I have looked over the CVS view that Aaron sent. I did not have time to go through it all, but I picked a few examples to explore. Here are some issues I discovered.

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.

For lack of an easy solution I did not check whether confidential data are being kept confidential. This is something Mike Lee could check with a more complete download when he returns from DownUnder in middle September.

I found it strange that the fields georeferenceProtocol\_\_@DwC\_\_@vegpath.org, & geovalid\_bien were blank for all records.

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.

There were some fields where I did not recognize the names, but they appeared blank. Not sure what to make of these: For example, what are occurrenceID\_\_@DwC\_\_@vegpath.org and recordedBy\_\_@DwC\_\_@vegpath.org

The morphospecies have non-alphabetical characters scrubbed out. For example Hypericum [graveolens + mitchellianum] is rendered as Hypericum [graveolens mitchellianum]

For every record I was sent identifiedBy\_\_@DwC\_\_@vegpath.org has the value Robert Peet. Most of these records I had nothing to do with. The person who collected the plot should be the default identifier.

I see lines duplicated for no obvious reason (eg 3952 & 3953)

Probably not a problem, but a rather large number of data fields have no home in the data extract I was given. Do were really want to discard all that site data, like soil data?

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 .1-1% cover. This needs to be indicated I some way

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. Unfortunatley, all are recorded as having the areas of the full plot, but a module is usually smaller than a plot , say .01 rathern 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.