

Complied peer-review feedback (Conference reviewers + Committee Members)
Adel Abdallah, May 17, 2014

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

My responses are in red

Conference Reviewer 1

The paper is well written. It proposes some nice ideas that at times are explained simplistically, but hopefully should prompt some discussion.

Thanks!

Title block at the top of the first page is wrong (old version).

Addressed

Comments:

- Abstract - "design of the WaM-DaM allows" This is the only place you use the article "the" before WaM-DaM and it feels out of place.

Addressed

- et al. This is a Latin term and is usually italicized.

This is the citation style of the *Journal of Environmental Modeling and Software*

- Section 2.2 Controlled Vocabulary - So you either use terms already there or add terms, but names are not "validated", just user validated. There needs to be a way to merge terms that are added and should not have been. This is not discussed.

I added a sentence that points out to the user responsibility to maintain the use of consistent terms throughout

- Section 3 paragraph 2 "the use can faster and easier query them" would read better as "the user can query them faster and easier"

The sentence is deleted

- Section 3 - the discussion between Figure 2 and Figure 3 is fairly simplistic OK.

I added more discussion

- Section 4 - weird page break with the section title on one page and all the text on another.

Addressed

- Section 4 first line of text - "as a standard to organizes, share" the s at the end of organizes needs to be removed.

Addressed

- References - Ames, D.P. - The journal title needs fixed "&"

Addressed

Conference Reviewer 2

The authors describe a Water Management Data Model they refer to as WaM-DaM that they propose as a water resources community standard. A premise is that of existing standards like ODM, Common Hydrologic Feature Model, ArchHydro, WaDE, and others – each has features useful to the community, but it is frequently difficult for water managers to deal with multiple standards and systems across multiple data sources and providers that convey individualized and often dissimilar methods to organize and store data. The authors also point out that certain metadata that would be useful to water managers is rarely documented in these systems. The problem statement that the proposed system is intended to address is nicely described and water managers will readily understand and relate to it and the need for a solution in this space.

The authors proposed solution, WaM-DaM, allows users to represent networks as nodes and links that are user-defined. Networks can be, for example, reservoirs and rivers and associated data including relevant operations. Nodes are representations of physical objects, and links connect two objects and conveys flow between them. The authors present the logical data model for WaM-DaM, and state that their relational database design follows precepts put forth in a textbook on database systems often used in academia. The authors group the model tables of WaM-DaM into five major features described in the paper and include a figure with a well thought-through detailed schematic of their logical data model. The paper discussion combined with the schematic of the logical data model is very effective in giving the water resource community sufficient detail to begin to consider WaM-DaM as a community standard.

The authors next discuss their implementation of WaM-DaM as a relational SQL Server database. As an example in populating WaM-DaM with a real-world example, they describe their initial work populating a single Little Bear River network with data from different sources that use different semantics, metadata, and data formats. Microsoft Excel was used to stage all data prior to individually importing into WaM-DaM. Using their initial implementation of Little Bear River network, the authors then proceed to describe some unique capabilities enabled by WaM-DaM as per their prior paper discussion of the positioning of WaM-DaM in the context of the community problems it is intended to address. The discussion here is a little light in that the example is only a single example of a relatively small network. Many assertions are made without demonstration as to what could be done in addition to what is being discussed and demonstrated with their Little Bear River network initial implementation. That said, the work does appear to be early and so the lack of additional more substantial examples perhaps can be justified given where they are in the award period of their NSF-funded CI-Water under which this work is funded.

Overall, the paper is well written and interesting work. I'm not aware of similar efforts and so the work is original, and the work does address what I feel most readers will readily identify with as a known community problem. I recommend this paper as a strong accept so that the proposed work can be considered by the broader community as a potential solution in this space. As a consequence of publishing this work, I look forward to more substantial vetting with more complex examples prior to its potential adoption as a community standard.

Thanks! In many places, I point out that this work is ongoing and further testing is needed. Also I am no longer using the term standard or community data model.

Notes/Questions on Adel's paper by David Tarboton April 13, 14

Responses to Notes/Questions are in red by Adel Abdallah. May 14, 2014

1. The problem that this paper addresses is unclear. You need to revise the paper to clearly state and formulate the problem that it addresses and the contribution to the solution that this paper makes. I revised the paper and I describe the problem as that the water resources data need to be integrated and synthesized in one system to allow integrated understating of systems data. The important features of water resources data are represented is different existing models and now WaM-DaM integrated all of them. In fact I changed the title of the paper to specifically communicate the contribution of the paper which is an integrated management of all aspects of water resources data.
2. How will you get the community to buy in to the standard? I am no longer refereeing my work as a standard. My work stands by its merits and contribution and the buy in might possibly come from the community if they find my work useful to them.

3. The claims of the contribution are overstated. Specifically with respect to creating a community standard. This is presumptive. What gives you the authority to write a community standard? What will you do to make this a community standard? I agree that the contribution was over stated. I removed the mention of "community standard" throughout
4. The work is still in a very preliminary stage (which makes the importance of clarifying the problem it will address all the more important). I agree. I revised the problem statement and addressed prior work clearly. I also summarized prior work and what feature it support in Table 1.
5. The review refers to features of existing models and states that none of the existing models support all of these features. What are these features specifically and which models support them. A table may be helpful. I agree, I added Table 1 that summarizes these features and mentioned specifically which existing model support of these features.
6. The paper refers to 20 questions. The examples given do not seem precise or focused enough. Can you give all 20 questions and explain how the model will facilitate answering them. (Should be able to as the paper says "we adopted"). I changed the wording "we adopted" to "we followed the concept of." I also added the categories of questions that could be asked and gave more generic and robust example questions. Due to the space limitation I couldn't add more questions. I plan to write a design specification document about WaMDaM and I will include all the 20 questions in this document.
7. In your paper you list CV's value type, data type and object type. Can you please explain the difference? Sure, I explained the difference between these concepts. Value type means like an observation or a model simulation result), object type means either a node or a link, and data type means like Max and average.
8. Can you elaborate overall on the role of controlled vocabularies in this work. What CV lists do you plan to establish as part of this work and what are the terms they will contain. I explained the overall role of the controlled vocabularies as to maintain consistency and enforce homogeneity throughout the data model. Since this work is in its early stage, controlled vocabularies are not fully compiled and I'm just introducing the concept and few examples of them in this work and I mentioned this fact in the controlled vocabulary section.
9. You refer to this work as data intensive. What do you mean by data intensive and how big do you think the databases that you are developing need to get. How are you evaluating their scalability with size?
I'm no longer referring to this term.
10. You mention the time water managers spend compiling data from scattered sources. Do you have any way to measure this and quantify the time that could be saved by using WamDam. How much time will it take to gather the data to put it in to wamdams. Will there be a net time saving?
I'm no longer stating that WaM-DaM saves time.
11. Why is reservoir - a general thing associated with a specific project (on data model slide in presentation).

The concept of a project in WaM-DaM is used as a container for the entire data for a user's work. So first, a user creates a project then they create objects (e.g., reservoir, demand site) and their associated

attributes. The user can choose any project name that makes sense to them. So the project name doesn't not necessarily go with a real project that is active in a particular area like the "Little Bear River Restoration Project" and it can be like "Adel's Project". Then the user creates a network that is located inside the project. The network contains all the instances and their connectivity. It's important to note that multiple networks can share many objects within a project.

Manuscript Title: WaM-DaM: A Data Model to Organize, Share, and Publish Water Management Data

By **Jeff Horsburgh**, April 15m 2014

Responses to Notes/Questions are in red by Adel Abdallah. May 14, 2014

Comments

1. Your title states that the data model is for organizing, sharing, and publishing water management data. I see evidence of the organization claim – do you substantiate the "share" and "publish" claims? How does WaM-DaM help me share and publish this kind of data? What software would I use to publish and share?

I agree! I changed the paper title to reflect the contribution of the work which is an integrated management of all aspects of water resources data.

2. Number 1 is an example of several places in this paper where you overpromise what the contribution of this paper is. You need to be very specific about what the contribution of this paper is and not overclaim. Other examples include:
 - a. Abstract: "We present a community standard..." – what is the community, and what makes this a standard for the community? I suggest calling it a "proposed standard" in all references throughout the document

I abandoned the use of the terms community model and standard.

- b. Abstract: "...open-source data management system..." – what open source code do you have? Where are you managing your open source code repository? What is the "data management system" to which you refer?

I am no longer referring to this data management system.

3. There are several places in the paper where you make a statement and then have multiple citations at the end of the sentence. It is not clear which citation goes with which part of the sentence. In general – move the in text citation as close to the statement with which it belongs as possible.

I fixed this issue

4. There are multiple instances of incorrect citations. I had nothing to do with ArchHydro, WaDE, HydroPlatform, or DSS, but you cite the ODM paper at the end of that sentence. Strassberg et al. did not write ArchHydro, they wrote ArchHydro Groundwater – these are two different data models and books.

Correct, I fixed this issue

5. The last paragraph on the first page should be combined with the first paragraph on the second page. It is repetitive.

I fixed it.

6. On page 2, you state that “...ArcHydro data model is loosely designed to accommodate entities, attributes, and metadata as needed with no standard that imposes a particular data structure.” This is not true. ArcHydro is a very specific data model that is not loosely designed. It is true that many people have extended ArcHydro for many different purposes. It is generally ok to ADD stuff, but if you were to delete elements of the base ArcHydro data model, the data model and any software based on it will break.

Thanks! I fixed this issue

7. On page 2, you state that “...the data model adopts a relational structure to store data values so others can access these values and their meanings through queries.” This get’s back to your claims about sharing – this statement is only true for users that have access to the database, which (due to security or file sharing limitations would be a small number of people). A relational database is only one potential component in a data sharing system (and it’s generally not the part that most users would see).

I am no longer referring to sharing the data. So I talk about accessing data from the perspective of the same user who owns the data database.

8. On page 2, you state that “...WaM-DaM can serve as a new opportunity to expand the capabilities of HydroDesktop software.” First, there’s no substantiation of this statement in the paper – e.g., HOW would WaM-DaM do this? A statement like this is more appropriate for discussion/conclusions/future work rather than the introduction. There is additional infrastructure required to pull this off, but you don’t describe it anywhere.

I’m no longer refereeing to HydroDesktop in this regard.

9. Figure 1: It would take me hours to really digest Figure 1. In a relatively quick look, I notice some potential issues with the data model design. You will only be able to resolve these issues through working with your data use cases. You need to load real data (lots of it) and then retrieve it using your 20 queries actually written as SQL code.

Correct, I already pointed out to this fact and that WaM-DaM is still in its development stage and added “proposed” to the figure caption.

- a. The labels on the relationships are a little hard to follow because some of them are out of place.

I agree. I tried to move the labels closer to the connections

- b. There are many potential issues with the design. Here’s some examples:

- i. Your CV tables are inconsistently used – with some CVs connected to other CVs???? This is problematic.

I resolved this issue and disconnected the connectivity between the CVs tables.

- ii. A good example of the previous issue is the circular relationship you have between ObjectTypeCV, ObjectNameCV, and Objects, where the ObjectType field appears in both the ObjectNameCV and Objects entities. And then, why do you have ObjectName in the AttributeNameCV and Objects entities? Any time you have circular relationships in your data model, there is a high potential for inconsistent data.

I resolved this issue!

- iii. Why do you have Units in InstanceAttributeScenarioData, but then have Units connected to TimeSeriesMetadata – why put units in both places when there is potential to get them wrong?
- iv. How do the Units specified in InstanceAttributeScenarioData apply to FileBased data? How do they apply to Binary data? How do they apply to Text data?
- v. There are others...

The units are connected to the InstanceAttributeScenarioData table because the unit is common for all data types. In case of Text Data and File Based, the user shall choose the unit = dimensionless and I pointed out to this fact in the manuscript. The unit Table is connected also to the Time Series Table so the user can choose the unit of the time.

10. Page 5, you state “...WaM-DaM enforces controlled vocabularies on names of objects (e.g., Reservoir and River) and attributes (e.g., Dam Owner and Storage)...” Do you mean that WaM-DaM enforces CVs on the *names* of attributes or on the *values* of attributes? I’m assuming you mean on the names of the attributes because you wouldn’t be able to enforce on the values.

Correct, I am enforcing them at the names of the attributes and objects. I indicated that now to reflect this fact.

11. Page 5, you state “...if the text has no controlled vocabulary, the user chooses the field called “TextValueFree” which has no controlled vocabulary enforcement.” How would you know which field to query when trying to retrieve text data (e.g., when retrieving text data, how would you know which attribute to query? TextValueFree? TextValueControlled?)

I am no longer controlling text values for this paper. In WaMDaM 0.2 version I have two separate text tables, one for controlled values and another one for free values. In case of the TextValueControlled, I want to enforce values like Dam owner (e.g., federal, state, private) and dam type (e.g., wood, masonry, concrete)

12. Table 1 is confusing because it isn't really one table. You should show this as multiple tables with the relationships between the columns. The relationships shown in Figure 1 for multiple column data seem problematic – you have InstanceAttributeScenarioData connected to both Columns and MultiColumns. Have you worked through more than one example? I agree. I deleted that confusing table and now I replaced it with excerpts tables from the database that shows how the reservoir bathymetry is organized and connected.

13. Your paper should be written in the past tense. For example, in the first paragraph of Section 3, I assume you mean to say “WaM-DaM was physically implemented in a relational SQL Server database...” In general – you are presenting a data model that should technically be implementable in any RDBMS right?

Probably you are right but the present tense is the style that I write with throughout my paper until I reach the conclusions which I then speak in the past tense.

Correct, WaMDaM can be implemented in any RDBMS.

14. The first paragraph on page 7 (just under Figure 2) is a bit of a stretch. Typically one does not expose a relational database on a publicly accessible web server – unless there is something like a website or web service that is providing the access. Exporting a network configuration would require some code to get it exported in the correct format wouldn't it? I wouldn't have any problem with you describing these functions as part of the additional work you are doing, but a relational database doesn't do all of that stuff on its own.

Correct, now I point out to this fact by mentioning that future work will develop programming scripts to export network configurations...

15. Is CUAHSI the right community to seek feedback from on CVs for water management data?

I am no longer mentioning this sentence.

16. Need to add the NSF disclaimer to your Acknowledgments section.

Thanks! I added it.

Title: WaM-DaM: A Data Model to Organize, Share, and Publish Water Management Data
Authors: Adel M. Abdallah and David E. Rosenberg
from Sarah Null, April 15, 2014

Responses to Notes/Questions are in red by Adel Abdallah. May 14, 2014

This is a nice start to a conference proceedings. Thanks!

Below are notes and ideas to improve it:

1. You describe other data standards and model/databases to manage water resources data in your introduction; however, I found it funny to lump some of these in the same sentence because they do very different things (e.g., ArcHydro, WEAP, ODM, and DSS have quite different goals). This makes it tough to figure out what exactly Wam-dam will do. This is compounded because wam-dam is not described consistently (alternatively as a model to share, organize, discover, retrieve data, as a community standard...). I am not clear as to what this model/database does.

The contribution of WaMDaM is now better articulated and focused as a data model to organize and synthesize systems data. WaMDaM borrows many concepts of the above models and integrates these concepts together. I am no longer talking about sharing and discovering data and I abandoned the use "community standard"

2. More importantly, it is unclear how Wam-dam is novel or what it does that the other platforms/databases/models do not. In the last paragraph of section 1, you write, "WaM-DaM can serve as a new opportunity to expand the capabilities of HydroDesktop software (Ames et al., 2012) to discover and retrieve data about networks of water management infrastructure." How? I deleted this paragraph that refers to HydroDesktop. I added Table 1 which summarized the features that we need to support water resources data and that none of them support all of those features.

3. First sentence of section 2 seems overly narrow. What about economic demands, spatial data, habitat indices...? Thanks! I added economic and ecologic data. WaM-DaM represents network-based spatial data and this fact is mentioned throughout

4. Figure 1 is fairly confusing if you're seeing it for the first time. It might help to take a dataset and show how it would be input to the data model to add clarity. Often the accompanying text feels vague. Add clarity and detail whenever possible (or describe examples). I agree that Figure 1 could be confusing. I added Figure 3 to show how data is organized in WaM-DaM.

5. The controlled vocabulary approach seems backward to me. What happens when the water resources community does not agree upon terms to describe objects/attributes? Will this approach collapse?

The purpose of the controlled vocabulary is to enforce consistency in the database. I am no longer mentioning that WaM-DaM is a community standard. We will let our collaborators test WaM-DaM and see if the provided controlled vocabulary work for them. The user though still can add their own controlled vocabulary if they don't like the provided one. The controlled vocabulary concept has been proved to be successful within the Hydrologic Community in the example of the Observation Data Model (ODM)

It seems a better way to go is for wam-dam to have flexibility to incorporate any vocabulary (but perhaps the more consistent/logical the vocabulary is, the more likely the model/data is to be used by others in the community).

That's correct as long as the user uses consistent vocabulary terms. I mentioned that we didn't formulate a full set of controlled vocabularies and future work will do so.

The example you provide using the term reservoir (versus dam) is particularly troubling because reservoirs and dams are not the same. The dam is the infrastructure and may have attributes like height, type (e.g., concrete arch), spillway capacity..., while reservoir is the water body behind the dam and would have attributes like surface area, capacity, dead storage...

You are right and technically a dam is different than a reservoir. However, the Army Corps of Engineers uses the term "Dam" to describe both the structure and the water body at the same time.

6. It would help if you could briefly discuss what data are required in wam-dam. For instance, are metadata fields required? I can see benefit in requiring them, but I could also see being frustrated as a user if I were required to enter fields when I just wanted to get a dataset input quickly.

The required fields of data are indicated in WaMDaM schema in Figure 1 in bold font. You are right, there will be an option in WaMDaM to let the user choose if they want to share or publish their WaMDaM database with others. If they choose to keep it for themselves then WaM-DaM will require less metadata so the user can populate their database quicker. Otherwise if they choose to share the data, then they are required to fill the metadata like units, sources, and methods so the other use can interpret the data correctly.

7. Work on the transition into the Bear River case study to improve readability.

Sure. I already improved this transition

8. I like that you included the Little Bear example as a case study. I would be careful of your language for importing data however. The process sounds rather cumbersome if you had a study area with many data sources (downloading them from models that already had data standards, putting them into excel, then one by one importing them to wam-dam). I would really emphasize the benefits of what wam-dam has to offer and why modelers should use it. As is, the utility isn't rising to the top, but the process seems labor-intensive.

Correct, I pointed out that the use of excel was for simplicity but future work will automate this process by using Python scripts to load data to WaM-DaM seemingly.

As we discussed in your meeting, you should also be wary of using terms like faster and easier that you probably won't be able to measure or monitor. Also, discussion of whether tools exist to import, format, and visualize data should be included.

Sure! I am no longer stating the terms "faster and easier". I already added more discussion on the tools to import, format, and visualize data.

9. Be consistent with your datasource names. You alternatively call the National Inventory of Dams, National Atlas of Dams and National Atlas of Major Dams. In a paper where naming conventions matter, this seems very sloppy.

Thanks! I addressed this comment throughout.

10. More discussion of how wam-dam will be adopted by the water community would improve this paper immensely. Especially since there are similar models which have not been widely accepted. Why will this effort be different? What is being done to ensure that it is adopted. Sentences like "Once WaM-DaM is adopted by the water resources community, there are four benefits to use..." sound naïve.

I am no longer referring to WaM-DaM as a community standard. However I already articulated how WaM-DaM is different than other models mainly in Table 1.

Thank you all for your critical feedback and led to significantly improve the manuscript
Adel