**Data Management Breakout Session: March 18, 2020**

**Attendance:** Esteban Muldavin, Julie Evens, Michael Lee, Robert Peet, Carol Spurrier, Chris Lea, Dave Tart, Chelsea Fowler, Don Faber-Langendoen

**ACTION ITEMS;** **Potential Future Activity**

**VegBank Updates**

* Option one: Rebuilding VegBank
  + Bob and Matt Jones had done a full VegBank rebuild proposal a few years ago.
* Option two: Finding a global solution
  + This would be a big undertaking
* Option three: Shorter term incremental steps.
  + Michael just finished mapping out the information from community types and plant information.

**Pouring Data Into VegBank**

* Pouring data into VegBank is a big step forward.
* Workflow: authors send the data package to Michael Lee (excel sheet or other), Michael coverts it and enters it into the database.
* Data upload doesn’t work through the web and it cannot be automated.
* **Potential Future Activity**: Create a plot template email authors can send their data packages to, then Michael Lee can enter it into the USNVC.
  + The interest in uploading data isn’t overwhelming, workload would be manageable (about a half hour per data package)
* Retain the submit data page, but encourage use of the Vegetation Plot Template

**R Outputting**

* This is a separate process from the data upload function.
* Users would access the plot data in R.
* We would have to write an API that would wrap around the functions in VegBank. So, when users request information through R it would send them back what they need.
* This is a big on-taking and is not currently top priority. Users can get data out of VegBank right now via the web interface, getting data into VegBank is more pressing.

**NCEAS Update and the Future of VegBank**

* Are we living on the edge of software usability? This is old software, did NCEAS just buy us a little time or is this usable into the long-term?
  + Since NCEAS worked on it, it should be good.
* Should we plan on using this system for another year or two as is? Should we be preparing to make a switch?
  + **ACTION ITEM:** Ask Bryce or Matt what they think.
  + Michael Lee believes we have some time on this software.
* Things to think about: If we rebuild the system, we could drive away current users. New interfaces can frustrate current users if tools, buttons, etc. have moved and they feel getting familiar with the new interface is difficult.
* **ACTION ITEM:** **Michael and Bob** will ask for a budget to develop an API

**Linking VegBank plots that are assigned to a type to the type description**

* There’s a field to find all the plots assigned with that plot.
* **Future Discussion Topic**: What would it take to rebuild this? What are the limitations?
  + This system worked both ways, click on a plot and it links to NatureServe explorer.
  + Michael Lee reloaded the new data into VegBank but we haven’t been able to resolve the existing annotations to the update.
  + Currently, there are plots that are not linked to the new types, they’re still linked to the old types from older NVC versions. So far, we’ve been updating the types, but we’re curious if that’s the best way forward. What if someone doesn’t agree with the NVC change?
  + Another consideration: We go through and see all the codes that haven’t changed and delete the new codes that are the same. But due to a cascade effect, we have to be aware how changes influence other linkages.
  + The only way to correct these incorrect linkages or disagreement in assigned type is for users to go in and report a mistake or we have to spot it. This is the role of a database manager.
  + **ACTION ITEM:** Michael will write up a one pager so we can review and determine business rules
  + **ACTION ITEM**: Work out business rules. We can’t expect authors who enter plots to personally update their plots to the newest NVC version. We need rules for updates.

**Discussion on a Database Manager**

* Currently, we don’t have a data manager but Michael could run queries.
* Does VegBank need a data manager? Part-time? Full-time?
  + US Forest Service VegBank budget
    - Budgeted for getting plot data into VegBank, and some for VegBank functionality. There is no budget dedicated to data curation.
    - Michael’s priorities are getting data in, fixing data, and usability
    - Quality Control has not been vital, but it is within our capacity.
    - The curation aspect may be of limited value if we don’t have the capacity to do anything about it
    - **Future Discussion Topic**: Regional Editorial board could have an additional role. We could send out queries and ask for REs to fix the inconsistencies.

**New NatureServe Explorer and Veg Bank**

* In the new Explorer you can click and get to the plots in VegBank.
* **ACTION ITEM:** Test to see if the linkages from NVC to explorer still work post-NS update

**What role does VegBank have in the Communications Plan?**

* One of the Panel’s founding goals is to provide plot-based descriptions.
* Having a type description that points to plots is foundational to the Veg Panel’s mission and VegBank is the tool by which we achieve this goal.
  + We need to find a way to do what we do better.
* **ACTION ITEM:** Revise author guidelines
  + Proposals for NVC revisions need to account for existing VegBank plot data.

One pager from Michael. Progress with the current structure and see if we can plan out major changes if we need them and getting APIs out there in the near future. Some budgetary frame if possible.

**The Global Perspective**

* We want users to be able to access global data.
* Data in different places aren’t so different
* We have similar uses for classification for broad scientific questions.
* A global solution would be best when brainstorming on storage and use issues.
* Europe has been successful.
* Pouring data out of VegBank into a universal data repository.
  + How would we achieve this? Bob plans to discuss at IAVS
    - **Future Discussion Topic**: Results of this conversation
  + GBIF is the group closest to developing this global network. But vegetation data isn’t necessarily the mission. GBIF is more focused on distribution than a data management.
  + GBIF will only ever be able to handle a subset of the data that we have in VegBank.
  + If someone wanted to assess community types at some broad global level GBIF would be a helpful tool.

There’s no solution to synchronizing everything, there will always be noise in the system, but we need to work with states and other partners to keep VegBank active, adaptive, and functional.

**Next Steps**

* Goal: VegBank is operational 3-5 years from now and we keep exploring what it means to maintain it.
* We still need to think about and implement updates and improvements.
  + Bob and Matt Jones have a proposal that describes improvements
  + **ACTION ITEM: Bob** will distribute this proposal to the group
* Continue to explore an international database (strong international support)