**Issues with CVS & VegBank plot data**. **Revised October 5**

**Priority 1**

**9.** The admittedly weird modules and subplots of the CVS protocol are very badly handled. In the current version it is not uncommon for a species to have on the order of 15 records recorded for a single plot. Some of these have different numbers of individuals and some have difference cover values. Unfortunately, all are recorded as having the areas of the full plot, but a module is usually smaller than a plot , say .01 rather than .1 ha. The area associated with the species and cover or count can refer to the module or the full plot. In addition there can be separate records for different size classes of trees with no indication of either the size or the area associated with the record. This we need to discuss so that you understand the structure of the data and whether it all needs to be retained.

* **Michael Lee will provide a fresh export of CVS data with subplots data omitted and cover and stem data summarized across the entire plot.**

**10.** I observe that some, but not all, plots in VegBank dump do not have the plot name populated ([plotName\_\_@VegX\_\_.plot@vegpath.org](mailto:plotName__@VegX__.plot@vegpath.org)). I have not seen this problem in CVS, but then I have seen only a small dump of data]

**1.** locality\_\_@DwC\_\_@vegpath.org: In some cases (eg 040-04-0144) this is actually presenting “Location Narrative”, and not “Author Location” as needed, whereas in other cases where Location narrative is not populated then we see the Author location data as needed. This should be strictly “Author Location”.

* **Regards #1, it appears that “Author Location” is always present in CVS, so all that needs to be done for CVS is to omit the Location Narrative from the “Locality” field.**
* **However, some plots in the VegBank view do lack Author Location. This may be the same problem as #10 with plotName missing and could be a problem with the VegBank data. Michael Lee needs to verify that these data are in VegBank, and if necessary correct VegBank or instruct Aaron as to conditions for omitting from BIEN plot data that are in VegBank.]**

**4.** Morphospecies have non-alphabetical characters scrubbed out. For example Hypericum [graveolens + mitchellianum] is rendered as Hypericum [graveolens mitchellianum]. Those extra characters in the morphospecies need to be retained. This may be a problem with all morphospecies in BIEN.

* **This is caused by a bug in TNRS but there is not sufficient time to get the bug fixed. Aaron should replace any “+” in a taxon name with “@” prior to passing the name through TNRS, and should then replace the @ with a + after the processing.**

**6.** I see many data lines duplicated for no obvious reason

* **Probably solved. Bob will check for duplicate lines in the next output.**

**Priority 2**

**3**. It was good to see communityConcept.name\_\_@VegX\_\_.communityDet@vegpath.org populated, but some critical associated fields were not present, including the Community Code and the fit and confidence values. Community code is very important and is needed for both VegBank and CVS.

* **At a minimum we need the CommunityCode added for both CVS and VegBank, and preferably Fit and Confidence. Michael Lee should clarify for Aaron where these fields are stored in the two databases.**

**5.** For every record “identifiedBy\_\_@DwC\_\_@vegpath.org” has the value “Robert Peet” and “dateIdentified\_\_@DwC\_\_@vegpath.org” has a value of “10/1/2008”. I think currently you are using the person who contributed the dataset (Peet) and the date of the contribution of the dataset (2008). I need clearer definitions of these two fields, but it appears they refer to the person who Identified the plant and on which date. These fields do have a home in CVS and we need to point you to these. If for some reason the field is blank the default should be the person who collected the plot and the date of the collection.

* **This is a problem with the CVS dataset. Michael Lee will correct this prior to providing a new copy of the data to Aaron.**

**Priority 3**

**7.** Do were really want to discard all soil data?

* **It is desirable but not essential to retain soil data. Low priority.**

**Priority unassigned**

**2.** I do not know the definitions of the fields georeferenceProtocol\_\_@DwC\_\_@vegpath.org & “geovalid\_bien”. However, they were blank for all records and I suspect they relate to geovalidation, in which case they should not be blank.

* **Seems simple, but not sure. Importance also not clear. Need a report from Aaron. I think this is a problem with other datasets as well.**

**8.** Cover values are given as the midpoint of a range with no indication of the range of the bin? A cover value of 0.505 seems very precise, but it is really the bin #2 in the CVS scale corresponding to 0.1-1% cover. This needs to be indicated in some way. It may be a problem with all the cover values in BIEN.

* **Not resolved. Further discussion needed.**