BIEN3.0 Database Meeting – 12 January 2012

Suggested agenda (Brad)

1. BIEN3.0 Milestones

* Disucss, review comments
* Reach decisions about appropriateness of goals and timeline

1. VegX Mapping tool

* Background:
  + For specimen data, BIEN3 uses the following import sequence: sourceDb🡪DwC\_extract🡪VegX🡪VegBIEN
  + This sequence works well because DwC is a simple and widely-adopted exchange schema, so majority of specimen data sources take care of the first step themselves, providing data extracts in the form of DwC. The remainder of the pipeline needs to be built only once and can be automated
  + For plot data, the import sequence would be: sourceDb🡪VegX🡪VegBIEN
  + This is much more difficult for a number of reasons. VegX is a new schema, therefore no one has adopted it yet & we (Aaron) must do the mapping ourselves. Furthermo, most data providers are NOT willing to give us access to their databases, but instead prefer to provide us with custom extracts, everyone different, so that the actual import pipeline looks something like this: sourceDb🡪custom\_extract🡪VegX🡪VegBIEN
  + If this is the case, then using VegX adds little to the import procedure, and in some ways makes it more complicated by adding an extra step.
  + Even for Aaron scripting custom mappings is an extremely difficulat and complicated task. This process is rapidly becoming unsustainable
* Possible solution:
  + Develop **mapping tool** that assist Aaron and data providers with mapping their data (either the original database, or a custom extract) to VegX
  + A mapping tool is listed in the Milestones document under "Additional Goals" (components we discussed doing but which are not officially in our timeline or in the list of requirements). Should we make this an official deliverable? When?

1. Progress on items from last time:
   1. Developing direct, generic VegX🡪VegBIEN mapping and import scripts
   2. Validation of SALVIAS and NYBG data
   3. Update of CTFS-VegX extract
   4. Other items?