12-02-24 BIEN db

Participants: Aaron, Mark, Brad, Bob, Mike, Martha, Jim

**Milestone Revise Schemas I**

* Next week, will we be finished with the first round of revisions to VegBIEN schema, DwC-VegBIEN mapping?
	+ *VegBIEN schema: Some things we might want to include in this round:*
		- *Converting all user-defined fields to first-class fields*
		- *Adding tables to store traits information*
	+ *DwC-VegBIEN mapping: We have mappings for DwC elements used by NYBG*
* Are the corresponding requests for changes to VegX, VegBank documented?
	+ *VegBank: The only change we have made to VegBank itself is to remove the length constraints on text fields. Everything else is a VegBIEN-only change.*
	+ *VegX: We have not started revising VegX yet, as we have been focusing on VegBIEN revision. Here are some possible changes:*
		- *Converting all user-defined fields to first-class fields*
		- *Allowing all top-level tables to be nested inside the ID elements that point to them*

**VegBIEN schema**

* Methods tables
	+ *Are there any more first-class fields we want to add? (Data providers can always create user-defined method fields for any fields that don't already exist.)*

**Shifting focus to data loading**

* Specimens
	+ *Full NYBG data has been loaded on VegBIENdev (not yet validated)*
* Plots
	+ *Full SALVIAS data has been loaded on nimoy*
	+ *Getting the validations to pass is in progress (involves adjusting mappings and configuring duplicate elimination on all relevant tables)*

**Notes**

**Picking up on VegBIEN schema discussion from email comments.**

* ML: Taxonomic inclusion/exclusion linking table
	+ BB: Growth form?
	+ ML: Just a generic inclusion/exclusion. What did you ignore/consider?
	+ Tree shrub liana, herb, woody, non, cultivated; The list added to the spreadsheet.
	+ ML: It would be a cross-reference table to handle inclusion/exclusion.
* MS: Relational model has limited ability to handle complicated rule sets. Well-known problem with databases.
	+ Aaron could use triggers or constraints to handle these situations.
	+ Might give us most flexible way to handle complexities of methods.
* BB: For now can we live with schema we have?
	+ A: Should we add min/max constraints (liana dbh) as additional optional field?
	+ BB: Do we store a link to method? Add it.
* BB: Is discussion of methods done?
	+ A: See if VegBIEN schema looks correct after making all these changes.
* BB: Current BIEN db has embedded a taxonomic hierarchy by using a left and right pointers traversal. Is there anything in plantconcept that supports that sort of hierarchy?
	+ A: yes. Plantparent\_id in plantstatus table
	+ MS: Out of scope for us to do that internally. Could link to it.
	+ BB: TNRS now working on APGIII vs NCBI view of hierarchy.
	+ MS: Should let the external source handle that.
	+ BB: Not sure how that would work performance wise.
	+ BB: Can walk away from it, but would be nice to have a classification table.
	+ BB: Mike said can’t point to an entire hierarchy that didn’t get changed all the time.
	+ ML: They can. Problem occurs when people don’t map into USDA plants.
	+ BB: Can get orphan taxa if genus is not in the hierarchy.
		- But it looks like the pieces are there.
	+ MS: is this to support a higher order taxonomic level query? Could include one or two higher taxonomic levels. Be clear it might not be the most recent version.
	+ BB: maybe such a hierarchy belongs in the analytical layer, not here.
	+ MS: Go to a separate service that does that.
	+ BB: Convinced. It’s out of scope. Hierarchy resolution service is a separate issue.

**Milestone Revise Schemas I**

* BB: Goal for end of February is a stable draft VegBIEN and import schemas.
	+ And comprehensive DwC schema.
	+ Are we happy enough with the VegBIEN schema?
	+ With DwC?
* BB: Wants to take another look at DwC.
* BB: Are we happy with the current state of VegX?
	+ A: We haven’t really started refactoring VegX.
* BB: Want to finish refactoring to DwC.
	+ A: We haven’t been adding DwC fields. Have just been using what is there. May need more than NYBG data to find limitations.
	+ MS: some of DwC fields seem to be under described.
		1. We may need to get back to them about those fields with feedback.
		2. It’s a well-known problem that different people interpret DwC fields differently.
	+ BB: Aaron, what changes do you think are needed in DwC?
		1. A: datasourcename is the only custom field we’ve added.
	+ MS: Need to have other DwC data sources to look at.
		1. BB: We have them MOBOT (not very DwC compliant), UA, CTFS
		2. Can load DwC data.
* But for plot data need to refactor VegX
	+ MS: Need to figure out whether Shash’s or Aaron’s approach is easier to use.
	+ BB: Eric’s TEAM plots are large and complex. Might be a good test for VegX schema
	+ MS: Did Rick or Dolans give access to raw schema?
		1. A: We have a local copy of db, but don’t know if it is most recent schema.
		2. MS: Aaron is having trouble interpreting Shasha’s records. Mapping VegX to VegBIEN.
			1. Cant’ check if Shasha’s CTFS mapping into VegX is accurate.
	+ BB: Will email Rick and copy Mark
	+ A: We have an April 2011 version of data dump from them.
	+ BB: Confirm with Shash that is most recent schema.
	+ BB: That April 2011 version looks fine. (so no need to email Rick)
	+ MS: Need the glossary or data dictionary
		1. A: Have the ERD
		2. MS: That’s not enough.
		3. Ask Steve for glossary or data dictionary.

**Loading Data**

* Mappings may need some adjustment as we load, but ready to begin loading both specimen and plot data after final sign-off on VegBIEN schema (ERD).
* Aaron send out final ERD and get sign off.
* Then ready to move on to refactoring VegX

**To Do List**

* Aaron make the changes to VegBIEN ERD, send it out.
* Brad, Bob, Mike review and sign off on the ERD.
* Aaron ask Steve for glossary or data dictionary for CTFS.