**Rough outline of a White Paper on BIEN 3.0 (almost the iPlant proposal of Jan 2009)  
For the working Group -- 8 Dec 2010, 16:00.**

**R.K. Peet. Based on notes of B. Boyle from breakout of 5 Dec**

1. **Introduction** (Revised section 1 of iPlant proposal)
   1. Taxon occurrence data (plots and individuals) as a critical data type for future science
   2. Where we are: ~BIEN 2.5 -- a prototype public archive
   3. Where we want to go: BIEN 3.0 – Cyberinfrastructure needed to support public storage, submission, search, capture and annotation of occurrence data, initially populated with BIEN 2.5 data.
2. **Need and Application – Questions to address** (revised section 2 of iPlant proposal)
   1. What grows where and why?
   2. Current distributions by species
   3. Future distributions?
   4. Who grows with whom?
   5. Current distributions of traits (across space)
   6. Phylogenetic patterns
   7. Change in communities over time
3. **Potential data**
   1. 4,000,000 vegetation plots (2,300,000 currently digitized)
   2. ? high-quality occurrence records (eg Heritage program records)
   3. 70,000,000 specimens (12,000,000 currently digitized)
4. **Core Components**
   1. Overview (including Figure 1 from iPlant proposal)
   2. Internal components
      1. Individual taxon specimen observations (with specimen metadata)
      2. Other high-quality individual occurrence records (with obs metadata)
      3. Aggregate occurrence observations (with plot metadata)
      4. Measurements of organisms
      5. Repeat observations
      6. Standard taxonomy with synonomy
      7. Metadata regards ownership and use conditions (method, ownership, agency)
   3. Linkages to supporting tools and services
      1. Environmental layers
      2. Taxonomic services
         1. Taxonomic concept resolution
         2. Support for alternative taxonomic standards
         3. Taxon concept relationship mapper
         4. Taxon agglomeration tool
      3. Trait data
      4. Geospatial discovery tool
      5. Data discovery for equivalent data housed elsewhere?
      6. IPToL and other phylogeny resources
5. **Use cases illustrating how a scientist would use the system**
6. **Data flows and work flows**
   1. Search & download records
   2. Submit records
   3. Annotate records
   4. Scrub records
   5. Feed back to data providers
   6. Manage system
7. **Design elements**
   1. Permissions, data access control, embargoes
      1. Ownership
      2. Permissions for submission and management
      3. Capability for embargo and geographic fuzzing
      4. censored data: original data vs. censored at point of access
      5. ability to give back to data provider information about access, downloads
      6. login required for data access
      7. Identity access management (confirmation) framework?
   2. Data upload issues
      1. specimen data: Darwin Core, (and possibly implementations of Specify)
      2. Vegetation data: tool for mapping into VegX
      3. VegBranch-like mapping tool for small datasets
      4. Automated import via schema?
      5. User-control of data refresh?
      6. Who is allowed to contribute?
   3. Web interface
   4. Versioning
      1. Data as viewed on web published (posted and archived) at fixed intervals. This information recorded on website.
      2. User will cite the particular version of the database. Can verify and walk back to that version at any time
   5. Annotations
      1. Taxonomic determination history
      2. Corrections
      3. Record of history of corrections
8. **Architectural approach**
   1. Operational (transaction) DB + multiple analytical (derivative) DBs
   2. Website: temporary select-optimized tables
   3. Probably modified VegBank model & VegX exchange within new system
9. **Staging & implementation**
   1. Core elements needed quickly
   2. Prioritized nice-to-have list
   3. Who and where