video is available on YouTube at: http://www.youtube.com/watch?v=-wWQvBC625E
- Co-presented a short educational movie "Get Involved with Science Activities" to encourage high school students to choose a career in Science Technology Engineering and Math (STEM). The video is available on YouTube at: http://www.youtube.com/watch?v=TGO-w0ovGkE
- Blog contribution, Utah and Western Water Blog. "Field Trip to Park City's Water Treatment Plants, Utah" January 2018. https://utahandwesternwater.wordpress.com/2018/01/12/field-trip-to-park-citys-water-treatment-plants-utah/

CONFERENCE AND PROFESSIONAL MEETING PRESENTATIONS

- Andrea Cominola**, Matteo Giuliania, Andrea Castellettia, David E. Rosenberg, **Adel M. Abdallah** (2018). "Can Data from Intelligent Water Meters Inform Water Demand Modelling and Management Accurately, Feasibly, and Cost-Effectively?." European Geosciences Union General (EGU) Assembly 2018.
- Jeffery S. Horsburgh**, Miguel E. Leonardo, **Adel M. Abdallah**, David E. Rosenberg (2018). "Inexpensive, High-Resolution Data for Quantifying Water Use, Conservation, and Differences by Gender." European Geosciences Union (EGU) General Assembly 2018.

- **Adel M. Abdallah**** and David E. Rosenberg (2017). "A New Method to Organize, Identify, and Compare Water Management Data for Systems Models." World Environmental & Water Resources Congress, Sacramento, CA: May 2125, 2017.

- **Adel M. Abdallah**** and David E. Rosenberg (2016) "Applying Best Data Practices to Work with Water Management Data (WaM-DaM)". The 9th Annual UC Davis Informal Water Management Workshop, University of California, Davis. Modernizing Data Management for System Modeling Discussion. December 13, 2015.

- **Adel M. Abdallah**** and David E. Rosenberg (2015) "Let's Target Collaborative Water and Energy Conservation Actions?" WaterSmart Innovations. Las Vegas, Nevada. September 7, 2015.

- Miguel Leonardo**, **Adel M. Abdallah**,** Jeffery Horsburgh, David E. Rosenberg (2015) "Low-Cost Smart Water Meter for Sustainable Water Monitoring and Conservation." 6th Annual Intermountain Sustainability Summit, Weber State University, Ogden UT, March 2015.

- **Adel M. Abdallah**** and David E. Rosenberg (2015). "A Relational Model to Organize and Synthesize Disparate Systems Water Management Data." 3rd CUAHSI Conference on HydroInformatics. Model and Data Interoperability: From Theory to Practice July 15-17, 2015, the University of Alabama and the National Water Center, Tuscaloosa, AL.

- **Adel M. Abdallah**** and David E. Rosenberg (2015). "WaM-DaM: A Data Model to Organize and Synthesize Water Management Data." Utah Water Data Users Group 2nd Meeting, Salt Lake City, Utah: Jan. 27, 2015.

- **Adel M. Abdallah**** and David E. Rosenberg (2014). "WaM-DaM: A Data Model to Organize and Synthesize Water Management Data." 8th International Congress on Environmental Modelling and Software (iEMSs)". San Diego, California, USA. June 15-19, 2014.

- **Adel M. Abdallah** and David E. Rosenberg** (2014). "WaM-DaM: A Data Model to Organize, Share, and Publish Water Management Data." World Environmental & Water Resources Congress - EWRI, ASCE, Portland, Oregon. June 1-5, 2014.

- **Adel M. Abdallah** and David E. Rosenberg** (2014). "Targeted and Collaborative Household Water and Energy Conservation Programs to Achieve City-Wide Goals." World Environmental & Water Resources Congress - EWRI, ASCE, Portland, Oregon. June 1-5, 2014.

- **Adel M. Abdallah**** and David E. Rosenberg (2014). "WaM-DaM: A Data Model to Organize and Synthesize Water Management Data." American Water Resources

Association (AWRA) Spring Specialty Conference". Snowbird, Utah, USA. May 12-14, 2014.

- **Adel M. Abdallah**** and David E. Rosenberg (2013). "A Proposed Water Management Data Model (WaM-DaM)." 2013 CUAHSI Conference on Hydroinformatics and Modeling. Logan, Utah. July 19-21, 2013.

- **Adel M. Abdallah**** (2013) "Design a Database to Manage Water Reservoir Data." CI-WATER Symposium, Salt Lake City, UT, May 10, 2013.

- **Adel M. Abdallah**** and David E. Rosenberg (2013). "Identifying Collaborative City-Wide Residential Water and Energy Conservation Programs." World Environmental & Water Resources Congress - EWRI, ASCE, Cincinnati, Ohio. May 19-23, 2013.

- **Adel M. Abdallah**** and David E. Rosenberg (2012). "Water and Energy Conservation Modeling and Planning: Stretching Resources to Save Money." Intermountain Section AWWA 2012 Annual Conference, Logan UT, September 12-14, 2012.

- **Adel M. Abdallah**** and David E. Rosenberg (2012). "Heterogeneous Water and Energy End-Uses and Implications for Water and Energy Conservation and Management." 7th Annual J. Paul Riley AWRA Utah Section, Student Water Conference & Paper Competition, Logan UT, April 10, 2012.

- **Adel M. Abdallah**** and David E. Rosenberg (2012). "Simulating Household-City Water and Energy Conservation Opportunities Modeling Stochastic Energy and Water Consumption to Manage Residential Water Uses." Spring Runoff Conference, Logan UT April 3-4, 2012.

- **Adel M. Abdallah** and David E. Rosenberg** (2011). "Modeling Stochastic Energy and Water Consumption to Manage Residential Water Uses." American Geophysical Union Annual Fall Conference, San Francisco CA, Dec 5-9, 2011.

- **Adel M. Abdallah****and David E. Rosenberg (2011). "Water System Water and Energy Linkages and Implications for Household and City-Scale Systems Modeling." Spring Runoff Conference, Logan UT March 29-30, 2011.

****Indicates the presenter

KEYWORDS

Hydroinformatics; smart meters; water conservation; data management; modeling; water resources; demand management; systems analysis; optimization; simulation; stochastic modeling, uncertainty, water-energy nexus, data modeling