

BIEN 3 - Bug #953

CVS and VegBank data

09/17/2014 09:37 AM - Martha Narro

Status:	New	Start date:	09/17/2014
Priority:	High	Due date:	
Assignee:	Aaron Marcuse-Kubitza	% Done:	60%
Category:		Estimated time:	0.00 hour
Target version:			
Description			
Martha,			
Here is a rehash of issues with CVS. Several of these are likely to apply to Vegbank as well. I view the first 5 as critical and 6-9 as important.			
If you wish me to call you in the morning soon, let me know when you are available and which numbers to try. My cell might work here = 919-368-4971			
Best, Bob			
Issues with CVS plot data.			
1. locality__@DwC__@vegbiendev.nceas.ucsb.edu: 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".			
2. 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.			
3. 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.			
4. It was good to see communityConcept.name__@VegX__communityDet@vegbiendev.nceas.ucsb.edu 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.			
5. For every record "identifiedBy__@DwC__@vegbiendev.nceas.ucsb.edu" has the value "Robert Peet" and "dateIdentified__@DwC__@vegbiendev.nceas.ucsb.edu" 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			
6. I see many data lines duplicated for no obvious reason			
7. Do we really want to discard all soil data?			
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.			
9. I do not know the definitions of the fields georeferenceProtocol__@DwC__@vegbiendev.nceas.ucsb.edu & "geovalid_bien". However, they were blank for all records and I suspect they relate to geovalidation, in which case they should not be blank.			

History

#1 - 10/16/2014 10:59 AM - Aaron Marcuse-Kubitza

- Tracker changed from Task to Bug

#2 - 10/16/2014 11:02 AM - Aaron Marcuse-Kubitza

- % Done changed from 0 to 50

#3 - 10/17/2014 03:09 PM - Aaron Marcuse-Kubitza

- % Done changed from 50 to 60

#4 - 04/28/2017 04:30 AM - Aaron Marcuse-Kubitza

- Description updated