

## BIEN3 Milestones and Development Timeline

Draft, 12 December 2011  
B. Boyle

The following is a list of the principal goals for the development of the BIEN3 database and website, roughly in order of development, followed by a proposed development timeline.

### Development milestones

1. Initial development of BIEN3 database and import utilities
  - a. Preliminary build of BIEN3 database (VegBIEN) based on Vegbank
  - b. Preliminary mapping script from VegX to VegBIEN
  - c. Additions to VegBIEN to support trait data
  - d. Develop trial loading and mapping scripts for source→VegX and VegX→VegBIEN using example data sources:
    - Specimens: NYBG (one specimen example should suffice for initial trial; DwC fairly consistent)
    - Plots (in increasing order of complexity: SALVIAS, TurboVeg, RAINFOR, NVS (New Zealand), CTFS, VegBank)
    - Traits
    - Full validation of all above mappings, working with someone familiar with each data source
2. Revise schemas I
  - a. Summarize shortcomings of VegX and VegBIEN schemas, and recommend changes. Issues to address include accurate modeling of individual observations (stems nested in individual trees), specimen vouchering of plot observations, morphospecies, metadata concerning ownership and access. Possibly also propose changes to DwC to support herbarium database fields such as "Cultivated".
  - b. Modify VegX and VegBIEN (Nick Spencer to implement changes to VegX and update TDWG)
3. Load data
  - a. Develop mappings (as needed) to all data originally in BIEN2, using complete original sources where possible, including all necessary metadata. For specimens, if get new dump, request 'Cultivated' field if available. Sources:
    - Specimens (DwC format): NYBG, ARIZ, CRIA (need to modify headings slightly), GBIF (request new dump), Canadian herbaria
    - Specimens not in DwC (request new dumps): REMIB (Can't get new dump, but Brad could modify original), Utrecht, NCU-NCSC, UNCC, MO (this is only a partial DwC dump; many fields missing, such as specimen description), GBIF
    - Plots: CTFS, CVS, FIA, Madidi, RAINFOR, SALVIAS, TEAM, VegBank

- b. Load all data to BIEN3
4. Revise schemas II
  - a. Make additional changes to VegX or VegBIEN based on issues discovered while loading data
5. Reload all data, using revised schemas and mappings
  - Load data
6. Develop validation utilities
  - Taxonomic validation
    - Automated taxon correction using TNRS web service
    - Validation scripts for double checking TNRS results as loaded to BIEN
  - Georeferencing
    - Develop new scripts that use existing web services to georeference additional records
  - Geovalidation
    - Replicate current geovalidations (BIEN2), increasing use of web services for reference data (BIEN2 version used numerous static tables of political division names, etc.)
  - Other validations
7. Analytical database
  - Construct variety of analytical views to be used for direct data download, as well as for querying data via web interface
  - Including but not limited to BIEN2's viewFullOccurrence and other summary tables developed by Brad
8. Data provider feedback
  - a. Scripts to assemble error reports as feedback to data providers. Should report on errors found during loading, taxonomic validation and geovalidation
9. Website
  - Develop website for serving raw data from BIEN database, as well as derived data products such as species maps. Will include following elements:
    - BIEN3 web database (extracted but separate from core and analytical databases developed above)
    - User database
    - Web front end, including different levels of read-only and write access, from general public to data owners and web administrators. Should support user-level management of data access and sharing, similar to current SALVIAS website.
    - Access to derived data products such as species range maps. Production of derived products is out-of-scope, but will still need to develop tools for accessing, storing and serving these products.
    - Web access to geovalidation tools (GNRS)
10. Acquire, map and load additional data

Martha Narro 12/15/11 2:38 PM

**Comment:** Do such georeferencing web services exist?

Martha Narro 12/15/11 2:38 PM

**Comment:** Based on John Donoghue's experience, currently there may not be a good service to use, so this could be a problematic dependency.

Martha Narro 12/15/11 2:38 PM

**Comment:** Will Aaron develop the website?

## **Additional goals**

The following goals would be important achievements, but may require more time, more personnel or additional financial support or outside collaboration.

1. Data-mapping tool
  - Desktop (or web?) tool for mapping new plot data to VegX
  - Would enable open-ended acquisition of new data by BIEN
  - Similar in concept to VegBranch
  - Possibly develop in cooperation with Landcare Research and the NVS?
2. Data entry tool
  - Would allow for controlled entry of plot data that would be both compatible with VegBIEN, and allow export as VegX
  - Support controlled taxonomy using TNRS web services
  - Other controlled vocabulary using BIEN web services
  - Possibly develop in cooperation or with support of Conservation International