12-03-02 BIEN db Conference Call

Participants: Aaron, Mark, Martha, Brad, Mike

Unavailable: Jim, Bob

**General Business**

* Daylight Savings Time
	+ Will change the time of the meeting for Tucsonans.
	+ Switching to 1 PM works for Brad.
* Rescheduling call week of April 1
	+ The Friday, April 6 call has been rescheduled at 1 PM PDT Mon April 9 to accommodate Aaron’s schedule.
	+ This is a one-time change.

**Sign off on first draft of schemas (VegBIEN, VegX, DwC) or more discussion?**

* Brad and Mike have signed off on all three draft schemas, though there was discussion of VegX (see below).
* We need Bob’s OK too.
* Martha will remind him to review them and sign off.

**Discussion of VegX schema**

* BB: Aaron, where are you on refactoring VegX?
* A: If want to officially add first class fields for the user defined fields, we should edit the XSD.
* MS: If edit the VegX XSD, we are forking the standard.
* BB: Maybe we now need to actively engage Nick, Mikael and \_\_\_.
	+ Crucial to have Nick’s input.
	+ If Nick signed off on the changes, would be good to have them rolled into VegX.
* MS: Yes essential.
* A: We can always user defined fields, but prefer to have them be first class fields.
* BB: Aaron, what are the next steps for VegX?
* A: to make it easy to import VegX row by row instead of having to import the whole thing.
	+ If each plot could be self contained, importing would be easier.
	+ It’s a large change to move things from separate tables to nest them inside each other.
	+ Would break other tools built on VegX.
	+ This actually impedes the mapping process even more than the import process.
* BB: Mapping is a major obstacle to adoption of VegX. It’s since so difficult.
	+ Would break the NVS developed tool, which is the only tool available.
* BB: sounds like a blocker to loading plot data, so really need to engage Nick. (later discussion shows it’s not really a blocker)
* MS: The import tool is a dot net tool, which is limiting in itself.
* BB: If we make necessary fixes to VegX, we’ll need to consider developing a tool for importing.
* MS: as long as we don’t put burden on Aaron to develop the tool.
	+ Aaron can do mappings ad hoc.
	+ But considering developing an import is OK for later.
* BB: We don’t need Nick to commit to a timeline for making changes.
	+ We don’t want to hold things up waiting for that.
* Aaron will update VegX schema to reflect the changes that have been discussed:
	+ First class fields
	+ Nesting of tables
* BB: How soon can this be dealt with so we can begin loading plots?
* A: We can load plots using user defined fields.
	+ The first class fields and nesting of tables would be “nice to have” changes.

**Loading data**

* BB: We should do two things this week:
	+ Decide about making the changes to VegX (Brad signing off on first class fields and the changes discussed).
	+ Move ahead on loading specimen data.
* There is raw data source for specimens and there are scripts that load those data into staging table.
* Look at php, inv files.
* A: Spreadsheet with all data sources is on Redmine: Bien\_data\_sources.xlsx
* BB: For plot data:
	+ Getting new CTFS dump, so don’t load that data.
	+ Don’t try loading FIA. It’s problematic, missing crucial meta data.
	+ Don’t reload TEAM, they’ll be sending new data.
* All specimen data can be loaded.
* MS: Keep track of performance issues. If taking over night, may want to talk to NIrav and Martha about iPlant resources.
* A: Have 4 more cores on VegBIEN dev, so can run 4-6 loads simultaneously.
* MS: That will help.
	+ Map and load the specimen data.
* BB; Need to do careful validations.
* MS: Issue some queries on SALVIAS and then issue them on VegBIEN. That should be the process.
* BB: Yes that’s the plan
* A: Already have the queries and outputs. There will be some tweaking.
* BB: We need to develop validation queries for specimen data.
	+ The raw files are the ones to use.
* A: I’ve been creating a separate little database for each source imported.
* BB: About validationas, I should write queries against the raw data files and make sure we get same result from VegBIEN?
* A: We have the queries for NYBG.
* BB: I can recreate those queries for other sources.
* MS: Since DarwinCore is a flat file data object, we should compare the DwC flat files.
* A: You want to export a flat file from VegBIEN?
* MS: yes
* BB: Need to remember that when we imported those, some were DwC, some were badly formed DwC, some not DwC.
	+ To reverse the process, will need to reverse the mapping.
* A: So I need to reverse the mapping?
* MS: Sorry, yes
* A: Designed to be reversible, but will be a little work.
* MS: If DwC went in, we should get DwC out.
	+ For VegX in, what comes out?
* BB: Export VegX xml and VegX flattened table
* MS: Essentially equivalent to view full occurrence with VegX tags.
* BB: Yes.
* BB: So were does that leave us on validating specimens?
* A: We have a dozen data sources
* MS: Let’s get NYBG right first.
* BB: Yes, for each data source, load it, validate it. Work through all data sources.

**Returning to VegX**

* MN: I want to be clear on what needs to be done to resolve the VegX blocker.
* MS: Remember it’s not actually a blocker since Aaron can do ad hoc mappings to load data.
	+ The plan is to review the VegX changes needed.
	+ Aaron will organize this discussion.
	+ Brad, Mark, Bob, Aaron will do that.
	+ Then, once can make the case for those changes, Aaron will contact Nick.

**To Do List**

* Martha will remind Bob to review revised schemas (VegBIEN, DwC, VegX) and sign off on them.
* All: The Friday, April 6 call has been rescheduled at 1 PM PDT Mon April 9 to accommodate Aaron’s schedule. This is a one-time change. Mark your calendars.
* Aaron will update VegX schema to reflect the changes that have been discussed:
	+ First class fields
	+ Nesting of tables
* Aaron: For VegX
	+ Organize the review of the VegX changes needed. (Brad, Mark, Bob, Aaron will participate).
	+ Once can make the case for those changes, Aaron will contact Nick.
* Aaron: Reverse the DwC mapping so validations can be done.