

Notes from Catalog Working Group Meeting, 7/13/2013

Attending: Julie LaRock (SFWMD), Susan Markley (Miami-Dade), Denise Miller (FDEP), Andy Woeber (FDEP), Lisa Van Houdt (FDEP, moderator), Gail Sloane (FDEP).

The goal of the meeting is to review questions and issues of FRD, and answer any questions, explain items to understand, take comments.

- 1. Approach for selecting stations that properly represent the water resource:** - *Rich Hammond, Water Atlas GIS Team Manager discussed how USF will perform station to water resource match and showed examples with such matching using a draft station and water resource map layer. At present, the approach for the Catalog is a one to one match (station to water resource).*

Andy Woeber, new DEP GIS analyst, and Rich Hammond made progress on how to move forward using the latest NHD layer, which was posted by FDEP on USGS website last week. Rich illustrated the issues associated with having unnamed canals in the Catalog, as well as 3 examples of conflicts assigning stations to waterbodies. Andy says there are three main layers in NHD to be concerned with: Area (watersheds/basins), Flowlines (streams) and Waterbodies (lakes, ponds). USF will begin matching sites using the NHD (3 layers), so the layer to select might be tricky. Andy suggested collapsing layers to a point layer. While this might make the matching process easier, it might also create other issues. There is a tradeoff with different water resource types. For example, being able to look at which of the three main layers a resource is listed on could help with matching. Will run some test matches to check this out.

Reach codes – a crosswalk is being developed for permanent water resource ID.

Comments/Proposed solutions:

- unnamed waterbodies: can use permanent waterbody ID. This will suffice for now, but the project is iterative, and better solutions may be identified in the future.
- Will need to make a decision during design to assign a site to one waterbody or multiple waterbodies (estuaries, harbors, inlets, etc) Will address during the spatial design phase.
- Waterbody site assignments for the Catalog will be long-term, over the life of the project.

If USF wants to share its coastal details, Edwin will have e-mail address for USF to send NHD updates or reports as they are found. The Estuary class has not been defined (yet) for NHD. There is an “S-code” but no “genus”. Estuary definition is a planned future enhancement. Bays have not yet been updated, and these, Gulf, and Oceans are in separate waterbody classes.

- 2. QA checks to include checks on free text.** – *Jason Scolaro, Database Team Manager and Keith Bornhorst, Web Team Manager, discussed the method that will be used to ensure the quality of metadata data for the live Catalog. This will include the use of data domains and word count (property of input form) and other automatic data checking approaches. USF discussed the possibility of adding a “Report Issues Button” that would allow data providers and others to communicate problems with a data record etc.*

Jason and Keith discussed three methods of metadata entry into the Catalog database:

- a. Providers enter data into online forms
- b. USF will allow batch uploads of CSV files from web form

- c. Working with STORET data providers, USF will harvest and load the metadata into the system.

All 3 methods will hit up against QA checks. It will be difficult to check free text using algorithms.

Comments/Proposed solutions:

- USF already checks expressions on URLs in its WaterAtlas, so will mimic this.
- QA checks can be run against batch uploads, which can then be sent to Rich for spatial checks (ArcQA validation).
- Will have specific information stored with data fields/forms to reflect no. of characters, character type, etc., as well as check boxes.
- USF will develop more complex algorithms to check fields against each other (Example: Station Type vs Water Resource Name)
- QA checks will evolve with project
- Batch uploads will be available via web forms using CSV files

3. **“Organization” data element and any set relationships and relationship types;** *Jason discussed the possible need for the Organizational User role and the functional diagram for the Logical Data Model (FRD section 2.2.2). Combined with Item #6 –data flow diagram*
Metadata emphasizes projects and organization, but need to focus on users.
Should Organization user be assigned to organization and have access to all projects, or only to individual projects?

Comments/Proposed solutions:

- Projects and stations are the major types of metadata elements
- Organization type and users are critical; may want to store other org information, but association of users to orgs is very important.
- Should user have access to all metadata of org, or permission only to individual projects?
Catalog Workgroup will need to address 8/28/13.
- Denise indicated that WIN uses Organization roles and Org User accounts; will send WIN info to USF.

4. **The roles for Metadata Provider, Metadata Administrator and System Administrator**
The Metadata Administrator role has been removed; *Jason discussed the reason for this role and approaches that might be taken to allow a similar role for FDEP.*

Originally, 4 types of users:

- a. Metadata Provider: ability to add or edit metadata
- b. System Admin: global power over metadata
- c. Metadata Administrator: DEP role to review & approve metadata
- d. public users: visitors to Catalogue

Metadata provider responsibilities will fall on many folks in many organizations

Comments/Proposed solutions:

- Metadata Administrator is now consolidated with System Admin., which assumes all MA roles

- Access log, as in who edited what, also called audit trail; trail will address changes to db.
- Metadata provider will have access to Administrator- flagged data, make corrections, etc.
- All data available to public, whether flagged or not.
- Public user can be anyone; do not need to log in or register.
- Can browse orgs, areas, or parameters, or do attribute or spatial search to obtain results.
- Change: Save searches or frequently-asked searches (FAS)? Will consider in design effort

5. Data Flow Diagrams in detail; *Jason discussed data flow using Figures 1-3 in the FRD section 2.2.1*

6. Documents as metadata and how the Catalog will accommodate if at all; *Documents are in the same category as data and are not managed by the Catalog. There are three data elements (Project, Station, Data) that have either a required or optional link field that could be used to link to documents that provide additional information. Keith will also discuss the website URL as a side discussion to this.*

Comments/Proposed solutions:

Will not currently document other metadata docs as not in SOS, but may eventually link to these . . . FFWCC has metadata portal (GAME)for researching documents, so USF could do this if desired eventually.

7. Bulk import and import of data in general as an FRD requirement: *Jason discussed. The term “bulk” has been removed. The initial intent of its use in the FRD was to indicate work that will be accomplished by the Water Atlas data team in gathering metadata from existing (STORET, USGS and possibly others) databases with a bulk download.*

Comments/Proposed solutions:

- Will do batch imports for station data only in Catalog
- Bulk import is one-time load from STORET, harvest metadata
- USF should harvest STORET metadata annually after FDEP does its annual STORET update. This is agreed upon by USF and FDEP
- Metadata Providers will enter STORET metadata manually in elements that are tinted red on Jason’s 7/11/13 Powerpoint presentation; e.g. no match with STORET
- Flag *non-required* blank fields as ok
- flag *required* blank fields as errors

8. The use of links in the metadata summary: *This item is also discussed in item 5 above. There are three link fields (URL fields) to allow user to link to a website for additional information on a project, a station or a data set (parametric data).*

USF has link checker that can be used on this project

Comments/Proposed solutions:

- Recommend eliminating station and project URL fields to minimize data provider effort
- Keep data URL
- *Catalog Workgroup will need to address 8/28/13.*

9. **Method of changing or adding to the metadata element domains and when the ability to make such changes will be frozen for the beta design phase:** *Jason discussed. Essentially, we can allow changes to domains as long as the project is being supported; however, changes to the database schema cannot be easily accommodated so we have stipulated the data schema freeze date of August 20, 2013. We may allow changes after this date to the schema but we must agree to the changes. We use the term “without contractor concurrence” to indicate this in the FRD. We can discuss what it would mean to change this date.*

Comments/Proposed solutions:

- USF and FDEP agreed that any element changes after the 8/20/13 deadline must be agreed upon by both parties.
- Timeline: final FRD will be agreed upon by Council at quarterly meeting.

Remaining items:

1. **Audit trail method:**
See No. 2, above.
2. **Update frequency and methods used by USF to ensure data security:**
Update frequency addressed above in #7. Jason confirmed with USF that nightly backup exists; sends tapes offsite; data is secure
3. **output formats, including shapefiles:**
Rick: shapefiles will be able to be extracted. Will be more tricky to extract text elements as shapefile; can download as text, converted into shapefile and re-projected by user.
4. **Help Sections to be added (then add a User Help requirement to the FRD):**
Keith to add User Help section requirement to FRD. FRD is general, so will flesh this out later.
5. **The responsibility of the FWRMC to provide content for catalog pages:**
Yes, there will be other elements needing to be added by FWRMC members. Desire is make this easy.
6. **The long term maintenance of the catalog:**
Contract is 2 years of maintenance after implementation; Catalog will need maintenance after that to be useful.
Andy: could metadata providers assist with maintenance? Jim doesn't know how long-term maintenance has been considered.
7. **Process for providing user permissions:**
All permissions will be made by USF; many metadata providers need to obtain permissions to do update work.
8. **The process for correcting metadata that are flagged:**
If data element is flagged, will be in db, but any user can see the flag (incorrect or incomplete). Metadata provider will get report of flagged data, so can correct. USF will hold different meetings to develop this.

9. Desired search methods used in creating a attribute search:

Pertains to searchable fields/options ; will need to have advanced search options. USF will work with Catalogue workgroup and FWRMC to develop. Cannot answer now, but will conduct analysis to see what makes sense.

Additional:

Julie LaRock expressed concern about the effort needed to update the elements

Susan Markley has concerns; provides data and metadata to state in different ways; MD county may not be able to keep up the Catalog. Everyone recognizes need for Catalog, but should be kept simple. Bulk uploads are OK, but the more fields there are, the more fields to enter.

Jason: could add pick list to ID upload formats if this would make it easier for data providers. Julie thought this would be helpful, but Jason hasn't really looked into this yet. USF expressed willingness to work with FDEP, FWRMC, and others to make this easy as possible.

Keith: To allow Metadata Providers to repurpose data upload files, it may be possible for USF to create templates for metadata upload that match formats for the data already being sent to agencies. This would not be practical for *all* agencies, but might be indicated for the most common data uploads.