**Water Information and Data Subcommittee**

**Data Exchange Workgroup (Workgroup #4)**

Minutes – January 23, 2012

**Attendees**: Steve Tessler (USGS), Dharhas Pothina (TX), Paul Houser (Reclamation), Sara Larsen (WSWC), Dwane Young (WSWC/WestFAST)

**Next Call:** Week of February 6th, look for a Doodle Poll

**Administrative:** The minutes from the prior call were approved. Action item #1 was closed. We also introduced Sara Larsen, who is a new hire with the WSWC. She will be working almost exclusively on the Water Use Data Exchange. Before joining the WSWC, she was a research engineer at Los Alamos National Laboratory in the Energy and Infrastructure Analysis Group, which models municipal water and energy infrastructure interactions and vulnerabilities during natural hazard events and scenarios. She has also worked for the State of Utah’s Division of Water Resources as an analyst, and for the Provo River Water Users Association – operator of the Deer Creek Division of the Provo River Project in Utah – as their GIS coordinator. She acquired a B.S. in Geography and a M.S. in Civil Engineering, with a water resources emphasis, from the University of Utah. She has written a thesis, several technical reports, and a book chapter on water and energy topics. The WSWC is fortunate to have her involved in this important project.

**Review Issues Document:** The group reviewed the ‘Issues Document’ that had been distributed by Dwane. The group agreed to skip straight to the issues section of the document, and to begin discussion around those issues. The first section of issues deal with System Design and Development. The first issue under this category is ‘Platform/Software’. The document describes three options:

1a) Develop using a Java Environment; 1b) Develop using a .Net environment; and 1c) Develop using both. Dharhas suggested adding a fourth option: 1d) Develop using Python. The group felt that it would be important that whatever was developed could be cross-platform. This requirement would almost exclude the .Net environment. For the Java environment, the group felt that some of the pros were that it is cross-platform, it might be in line with a number of state environments, as well as USGS. Some cons to this approach were that many states may not have the in-house expertise to maintain software written in Java, and that there weren’t as many scientific libraries available as compared with something like Python. For the .Net environment, some of the pros were that it might be used more broadly by the states, however, because it is not cross-platform, there may be significant issues for states that do not run Microsoft Servers. The group felt that developing in both .Net and Java would be very difficult, even if it did promote broader adoption. There was considerable discussion about the Python recommendation. Texas has a done a great deal of web service development using Python, and would be able to share some of their code with the group. Python would also support a cross-platform implementation. Texas has had some issues convincing their IT staff that Python is a mainstream language. They also have not yet had success in getting the services to run on an ISS server. **Recommendation:** *The WSWC will leverage the work that Texas has already done and look at their Python services. We also will consider Java as a fallback position. We will need to evaluate the State Capabilities Assessment Survey responses to get a better sense on what the current environments are within the states to see if Python would be a viable language to develop in.*

The next issue dealt with which database platform to use for the system. Dwane had identified 5 options: 2a) Oracle; 2b) SQL Server; 2c) Open Source (i.e. PostGreSQL); 2d) Multi-platform approach; and 2e) target like-platform states to include in ‘proof-of-concept’ project. The group spent most of the time discussing option 2d. Texas has implemented a Python application that can connect to any number of different database platforms and provide services. The original concept of the system would have involved using a ‘staging database’ that would be deployed in each state. The states would then develop extract-transform-and-load (ETL) procedures that would migrate data to this staging database on a routine basis (daily, weekly, monthly – timeframe to be determined by the state). The web services and catalog services would run off this staging database. The approach that Texas has taken removes the need for these staging databases. The states would develop queries off their existing database that would provide the information needed for a ‘Data Access Object’ (DAO) to connect to the database and provide the data to the services. The DAO has to be recoded for each database platform, and potentially for each database structure, but the core is similar enough that it could be maintainable across the many platforms. Further research will be needed to determine whether it would be easier to maintain this DAO or the staging database. **Recommendation:** *As an experiment, the WSWC will download the code from TX and attempt to implement it. That exercise will inform the group on how easy it would be to deploy these DAO’s to other states.* The group also discussed the implementation and design of the catalog services. Dwane explained that the catalogs were meant to provide a higher level summary of the data that are available. They should respond almost instantaneously to a service request. The catalogs will most likely be organized around geographic location. Texas pointed out that there will still likely be scenarios where the catalogs wouldn’t help. For example, it would be difficult to build a catalog that provided the answer to the question: give me all the stations with a flow rate higher than 10 CFS. The design of the catalogs would be driven by the Portal user interface that will be developed by the WSWC. Dwane envisions that each state would maintain a catalog of their data. That catalog could then be consumed by the central portal to populate a central catalog on some routine basis. A question was asked if the OGC Catalog Services would be of help. Dwane said that his understanding of the OGC services were that they defined a different type of catalog. The OGC catalog describes a catalog of services that are available, as opposed to a catalog of data that are available.

The next issue dealt with the security of the data in the system. There is an outstanding question of whether or not the data provided by the system would need to be enveloped within a security model. The group did not have a definitive answer to the question of whether or not there would be data that would not be publicly available, but would be available to other secure partners (i.e. other states). The feeling was that if the data weren’t available to the public, then the data probably wouldn’t be available to other partners either. The overall feeling of the group was that the services should be completely open, and that if a state has data that they don’t want to be made publicly available, that they wouldn’t include that data in the services. This keeps the control of what data are shared with the service provided. Dwane said that he would raise this issue with the Water Information and Data Subcommittee at their next call in February. Dwane also explained that if a security model is needed, we would probably benefit from leveraging the existing security model provided by EPA’s Exchange Network. However, if a security model is not needed, then this approach may be overkill. The other benefit in using an Exchange Network approach is that we might be able to leverage existing capability in the state DEQ agencies. Texas pointed out that there may be significant hurdles in trying to make that collaboration work between the various agencies. Their IT systems are likely not integrated enough to provide that type of collaboration. For example, it would be difficult for TWDB to make use of TCEQ’s Exchange Network Node. **Recommendation:** *For now, we will work under the assumption that a security model is not needed. Dwane will explore this further with the Water Information and Data Subcommittee.* The group also discussed other types of security (for example securing against a denial of service attack). The group felt that the services needed to be designed such that if a request is too large that the user be notified that the request is too large and that the response to that request be processed in an asynchronous manner as opposed to a synchronous one.

**Call Schedule:** The next call would be scheduled for the week of February 6th. Dwane said that the only days that worked for him that week were Monday and Tuesday. He will send a Doodle Poll to the workgroup to check on their schedules.