**BIEN DB Meeting – 5 January 2012**

Brad's notes

In order of expected completion. The priorities below cover most of this month. Top priorities for the next week or two are 1-5.

Priorities:

1. Modifiy VegBIEN schema to incorporate all of Bob suggested changes to support correct mapping between individuals and stems, etc.
2. Create direct mapping & import scripts from VegX🡪VegBIEN
3. Identify critical fields in VegBIEN, and modify constraints **(Brad & Aaron; may need input from Bob or Mike Lee)**

These are fields which block a record from importing if it cannot be parsed or otherwise violates constraints. All other fields should be set to NULL if value cannot be imported, and the import error reported and logged. In most cases, it is better to set a particular value to NULL than to skip an entire record (in other words, participation [relational constraints] are more important than business rules or data type constraints for particular columns). IMHO, at least for plot data. It may be necessary to modify some FKs and relations in the VegBIEN schema to accommodate these changes; mostly I suspect we will be "loosening" constraints. In my experience, VegBank has a number of mandatory participations and requires fields would should be set to optional for VegBIEN.

1. SALVIAS data (plots)
   1. Complete mapping of SALVIAS🡪VegX
   2. Expand VegX🡪VegBIEN import utility to accommodate all elements in SALVIAS VegX extract
   3. Run complete import of entire SALVIAS database
   4. Run all validations (**with help from Brad**)
   5. Makes changes as necessary to schemas and import scripts to fix any issues found
2. NYBG data (specimens; DwC)
   1. Complete mapping of NYBG(DwC)🡪VegX
   2. Expand VegX🡪VegBIEN import utility to accommodate all elements in NYBG VegX extract
   3. Run complete import of entire NYBG database
   4. Run all validations (**help from Brad**)
   5. Makes changes as necessary to schemas and import scripts to fix any issues found
3. CTFS (plots)
   1. **Work with Shash** to expand VegX🡪VegBIEN scripts to cover any elements present in the CTFS Panama data not previously included
   2. Import Panama plot data
   3. Validate (with help from **Rick**)
   4. **Work with Shash, Steve** to develop CTFS🡪VegX separate mappings for species-level inventories (this is a separate data set which Shash has not yet mapped to VegX; should be done separately, no reason to delay import of Panama plots)
   5. Modify VegX🡪VegBIEN scripts if necessary
   6. Import species inventory data
   7. Validate (with help from **Rick**)

Other data sources to be added (lower priority, after above is completely; roughly by the end of January):

1. NCU (specimens)
2. Aaron to work directly with **Mike Lee** to develop mapping for DwC dump from NCU database. **Brad & Bob** may be able to help as well.

**NCU data should be mapped to DwC, NOT VegX**. This is because NCU is herbarium data, which is much simpler than plot data. Most herbaria will be able to provide us with data dumps in this form; if they cannot (as with NCU) we should help them map to DwC, which they can use for other purposes. No herbarium database manager is going to be interested in mapping their data to something as complex as VegX. For this reason, we need a separate, generic DwC🡪VegX mapping. Thus, the import route for herbarium data should always be:

Herbarium DB 🡪 DwC 🡪 VegX 🡪 VegBIEN

As most herbaria will provide us with data already in DwC format, we will rarely have to do step one. The rest should be totally generic.

1. TurboVeg (plots)
   1. **Bob** will work on obtaining access
2. RAINFOR (plots)
   1. **Brad** to pester Gaby to respond to Aaron