**BIEN Database Meeting – Dec. 15, 2011**

- Brad’s notes

**Mapping to VegBIEN via VegX**

* important to keep as priority mapping external sources to VegX, and importing to VegBIEN via VegX
* Will free us to make changes to core VegBIEN schema without having to change dozens of mappings to individual databases; would only have to adjust VegX🡪VegBIEN mapping
* Will increase use and support of VegX
* Therefore, any issues we identify with VegX should be fixed as soon as possible

**Ensure we can import complete data sets**

* Top priority right now is to ensure we can import complete datasets (for example, all of NYBG, or all of SALVIAS) via the schema
* The validation schemas Brad has provided work on the entire dataset; thus we will need the entire SALVIAS dataset in VegBIEN before we can begin validating
* It would be putting the cart before the horse to go after additional datasets (such as RAINFOR or GBIF) if we can't validate.
* Important to make sure we are ready and able to (a) import entire dataset and (b) validate entire dataset before we begin working with data providers, otherwise they may lose interest if they volunteer to help but we are not ready

**Obstacles to making changes to core schema?**

* Aaron: Changes to VegBIEN schema risk "breaking" VegBank web interface, which we were proposing to use
* Brad: Priority is to make sure core schema meets our needs and can handle all source data, even if that breaks link to VegBank interface. Jim: similar comment, we should be willing to improve schema as needed. Now is the time to do it.
* Recommend we adopt a **two-schema architecture:** (1) core "transactional" schema [=VegBIEN] for faithfully importing, scrubbing, validating and exporting data, (2) web schema, select-optimized (=VegBank?) to be used by web interface. Core schema would be periodically transformed to web schema by a separate set of mappings
* Having a separate web schema "protects" the web interface from changes made to core schema, and conversely, frees us to make any changes needed to core schema.
* If core schema is changed, only mapping to web schema would need to be changed

**Obstacles to full import & validation of SALVIAS plots database:**

1. Inability to correctly represent relationship between stems and individual organisms
   * Bob: Will send ASAP the VegBank model revisions so Aaron can fix the core schema
2. Difficulty of scaling up import via VegX:
   * Aaron: Can't move all attributes of a single record as a unit without loading the entire contents to memory
   * Brad: Important to get these comments back to Nick so he can make changes and commit to TDWG as well. Nick is very limited for time. I suggest we also involve the other two main co-authors (Miquel Cáceres and Martin Kleinkampf). Perhaps if all three are involved the changes will happen more quickly. Also, it would be helpful to copy Susan Wiser ([WiserS@landcareresearch.co.nz](mailto:WiserS@landcareresearch.co.nz)) onto any correspondence concerning VegX as she is good at keeping things moving along. Here are the emails I have for Miquel and Martin (Nick may have more up-to-date emails if these don't work):

Miquel De Cáceres Ainsa [miquelcaceres@gmail.com](mailto:miquelcaceres@gmail.com)

Martin Kleikamp [Martin.Kleikamp@web.de](mailto:Martin.Kleikamp@web.de)

**Deliverables (in order):**

1. send revised VegBIEN schema to Aaron (Bob)
2. Make changes to VegBIEN schema (Aaron)
3. Mapping from VegBIEN to original VegBank (the latter to be used as web schema for BIEN web interface)
4. Complete full dataset imports to VegBIEN via VegX of NYBG and SALVIAS (Aaron)
   * Identify and make changes to VegX needed to enable full-dataset imports
   * Or report changes needed to Nick, Miquel and Martin and have them make changes
   * Even if you make changes yourself, but sure to report to Nick et al. so these changes change be incorporated to the draft schema on TDWG
5. Complete full-dataset validations for NYBG & SALVIAS (Aaron and Brad)
6. Acquire additional plot data sets from providers willing to work with Aaron on mappings and validations (Bob: TurboVeg; Brad: RAINFOR, CTFS).
7. Acquire additional specimen data sets in both DwC and DwCA format, esp. GBIF (Brad)