NEON ALGORITHM THEORETICAL BASIS DOCUMENT (ATBD): TOS PLANT DIVERSITY - QA/QC OF RAW FIELD AND LAB DATA

PREPARED BY	ORGANIZATION	DATE
David Barnett	TOS	09/03/2014

APPROVALS	ORGANIZATION	APPROVAL DATE
Hanne Buur	PSE	02/27/2015

RELEASED BY	ORGANIZATION	RELEASE DATE
Jennifer DeNicholas	CM	03/03/2015

See configuration management system for approval history.

The National Ecological Observatory Network is a project solely funded by the National Science Foundation and managed under cooperative agreement by NEON, Inc. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Change Record

REVISION	DATE	ECO #	DESCRIPTION OF CHANGE
A	03/03/2015	ECO-02664	Initial release

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

1 DESCRIPTION

1.1 Purpose

This document details the algorithms used for creating a subset of NEON Level 1 data products that are the quality controlled products generated from raw Level 0 data, and associated metadata. In the NEON data products framework, the raw data collected in the field, for example, percent cover or a certain plant species, are considered the lowest level (Level 0). Raw data that have been quality checked via the algorithms detailed herein, as well as simple metrics that emerge from the raw data, such as total plant species richness at a particular site, are considered Level 1 data products. This document relates only to the former group of L1 data products, the quality controlled pass-through products from the Level 0 data products.

This document includes a detailed discussion of measurement theory and implementation, appropriate theoretical background, data product provenance, quality assurance and control methods used, approximations and/or assumptions made, and a detailed exposition of uncertainty resulting in a cumulative reported uncertainty for this product.

1.2 Scope

This document describes the theoretical background and entire algorithmic process for creating a subset of quality controlled and calibrated L1 data products and associated metadata from input data. These data products include such measurements as the presence and percent cover of plant species in NEON plots, at multiple spatial scales (Table 1). It does not provide computational implementation details, except for cases where these stem directly from algorithmic choices explained here. This document also provides details relevant to the publication of the data products via the NEON data portal (NEON Data Publication Workbook for TOS Plant Diversity: QA/QC of Raw Field and Lab Data (AD[12])).

This document describes the algorithms for ingesting and performing automated quality assurance and control procedures on the data collected in the field pertaining to the NEON Field and Lab Protocol for Plant Diversity (AD[10]). The raw data that are processed in this document are detailed in the NEON Raw Data Ingest Workbook for TOS Plant Diversity (AD[11]).

2 RELATED DOCUMENTS AND ACRONYMS

2.1 Applicable Documents

The applicable documents listed below include documents that are cited throughout the body of this text, and that contain provisions or other pertinent requirements that directly relate to, and are necessary for, the performance of the activities specified in this document.

AD[01]	NEON.DOC.000001	NEON Observatory Design (NOD) Requirements
AD[02]	NEON.DOC.005003	NEON Scientific Data Products Catalog
AD[03]	NEON.DOC.005004	NEON Level 1-3 Data Products Catalog
AD[04]	NEON.DOC.005005	NEON Level 0 Data Product Catalog
AD[05]	NEON.DOC.005011	NEON Coordinate Systems Specification
AD[06]	NEON.DOC.004309	NEON Field Site Information
AD[07]	NEON.DOC.002162	NEON ATBD QA/QC Taxonomic Identification of Organisms – in prep
AD[08]	NEON.DOC.000912	TOS Science Design for Plant Diversity
AD[09]	NEON.DOC.014042	NEON Field and Lab Protocol for Plant Diversity
AD[10]	NEON.DOC.001399	NEON Raw Data Ingest Workbook for TOS Plant Diversity
AD[11]	NEON.DOC.001410	NEON Data Publication Workbook for TOS Plant Diversity: QA/QC of Raw Field and Lab Data
AD[12]	NEON.DOC.004309	NEON Field Site Information
AD[13]	NEON.DOC.000261	TOS Spatial Data
AD[14]	NEON.DOC.002253	List of Valid Identification Qualifiers
AD[15]	NEON.DOC.002257	NEON Taxonomic Name List for Plants

2.2 Reference Documents

Reference documents contain information complementing, explaining, detailing, or otherwise supporting the information included in the current document.

RD[01]	NEON.DOC.000008	NEON Acronym List
RD [02]	NEON.DOC.000243	NEON Glossary of Terms

3 DATA PRODUCT DESCRIPTION

Plant diversity will be quantified during one to two sampling bouts per year, by NEON field technicians who collect plot-based measurements at core and relocatable sites. Sampling bout frequency will vary among sites, in accordance with site-specific intra-annual distributions of growing seasons. For an overview of the data collection strategy see: TOS Science Design for Plant Diversity (AD[08]) and NEON Field and Lab Protocol for Plant Diversity (AD[09]). Products resulting from plant sampling events include observed species presences at multiple spatial scales within each NEON plot (i.e., 400 m^2 and 100 m^2 subplots, and 10 m^2 and 1 m^2 nested subplots), and the percent cover of plant species in 1-m^2 nested subplots (Figure 1).

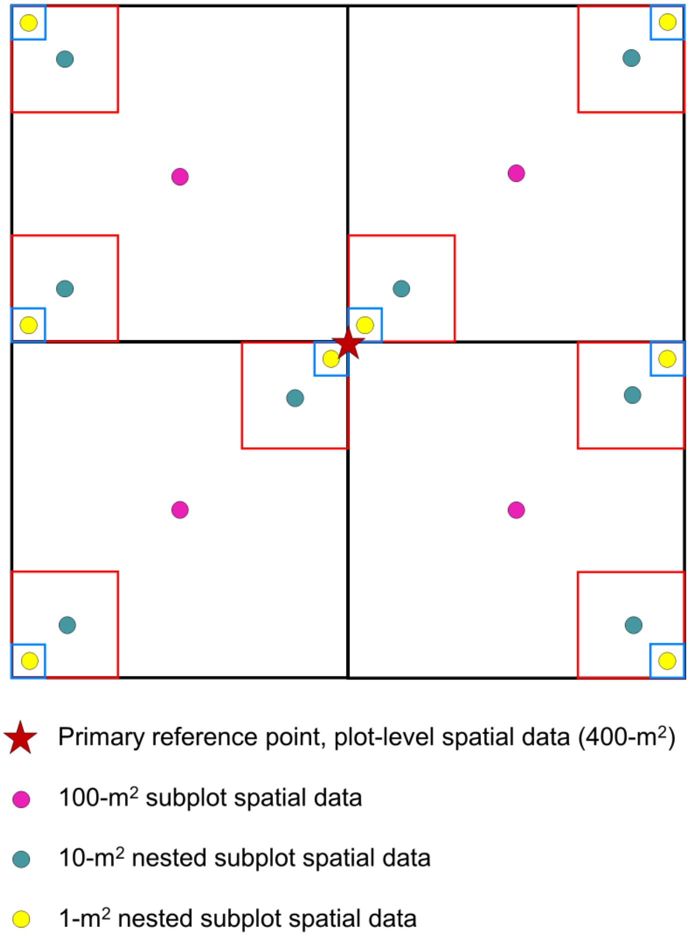

Figure 1. The plot design for observing plant diversity; the reference location of the spatial data information associated with the plant diversity data as populated by and found in the pointSpatialData table.

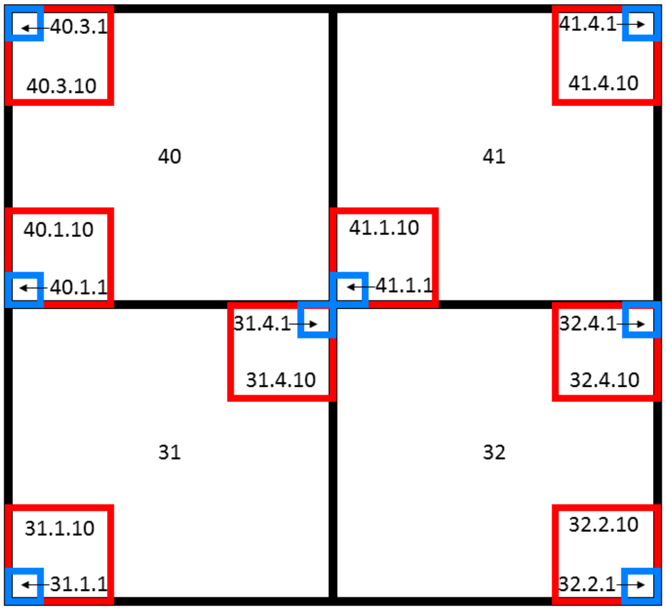

Figure 2. The plot design for observing plant diversity including subplot numbering system.

3.1 Variables Reported

This ATBD describes the steps needed to generate the L1 data product plant presence/absence and percent cover (NEON.DOM.SIT.DP1.10058). Subproducts for this data product are listed below (Table 1). Detailed lists of the associated subproducts and metadata products are provided separately, along with example data in publication-ready spreadsheets (NEON Data Publication Workbook for TOS Plant Diversity: QA/QC of Raw Field and Lab Data (AD[11])). Field names have been standardized with Darwin Core terms (http://rs.tdwg.org/dwc/; accessed 16 February 2014), the Global Biodiversity Information Facility vocabularies (http://rs.gbif.org/vocabulary/gbif/; accessed 16 February 2014), the VegCore data dictionary (https://projects.nceas.ucsb.edu/nceas/projects/bien/wiki/VegCore; accessed 16 February 2014), and with the Bird Monitoring Data Exchange standards (http://www.avianknowledge.net; accessed 16 February 2014), where applicable. Geospatial data shall conform to the standards set forth in the NEON Coordinate Systems Specification (AD[05]).

Table 1. List of subproducts produced in this ATBD for the data product, plant presence/absence and cover (NEON.DOM.SIT.DP1.10058). The list is not exhaustive and a variety of supporting data will also be made available.

Number	Field Name	Description
NEON.DOM.SITE.DP1.10058.001.00077.001.001.001	targetTaxaPresent	Indicator of whether the sample contained individuals of the

		target taxa
NEON.DOM.SITE.DP1.10058.001.00124.002.001.001	otherVariablesPresent	Indicator of whether the sample contained other variables, such as abiotic, biotic, and non-vascular plant cover
NEON.DOM.SITE.DP1.10058.001.00012.001.001.001	taxonID	Species code, based on one or more sources.
NEON.DOM.SITE.DP1.10058.001.00055.001.001.001	scientificName	Scientific name, associated with the taxonID. This is the name of the lowest level taxonomic rank that can be determined.
NEON.DOM.SITE.DP1.10058.001.00056.001.001.001	taxonRank	The lowest level taxonomic rank that can be determined for the individual or specimen.
NEON.DOM.SITE.DP1.10058.001.00013.001.001.001	identificationQualifier	A standard term to express the determiner's doubts about the Identification.
NEON.DOM.SITE.DP1.10058.001.00081.001.001.001	morphospeciesID	Identifier for morphospecies
NEON.DOM.SITE.DP1.10058.001.00108.001.001.001	percentCover	Ocular estimate of cover of the index (e.g., species) as a percent
NEON.DOM.SITE.DP1.10058.001.00127.001.001.001	heightPlantOver300cm	An indicator for whether individuals of the species in the sample are taller than 300 cm
NEON.DOM.SITE.DP1.10058.001.00128.001.001.001	heightPlantSpecies	Ocular estimate of the height of the plant species

3.2 Temporal Resolution and Extent

The finest temporal resolution at which plant diversity data will be tracked is day of sampling event. Sampling will occur in bouts comprised of sampling plots or a subset of plots at a site. A bout is expected to occur over a period of one to two months. In most cases, there will be only one bout per site per year. Sites with multiple peaks in phenology, periods of time when multiple species are in flower and easier to identify, may require more than one sampling bout per year.

3.3 Spatial Resolution and Extent

The finest spatial resolution at which plant presence and percent cover will be tracked is within 1-m^2 nested subplots. The species present, but not their percent cover, will also be recorded in 10-m^2 nested subplots and 100-m^2 subplots. Data from the four 100-m^2 subplots will be aggregated to describe the species present at the scale of each 400-m^2 plot. Plots are distributed across the study site in vegetation types that occupy more than 5% of the total site extent, and the number of plots at a site will vary with the size, heterogeneity, and relative cover of vegetation types at each site (see TOS Science Design for

Plant Diversity (AD[08]) for further details). Overall, plant diversity subproducts relate to one another via the following spatial hierarchy (with field names, data used to generate entries, and example entries in parentheses):

nested 1m^2 subplot ("**subplotID**"; subplot.corner.size, 31.1.1) → nested 10m^2 subplot ("**subplotID**"; subplot.corner.size, 31.1.10) → 100m^2 subplot ("**subplotID**"; subplot, 31) → 400-m^2 plot ("**plotID**"; plotID, CPER_001) → siteID (ID of NEON site) → domainID (ID of a NEON domain)

3.4 Associated Data Streams

Although NEON-collected remote sensing layers cover the sampling area, and other NEON data products are collected at the plots sampled for plant diversity, individual measurement streams within the plant diversity module are not otherwise linked to other NEON data products.

3.5 Product Instances

The raw field data generated from observing plant diversity and abundance (percent cover) are predominantly per species, per plot (per bout, per site). Plant diversity plots will be set at all core and relocatable sites throughout NEON every year. Sampling effort for plant diversity is typically measured in number of plots. The number of species per plot will vary across NEON sites, and, consequently, the number of records to be generated is not pre-determined.

4 SCIENTIFIC CONTEXT

4.1 Theory of Measurement/Observation

Observations of plant diversity have played a central role in the development of ecological theory and practice. Darwin documented the distribution of plant species assemblages in his backyard (Magurran and McGill 2010) prior to defining theories that described species interactions and species-environment relationships, during subsequent global exploration (Darwin 1859). The science has evolved, furthering understanding of the processes - mutation, drift, selection, dispersal, speciation, and extinction - that govern the interactions of species and species-environment relationships (Vellend 2010). Measures of plant diversity now play a central role in several branches of ecological research. For example, research aimed at understanding the causes of plant species distributions, fecundity, and persistence dominates population ecology (Clark et al. 2004). Community ecology studies commonly focus on the interactions of two or more plant species, and the resulting impact of these interactions on species composition in time and space. Other approaches to studying plant diversity focus on the importance of regional species pools, and the relationship between environmental factors and the distributions, occurrences, and abundances of species (Stohlgren 2007).

Plant diversity is sensitive to changes in climate (Ibanez et al. 2006, Magurran and Dornelas 2010), species invasion (Crall et al. 2006, Barnett et al. 2007), land use change, and disturbance (Dornelas 2010). Paleoecological records demonstrate the influence of shifting climate on species distributions (Wagner and Lyons 2010). Since natural selection is influenced by natural and anthropogenically-induced climate change, species not suited to emerging conditions will be forced to adapt or track change through a combination of dispersal and adaptation to novel conditions and interactions (Clark et al. 2012a). Even without directional changes in climate, plant species composition and diversity will change as species invade native ecosystems and alter resource availability, species interactions, and disturbance regimes. Land use may drive the most pronounced changes. Disturbance to the structure of soil and species, changing disturbance regimes, and inputs to systems have direct and indirect impacts on plant diversity (Pickett and White 1985, Pickett et al. 1989). Collectively, many factors influence the direction and magnitude of changes in plant diversity, species composition and abundance, and the distribution and interactions of other species in a complex environment.

Changes in patterns of plant diversity and associated processes are best quantified through repeated observations through time and space. However, there is little in the way of a standardized protocol for measuring plant diversity (Stohlgren 2007). The development of such a standardized method for sampling at multiple spatial and temporal scales would: allow for an improved understanding of processes driving patterns of diversity at various spatial extents, facilitate an understanding of local and regional patterns of diversity, and provide data to validate airborne observations at multiple grain sizes. One goal at NEON is to develop standardized plant diversity sampling methods, including sampling plant diversity within consistently sized plots at all sites within the NEON domains. This will allow for the comparison of plant diversity across sites, which is crucial for furthering understanding of patterns and processes at coarse spatial scales. A multi-scale approach, documenting plant species richness at a variety of nested spatial scales, allows for a consistent and comparable plot sizes that are also capable of characterizing patterns of diversity in grasslands and large-stature, well-spaced tree forests.

NEON will sample plant species with a multi-scale plot design developed by the Carolina Vegetation Project. NEON will sample a 20 x 20 m square plot comprised of four 100-m^2 subplots. Additionally, these subplots will contain 10-m^2 and 1-m^2 nested subplots. Information specific to NEON data products is collected as follows:

- Vascular plant taxonomic composition, i.e., the species present, including woody and herbaceous species but not non-vascular species, will be recorded in each subplot (100-m^2) and each nested subplot (1-m^2 and 10-m^2). Plants will be identified to the species level wherever possible; however in some cases (such as when a plant lacks reproductive structures at the time of survey) only a designation to a higher level of taxonomic resolution is possible.
- Estimates of abundance of herbaceous plants, made through ocular visual estimation of percent cover per taxon, will be made within the 1-m^2 subplots.

- Estimates of the total percent cover of nonvascular plant species, as well as aerial cover of substrate other than living herbaceous vascular plant tissue (e.g. litter, standingDead, moss, wood) will be made within the 1-m^2 subplots.

4.2 Theory of Algorithm

This document describes the algorithms for assessing the integrity of the L0 data stream generated by the field sampling of plant diversity. The conversion from the L0 to L1 data products requires simple, automated quality control and quality assurance procedures, including: the verification that all required data are recorded for each sampling event, comparing data values to pre-defined (through provided validation rules or specified lookup tables) ranges, and tracking the individual-level data for consistency and accuracy through time. Additionally, data products representing coarser scales than those directly sampled need to be generated through the concatenation of species lists obtained during fine resolution sampling.

4.2.1 Summary of the Algorithm

1. Create multi-scale temporary tables from ingest, splitting plant and other variables cover into separate tables for publication by size of plot and type of data (taxonomic or other cover)
2. Generate eventID (format DIV.SITE.YEAR.BOUTNUMBER)
3. Check all data for duplicates. Exact duplicates are removed algorithmically but flagged with '1' to indicate duplicates that were resolved; remaining records are flagged as potential duplicates (flag value ='2') when the same taxonID and/or {morphospeciesID,measuredBy combinations} are recorded more than once within the same subplot during a single sampling bout.
4. Check all data for missing records. During each bout, the protocol specifies that for each sampled plot, technicians should survey each of the 8 nested 1m^2 subplots enumerated in Figure 2, each of the 8 nested 10m^2 subplots enumerated in Figure 2, and each of the 4 nested subplots enumerated in Figure 2. Missing record flag values indicate the total number of missing subplots encountered at each spatial scale, for a maximum of 4 missing 100m2 subplots and 8 missing subplots at 1m^2 and 10m^2 scales. Missing record flags are only applied where technicians failed to **enter** data for a subplot; if a technician records the data in a plot as 'targetTaxaPresent =N' (i.e. where a subplot was surveyed but contains no vascular plants [or potentially no additional vascular plants, in the case of the larger subplots] , this does not generate a missing record flag.
5. Provide scientific names associated with each taxon code (taxonID). Technicians may enter either the current (valid)taxon code for each taxon identified, or a synonym. Synonymies are resolved, and the accepted taxon code, and associated scientific name and taxonomic rank are reported in the L1 data.

6. Create full taxon lists at each spatial scale by concatenating lists from nested subplots. For example, the plant taxon list for subplot 31 will contain all taxa (scientificNames and {morphospeciesID,measuredBy combinations} within subplot 31, as well as any additional taxa recorded within subplots 31.4.1, 31.4.10, 31.1.10, and 31.1.1.
7. Remove taxa of conservation concern from the L1 data prior to publication
8. Generate spatial data
9. Process 400-m^2 data by concatenating unique taxa (defining uniqueness based on taxonID and/or {morphospeciesID,measuredBy combinations} from the four 100-m^2 subplots

5 ALGORITHM IMPLEMENTATION

Throughout the algorithm implementation section of this ATBD, 'nodata', 'null', and/or 'NA' indicates a blank cell. All variables reported from the field or laboratory technician (L0 data) are listed in the data ingest workbook (AD[10]), notated here as div_tablename_in. Unless otherwise specified in the algorithm below, all variables that appear in tables div_tablename_pub (L1 data) have been passed directly from the L0 variables with the same name, as listed in the data publication workbook (AD[11]). Algorithm implementation should proceed in the order of the subsections provided here, e.g., the processing of data in subsection 5.1 should occur prior to that of data in subsection 5.2, and so on.

Automated Processing Steps for Validation of Field Collected Data

5.1 Run the following processing steps for all data

1. Tables to be ingested: *div_1m2Data_in, div_10m2Data100m2Data_in*
2. Temporary tables created: *div_1m2Data_temp, div_1m2Variables_temp, div_10m2Data_temp, div_100m2Data_temp, div_400m2Data_temp*
3. Tables to be published: *div_1m2Data_pub, div_1m2Variables_pub, div_10m2Data_pub, div_100m2Data_pub*
4. Generate **eventID**
    a. For each record, concatenate: 'DIV' + value in **siteID** field + Year component of **date** + value in **boutNumber** field, separated by '.'
        i. Example: DIV.OSBS.2014.1
    b. Insert string into **eventID** field of *div_1m2Data_in, div_10m2Data100m2Data_in*

5.2 Separate ingest data into temporary tables for validation prior to passing data to publication tables

1. Check that sampling bouts were sufficiently documented at the **1m2** scale, prior to creating temporary tables
    a. Create quality flag field in *div_1m2Data_in*, and populate with zeroes
        i. QF Name = **missingRecords1m2QF**
    b. For each {**plotID**, **eventID**} combination that occurs in either ingest table (*div_1m2Data_in, div_10m2Data100m2Data_in)*, verify that there is at least one

record with of each of the following values in the **subplotID** field: '31.1.1', '31.4.1', '32.2.1', '32.4.1', '40.1.1', '40.3.1', '41.1.1', '41.4.1'. If any of these subplotIDs are missing:

i. Create *n* new records in *div_1m2Data_in*, where *n* = number of missing subplotIDs from list in (b), and assign the value(s) of the missing subplotID(s) from list (b) into the **subplotID** field(s) of the new record(s)
ii. assign the values for **eventID**, **domainID**, **siteID**, **plotID** and **boutNumber** accordingly, using information in the given {**plotID, eventID**} combination
iii. assign to the **date** field: the last value in the sorted list of dates (from non-missing records) corresponding to the given {**plotID, eventID**} combination
iv. insert ‘1’ into the **missingRecords1m2QF** field of all new records for the given {**plotID, eventID**} combination
v. generate new uids for the generated records and populate remarks with ‘This record is generated by the system’

2. Create *div_1m2Data_temp* from fields in *div_1m2Data_in*
    a. When **missingRecords1m2QF** = 1 in *div_1m2Data_in*, copy all fields from *div_1m2Data_in* to *div_1m2Data_temp* except **otherVariablesPresent**, **otherVariables,** and **enteredBy**
    b. When **divDataType** = ‘plantSpecies’ in *div_1m2Data_in,* copy all fields from *div_1m2Data_in* except **otherVariablesPresent**, **otherVariables,** and **enteredBy** to *div_1m2Data_temp*
3. Create *div_1m2Variables_temp* from fields in *div_1m2Data_in*
    a. When **divDataType** = ‘otherVariables’ in *div_1m2Data_in,* copy all fields from *div_1m2Data_in* except **targetTaxaPresent, taxonID, acceptedTaxonID, identificationQualifier, taxonIDRemarks, morphospeciesID, heightPlantOver300cm, heightPlantSpecies, identificationReferences, and enteredBy** to *div_1m2Variables_temp*
4. Create *div_10m2Data_temp* from fields in *div_10m2Data100m2Data_in*
    a. When **subplotID** = (31.1.10, 31.4.10, 32.2.10, 32.4.10, 40.1.10, 40.3.10, 41.1.10, OR 41.4.10) in *div_10m2Data100m2Data_in*, copy all fields except **enteredBy** from *div_10m2Data100m2Data_in* to *div_10m2Data_temp*
5. Create *div_100m2Data_temp* from fields in *div_10m2Data100m2Data_in*
    a. When **subplotID** = (31, 32, 40, OR 41) in *div_10m2Data100m2Data_in*, copy all fields except **enteredBy** from *div_10m2Data100m2Data_in* to *div_100m2Data_temp*

5.3 Run the following processing steps for *div_1m2Data_temp*

1. Check for duplicate records - I

  - a. Create quality flag field and populate with zeroes
    - i. QF Name = **duplicateTaxonIDBoutPer1m2QF**
  - b. Check for duplicate records, based on exact matches all fields except **uid**, treating NULL values as 'matches' to other NULL values.
  - c. If exact duplicates are found
      - A. Pass only one of the records into *div_1m2Data_temp*
      - B. Enter a '1' into the corresponding cell in **duplicateTaxonIDBoutPer1m2QF**
    - i. Else:
      - A. Pass both records into *div_1m2Data_temp*
      - B. Enter a '1' into the corresponding cell in **duplicateTaxonIDBoutPer1m2QF**

2. Check for duplicate records – II – flag any remaining duplicate instances of **acceptedTaxonID** or {**morphospeciesID**,**measuredBy**} combinations within the same {**eventID, plotID, subplotID**}.
   - a. If morphospeciesID is NULL: List of fieldNames = [**eventID, plotID, subplotID, acceptedTaxonID]**
     - i. Check for duplicate records, based on exact matches in the fields listed in (a), excluding records with NULL values in any of the fields
     - ii. If duplicate records are identified:
       - A. Retain both records in *div_1m2Data_temp*
       - B. Enter a '2' into the corresponding cells in **duplicateTaxonIDBoutPer1m2QF**
   - b. If morphospecies ID is not NULL: List of fieldNames = [**eventID, plotID, subplotID, morphospeciesID, measuredBy]**
     - i. Check for duplicate records, based on exact matches in the fields listed in (b), excluding records with NULL values in any of the fields
     - ii. If duplicate records are identified:
       - A. Retain both records in *div_1m2Data_temp*
       - B. Enter a '2' into the corresponding cells in **duplicateTaxonIDBoutPer1m2QF**
   - c. Set **duplicateTaxonIDBoutPer1m2QF to** 0 for all records where **missingRecords1m2QF** =1
3. Generate spatial location and uncertainty information
   - a. Insert corresponding values of **decimalLatitude, decimalLongitude, geodeticDatum, coordinateUncertainty** values from the TOS point-level spatial data lookup table (AD[13]) into fields of the same name in *div_1m2Data_temp* by matching each {**plotID**, **subplotID**} combination in *div_1m2Data_temp* with a {**plotID, pointID**} combination in AD[13] (and where AD[13]:**applicableModule** contains 'div').
4. Assign taxonomic information
   - a. If *div_1m2Data_temp*:**acceptedTaxonID** is not NULL, locate the value in the **taxonID** field of the NEON taxonomic name list for plants (AD[15]) and assign:
     - i. The value in the **taxonID** field of AD[15] to the **taxonID** field *div_1m2Data_temp* (i.e., replacing the L0 taxonID)

            ii. Corresponding values of **scientificName** and **taxonRank** from AD[15] to the same fields in *div_1m2Data_temp*
        b. Else:
            i. Assign NULL to the **taxonID**, **scientificName** and **taxonRank** fields of *div_1m2Data_temp*

5. Do not pass records of species of conservation concern to the L1:
    a. For each record in *div_1m2Data _temp*:
        A. Use *div_1m2Data _temp*.**taxonID** and *div_1m2Data _temp*.**domainID** to lookup the dXXNativeStatus for the relevant domain (e.g., use AD[15].d10NativeStatus where *div_1m2Data _temp*.**domainID**=D10) where AD[15].taxonID= *div_1m2Data _temp*.**taxonID**.
        B. If dXXNativeStatus = 'STATE ' or 'FEDERAL':
            A. Remove record from *div_1m2Data _temp*
6. Create **morphospeciesIDRemarks** field, and set to NULL. Data portal functionality for entering extensive comments on morphospecies has not yet been implemented
7. Pass L0 data from *div_1m2Data _temp* to *div_ 1m2Data _pub*

5.4 Run the following processing steps for *div_1m2Variables_temp*

1. Check for duplicate records -1
    a. Create quality flag field and populate with zeroes
        i. QF Name = **duplicateVariablesBout1m2QF**
    b. Check for duplicate records, based on exact matches all fields except **uid and enteredBy**, treating NULL values as 'matches' to other NULL values.
    c. If exact duplicates are found
        i. Pass only one of the records into *div_1m2Variables_temp*
        ii. Enter a '1' into the corresponding cell in **duplicateVariablesBout1m2Q**
2. Check for duplicate records -2
    a. List of fieldNames = [**eventID, subplotID, plotID, otherVariables]**
    b. Check for duplicate records, based on exact matches in the fields listed in (a), treating NULL values as 'matches' to other NULL values.in **otherVariables** as 'matches' for evaluating duplicates in those fields.
    c. If duplicate records are identified:
            A. Pass both records into *div_1m2Variables_temp*
            B. Enter a '2' into the corresponding cell in **duplicateVariablesBout1m2QF**
2. Generate spatial location and uncertainty information
    a. Insert corresponding values of **decimalLatitude, decimalLongitude, geodeticDatum, coordinateUncertainty,** values from the TOS point-level spatial data lookup table (AD[13]) into fields of the same name in *div_1m2Variables_temp* by matching each

{**plotID**, **subplotID**} combination in *div_1m2Variables_temp* with a {**plotID, pointID**} **combination** in AD[13] (and where AD[13]:**applicableModule** = 'div').

3. Pass L0 data from *div_1m2Variables_temp* to *div_1m2Variables_pub*

5.5 Run the following processing steps for *div_10m2Data_temp*

1. Check for duplicate records. Repeat check for duplicates – I from section 5.3.1 (above) on *div_10m2Data_temp,* to generate **duplicateTaxonIDBoutPer10m2QF**
2. If taxa are recorded in 1-m^2 nested subplots, but not in <u>corresponding</u> 10-m^2 nested subplots, add records from *div_1m2Data_pub* to *div_10m2Data_temp*
    a. For each record in *div_1m2Data_pub* and each record in *div_10m2Data_temp*:
        i. Define taxMorphoMeas_temp according to the following rules:
            A. If morphospeciesID is NULL: taxMorphoMeas_temp=**taxonID**
            B. Else: taxMorphoMeas_temp=concatenate (**morphospeciesID. measuredBy)**
    b. For each record in *div_1m2Data_pub* (i)
        i. Determine 10m2subplot(i) that corresponds to the *div_1m2Data_pub* (i) record , as the value in the subplotID (i) field of the record from *div_1m2Data_ temp* + '0'
            A. Example : subplotID 31.1.1 from *div_1m2Data_pub* becomes 31.1.10
        ii. Determine the set (j) of taxMorphoMeas_temp measured in the 10m2subplot, as the unique set of *div_10m2Data_temp*:taxMorphoMeas_temp where *div_10m2Data_temp*:plotID= *div_1m2Data_pub:*plotID (i) and *div_10m2Data_temp:*eventID= *div_1m2Data_pub:*eventID(i) and *div_10m2Data_temp:*subplotID= *div_1m2Data_pub:*10m2subplot(i)
        iii. If *div_1m2Data_pub:* taxMorphoMeas_temp is NOT an element of (j)
            A. Create a new record in *div_10m2Data_temp*
            B. Insert into the **subplotID** field of the new record: 10m2subplot from step b.i
            C. Fill data in the **taxonID, scientificName, taxonRank, identificationQualifier, taxonIDRemarks, morphospeciesID, morphospeciesIDRemarks, measuredBy, recordedBy** and **date** fields of the new record with values from the same fields in the *div_1m2Data_pub(i)*
            D. Generate new **uids** for these records

3. Check that sampling bouts were sufficiently documented at the **10m2** scale, prior to creating temporary tables
    a. Create quality flag field and populate with zeroes

            i. QF Name = **missingRecords10m2QF**

    b. For each {**plotID**, **eventID**} combination that occurs in either ingest table (*div_1m2Data_in, div_10m2Data100m2Data_in)*, verify that there is at least one record with of each of the following in the **subplotID** field the raw ingest table *div_10m2Data100m2Data_in*: '31.1.10', '31.4.10', '32.2.10', '32.4.10', '40.1.10', '40.3.10', '41.1.10', '41.4.10'. If any of these subplotIDs are missing (indicating the 10m2 subplot was not sampled:
        i. Create *n* new records in *div_1m2Data_in*, where *n* = number of missing subplotIDs from list in (b), and assign the value(s) of the missing subplotID(s) from list (b) into the **subplotID** field(s) of the new record(s)
        ii. assign the values for **eventID**, **domainID**, **siteID**, **plotID** and **boutNumber** accordingly, using information in the given {**plotID, eventID**} combination
        iii. assign to the **date** field: the last value in the sorted list of dates (from non-missing records) corresponding to the given {**plotID, eventID**} combination
        iv. insert '1' into the **missingRecords10m2QF** field of all new records for the given {**plotID, eventID**} combination
        v. generate new uids for the generated records and populate remarks with 'This record is generated by the system'

4. Copy the **missingRecords1m2QF** from *div_1m2Data_pub* into *div_10m2Data_temp*
    a. For each record in which the value in the **missingRecords1m2QF** field of *div_1m2data_pub* is > 0:
        i. Append a '0' to the end of the value in the **subplotID** field to get the corresponding 10m2 **subplotID**
            A. Example: **subplotID** 31.1.1 from *div_1m2Data_pub* corresponds to 10m2 **subplotID** 31.1.10
        ii. Find the corresponding {**plotID,eventID**, **10m2 subplotID** combination, and set missingRecords1m2QF=1.
        iii. Assign **eventID**, **domainID**, **siteID**, **plotID, subplotID,** and **boutNumber** to this record accordingly, using information in the given {**plotID, eventID**} combination
        iv. Assign to the **date** field: the latest date in the list of dates (from non-missing records) corresponding to the given {**plotID, eventID**} combination
        vi. Generate a new uid for each newly created record and populate remarks with 'This record is generated by the system'

5. Assign taxonomic information

    a. If *div_10m2Data_temp*:**acceptedTaxonID** is not NULL, locate the value in the **taxonID** field of the NEON taxonomic name list for plants (AD[15]) and assign:

            i. The value in the **taxonID** field of AD[15] to the **taxonID** field *div_10m2Data_temp*
            ii. Corresponding values of **scientificName** and **taxonRank** from AD[15] to the same fields in *div_10m2Data_temp*
        b. Else:
            i. Assign NULL to the **taxonID**, **scientificName** and **taxonRank** fields of *div_10m2Data_temp*
    vii. Check for duplicate records – II – flag any remaining duplicate instances of **taxonID** and {**morphospeciesID**,**measuredBy**} combinations within the same {**eventID, plotID, subplotID**}. Repeat logic from section 5.3.2 above, setting **duplicateTaxonIDBoutPer10m2QF** to '2' for records with matching **taxonID** or {**morphospeciesID, measuredBy**} within the same 10m2 subplot.

6. Generate spatial location and uncertainty information
    a. Insert corresponding values of **decimalLatitude, decimalLongitude, geodeticDatum, coordinateUncertainty** values from the TOS point-level spatial data lookup table (AD[13]) into fields of the same name in *div_10m2Data_temp* by matching each {**plotID**, **subplotID**} combination in *div_10m2Data_temp* with a {**plotID, pointID}** combination in AD[13] (and where AD[13]:**applicableModule =** 'div').

7. Do not pass records of species of conservation concern to the L1:
    a. For each record in *div_10m2Data _temp*:
        i. Use *div_10m2Data _temp*.**taxonID** and *div_10m2Data _temp*.**domainID** to lookup the dXXNativeStatus for the relevant domain (e.g., use AD[15].d10NativeStatus where *div_10m2Data _temp*.**domainID**=D10) where AD[15].taxonID= *div_10m2Data _temp*.**taxonID**.
        ii. If dXXNativeStatus = 'STATE ' or 'FEDERAL':
            1. Remove record from *div_10m2Data _temp*
8. Create **morphospeciesIDRemarks** field, and set to NULL
9. Pass L0 data from *div_10m2Data _temp* to *div_ 10m2Data _pub*

5.6 Run the following processing steps for *div_100m2Data_temp*

1. Check for duplicate records. Repeat check for duplicates – I from section 5.3.1 (above) on *div_100m2Data_temp,* to generate **duplicateTaxonIDBoutPer100m2QF**
2. If taxa are recorded in 10-m^2 nested subplots, but not in <u>corresponding</u> 100-m^2 nested subplots, add records from *div_10m2Data_pub* to *div_100m2Data_temp*
    a. For each record in *div_100m2Data_temp*:
        i. Define taxMorphoMeas_temp according to the following rules:
            A. If morphospeciesID is NULL: taxMorphoMeas_temp=**taxonID**

            B. Else: taxMorphoMeas_temp=concatenate (**morphospeciesID. measuredBy)**
    b. For each record in *div_10m2Data_pub* (i)
        i. Determine 100m2subplot(i) that corresponds to the *div_1m2Data_pub* (i) , as the first two characters from the *div_10m2Data_ temp:***subplotID**
        ii. Example : **subplotID** '31.1.10' from *div_10m2Data_pub* becomes 100m2-subplot '31'
            A. Determine the set (j) of taxMorphoMeas_temp measured in the 100m2subplot, as the unique set of *div_100m2Data_temp*:taxMorphoMeas_temp where *div_100m2Data_temp*:plotID= *div_10m2Data_pub:*plotID (i) and *div_100m2Data_temp*:eventID= *div_10m2Data_pub*:eventID(i) and *div_100m2Data_temp*:subplotID= *div_10m2Data_pub*:10m2subplot(i)
            B. If *div_10m2Data_pub* taxMorphoMeas_temp is NOT an element of (j)
                1. Create a new record in *div_100m2Data_temp*
                2. Insert into the **subplotID** field of the new record: *div_10m2Data_pub:*100m2subplot(i)
                3. Fill data in the **taxonID, scientificName, taxonRank, identificationQualifier, taxonIDRemarks, morphospeciesID, morphospeciesIDRemarks, measuredBy, recordedBy** fields of the new record with values from the same fields in the *div_10m2Data_temp(i)*
                4. Set the **missingRecords100m2Q**F for these new records to 0
                5. Generate new **uids** for these records
                6. Algorithm should be completed sequentially by such that taxMorphoMeas_temp recorded in more than one 10m2 nested subplot generate only ONE additional record within a 100m2 plot.

3. Check that sampling bouts were sufficiently documented
    a. Create quality flag field and populate with zeroes
        i. QF Name = **missingRecords100m2QF**
    b. For each {**plotID**, **eventID**} combination, verify that there is one of each of the following values in the **subplotID** field of field the raw ingest table *div_10m2Data100m2Data_in*: '31', '32', '40', '41'. If any of these subplotIDs are missing:

      i. Create *n* new records in *div_100m2Data_temp*, where *n* = number of missing subplotIDs from list in (b) and assign the value(s) of the missing subplotID(s) from list (b) into the **subplotID** field(s) of the new record(s)
      ii. assign the values for **eventID**, **domainID**, **siteID**, **plotID** and **boutNumber** accordingly, using information in the given {**plotID, eventID**} combination
      iii. assign to the **date** field: the last value in the list of dates (from non-missing records) corresponding to the given {**plotID, eventID**} combination
      iv. insert '1' into the **missingRecords100m2QF** field of all new records for the given {**plotID, eventID**} combination
      v. generate a new uid for the generated record and populate remarks with 'This record generated by the system'

4. Copy **missingRecords1m2QF** and **missingRecords10m2QF** from *div_10m2Data_pub* and *div_1m2Data_pub* into *div_100m2Data_temp*
   a. For each combination of {**plotID**, **subplotID**, **eventID**} in *div_100m2Data_pub*:
      i. Set missingRecords10m2QF as the count the total number of unique nested subplotIDs where **missingRecords10m2QF** =1 in *div_10m2Data_pub*
      ii. Set missingRecords1m2QF as the count the total number of unique nested subplotIDs where **missingRecords1m2QF**=1 in *div_1m2Data_pub*
         A. Example: For CPER_001, subplot 31, count **missingRecords10m2QF**=1 from 10m2 subplots 31.1.10 and 31.4.10, and count **missingRecords1m2QF**=1 from subplots 31.1.1 and 31.4.1
      iii. Find the corresponding {**plotID,eventID**, 100m2 **subplotID** combination, and insert flags from step 4.a.i and 4.a.ii
      iv. Assign **eventID, domainID, siteID, plotID,** subplotID, and **boutNumber** to this record accordingly, using information in the given {**plotID, eventID**} combination
      v. Assign to the **date** field: the latest date in the list of dates (from non-missing records) corresponding to the given {**plotID, eventID**} combination
      vi. Generate a new uid for each newly created record and populate remarks with 'This record is generated by the system'

5. Assign taxonomic information
   a. If *div_100m2Data_ temp*:**acceptedTaxonID** is not NULL, locate the value in the **taxonID** field of the NEON taxonomic name list for plants (AD[15]) and assign:
      i. The value in the **taxonID** field of AD[15] to the **taxonID** field *div_100m2Data_temp*
      ii. Corresponding values of **scientificName** and **taxonRank** from AD[15] to the same fields in *div_100m2Data_temp*
   b. Else:
      i. Assign NULL to the **taxonID**, **scientificName** and **taxonRank** fields of *div_100m2Data_temp*

6. Check for duplicate records – II – flag any remaining duplicate instances of **taxonID** and {**morphospeciesID**,**measuredBy**} combinations within the same {**eventID, plotID, subplotID**}. Repeat logic from section 5.3.2 above, setting **duplicateTaxonIDBoutPer100m2QF** to '2' for records with matching **taxonID** or {**morphospeciesID, measuredBy**} within the same 100-m2 subplot.

7. Generate spatial location and uncertainty information
    a. Insert corresponding values of **decimalLatitude, decimalLongitude, geodeticDatum, coordinateUncertainty** values from the TOS point-level spatial data lookup table (AD[13]) into fields of the same name in *div_100m2Data_temp* by matching each {**plotID**, **subplotID**} combination in *div_100m2Data_temp* with a **{plotID, pointID}** combination in AD[13] (and where AD[13]:**applicableModule =** 'div').

8. Do not pass records of species of conservation concern to the L1:
    a. For each record in *div_100m2Data _temp*:
        i. Use *div_100m2Data _temp*.**taxonID** and *div_100m2Data _temp*.**domainID** to lookup the dXXNativeStatus for the relevant domain (e.g., use AD[15].d10NativeStatus where *div_100m2Data _temp*.**domainID**=D10) where AD[15].taxonID= *div_100m2Data _temp*.**taxonID**.
        ii. If dXXNativeStatus = 'STATE ' or 'FEDERAL':
            A. Remove record from *div_100m2Data _temp*
9. Create **morphospeciesIDRemarks** field and set to Null
10. Pass L0 data from *div_100m2Data _temp* to *div_ 100m2Data _pub*

5.7 Run the following processing steps to create *div_400m2Data_pub*

1. Populate *div_400m2Data_temp* with records from all subplots and nested subplots in a given site
    a. For each {**plotID, eventID**} combination in *div_100m2Data_pub*, copy all records to *div_400m2Data_temp* where all values in missing records flags are '0' and **targetTaxaPresent** = 'Y', keeping **eventID, domainID, siteID, plotID, boutNumber, date, targetTaxaPresent, taxonID, scientificName, taxonRank, identificationQualifier, taxonIDRemarks, morphospeciesID, morphospeciesIDRemarks, samplingProtocol, identificationReferences, remarks, and measuredBy**
2. Check for duplicate records and remove remaining duplicate instances of **taxonID** or {**morphospeciesID**,**measuredBy**} combinations within the same {**eventID, plotID**}.
    a. Define taxMorphoMeas_temp according to the following rules:
        i. If morphospeciesID is NULL: taxMorphoMeas_temp=**taxonID**

        ii. Else: taxMorphoMeas_temp=concatenate (**taxonID, morphospeciesID. measuredBy)**
    b. Remove all but one record from *div_400m2Data_temp* for each duplicate set of taxMorphoMeas_temp for a given {**eventID**, **plotID**} combination
    c. Generate a **uid** for the retained records

1. Calculate the total number of missing records at the 1, 10, and 100m2 scales contained within each 400m^2 plot.
    a. Generate *div_400m2Data_temp:***missingRecords1m2QF**, *div_400m2Data_temp:***missingRecords10m2QF**, and *div_400m2Data_temp:***missingRecords100m2QF** for all existing records in *div_400m2Data_temp* and populate with '0'
    b. For each unique {**plotID**, and **eventID**} combination {i,j} in *div_400m2Data_temp*
        i. **missingRecords1m2QF**=count of unique (*div_1m2Data_temp:***subplotID**)) where *div_1m2Data_temp:***plotID**=*i and div_1m2Data_temp:***eventID**=*j* and *div_1m2Data_temp:***missingRecords1m2QF**=1
        ii. **missingRecords10m2QF**=count of unique (*div_10m2Data_temp*:**subplotID**) where *div_10m2Data_temp:***plotID**=*i and div_10m2Data_temp:***eventID**=*j* and *div_10m2Data_temp:***missingRecords10m2QF**=1
        iii. **missingRecords100m2QF**=count of unique (*div_100m2Data_temp:***subplotID**) where *div_100m2Data_temp:***plotID**=*i and div_100m2Data_temp:***eventID**=*j* and div_100m2Data_temp:**missingRecords100m2QF**=1
    c. For each combination {i,j}, if any of {**missingRecords100m2QF, missingRecords10m2QF**, **missingRecords1m2QF** from step b (above)} are >0:
        i. Create *1* new record in *div_400m2Data_temp* and assign counts calculated in step b (above) to the corresponding columns
        ii. Assign **eventID**, **domainID**, **siteID**, **plotID** and boutNumber to this record accordingly, using information in the given {**plotID, eventID**} combination
        iii. Assign to the **date** field: the latest date in the list of dates (from non-missing records) corresponding to the given {**plotID, eventID**} combination
        iv. generate a new **uid** for each newly created record

2. Generate spatial location and uncertainty information
    a. Insert corresponding values of **decimalLatitude, decimalLongitude, geodeticDatum, coordinateUncertainty , nlcdClass, elevation,** and **elevationUncertainty** values into fields of the same name, in *div_400m2Data_temp*, for each **plotID** (where **subType** = 'Base Plot'), using the TOS plot-level spatial data lookup table (AD[13]).
3. Pass L0 data from *div_400m2Data _temp* to *div_ 400m2Data_pub*

6 UNCERTAINTY

6.1 Analysis of Uncertainty

6.2 Reported Uncertainty

7 VALIDATION AND VERIFICATION

7.1 Algorithm Validation

7.2 Data Product Validation

7.3 Data Product Verification

8 SCIENTIFIC AND EDUCATIONAL APPLICATIONS

9 FUTURE MODIFICATIONS AND PLANS

Plans include: (1) an improved process for updating data products based on post-processing taxonomic identification of morphospecies (2) a flag for remaining potential taxonomic duplicates in *div_400m2Data _pub* (3) a flag for missing div1m2variables (4) removal of records in 10m2 and 100m2 where targetTaxaPresent –'N' was used to indicate that no additional plant taxa were recorded in 10m2 and/or 100m2 plots (currently, targetTaxaPresent ='N' may be used in some cases where plants are present in larger plots, no further plants were added to the list beyond what technicians found in the nested subplots) (5) ability for technicians to access previously entered data to prevent inadvertent duplicate data entry, and update information where appropriate.

10 BIBLIOGRAPHY

Adler, P. B., E. W. Seabloom, E. T. Borer, H. Hillebrand, Y. Hautier, A. Hector, W. S. Harpole, L. R. O'Halloran, J. B. Grace, T. M. Anderson, J. D. Bakker, L. A. Biederman, C. S. Brown, Y. M. Buckley, L. B. Calabrese, C. J. Chu, E. E. Cleland, S. L. Collins, K. L. Cottingham, M. J. Crawley, E. I. Damschen, K. F. Davies, N. M. DeCrappeo, P. A. Fay, J. Firn, P. Frater, E. I. Gasarch, D. S. Gruner, N. Hagenah, J. H. R. Lambers, H. Humphries, V. L. Jin, A. D. Kay, K. P. Kirkman, J. A. Klein, J. M. H. Knops, K. J. La Pierre, J. G. Lambrinos, W. Li, A. S. MacDougall, R. L. McCulley, B. A. Melbourne, C. E. Mitchell, J. L. Moore, J. W. Morgan, B. Mortensen, J. L. Orrock, S. M. Prober, D. A. Pyke, A. C.

Risch, M. Schuetz, M. D. Smith, C. J. Stevens, L. L. Sullivan, G. Wang, P. D. Wragg, J. P. Wright, and L. H. Yang. 2011. Productivity Is a Poor Predictor of Plant Species Richness. Science **333**:1750-1753.

Barnett, D. T. and T. J. Stohlgren. 2003. A nested-intensity design for surveying plant diversity. Biodiversity and Conservation **12**:255-278.

Barnett, D. T., T. J. Stohlgren, C. Jarnevich, S., G. W. Chong, J. Ericson, A., T. Davern, R., and S. Simonson, E. 2007. The art and science of weed mapping. Environmental Monitoring and Assessment **132**:235-252.

Clark, J. S., S. LaDeau, and I. Ibanez. 2004. Fecundity of trees and the colonization-competition hypothesis. Ecological Monographs **74**:415-442.

Clark, J. S., D. M. Bell, M. Kwit, A. Stine, B. Vierra, and K. Zhu. 2012. Individual-scale inference to anticipate climate-change vulnerability of biodiversity. Philosophical Transactions of the Royal Society B-Biological Sciences **367**:236-246.

Crall, A., W., G. Newman, T. J. Stohlgren, C. Jarnevich, S., P. Evangelista, and D. Guenther. 2006. Evaluating dominance as a component of non-native species invasions. Diversity and Distributions **12**:195-204.

Darwin, C. 1859. The Origin of Species. Reprinted by Penguin Books, London.

Dornelas, M. 2010. Disturbance and change in biodiversity. Philosophical Transactions of the Royal Society B-Biological Sciences **365**:3719-3727.

Ibanez, I., J. S. Clark, M. C. Dietze, K. Feeley, M. Hersh, S. LaDeau, A. McBride, N. E. Welch, and M. S. Wolosin. 2006. Predicting biodiversity change: Outside the climate envelope, beyond the species-area curve. Ecology **87**:1896-1906.

Magurran, A. E. and M. Dornelas. 2010. Biological diversity in a changing world. Philosophical Transactions of the Royal Society B-Biological Sciences **365**:3593-3597.

Magurran, A. E. and B. McGill. 2010. Challenges and opportunities in the measurement and assessment of biological diversity.*in* A. E. Magurran and B. McGill, editors. Biological diversity: Frontiers in measurement and assessment. Oxford University Press, New York.

Pickett, S. T. A., J. Kolasa, J. J. Armesto, and S. L. Collins. 1989. THE ECOLOGICAL CONCEPT OF DISTURBANCE AND ITS EXPRESSION AT VARIOUS HIERARCHICAL LEVELS. Oikos **54**:129-136.

Pickett, S. T. A. and P. S. White, editors. 1985. The Ecology of Natural Disturbance and Patch Dynamics. Academic Press, New York, NY.

Stohlgren, T. 2007. Measuring plant diversity, lessons from the field. Oxford University Press, New York.

Stolte, K. W. 1997. 1996 National Technical Report on Forest Health. Admin. Report FS-605, U.S. Department of Agriculture, Forest Service, Southern Research Station, Asheville, NC.

Vellend, M. 2010. CONCEPTUAL SYNTHESIS IN COMMUNITY ECOLOGY. Quarterly Review of Biology **85**:183-206.

Wagner, P. J. and S. K. Lyons. 2010. Estimating extinction with the fossil record.*in* A. E. Magurran and B. McGill, editors. Biological diversity: Frontiers in measurement and assessment. Oxford University Press, New York.

11 CHANGELOG

12 APPENDIX: DATA ENTRY VALIDATION VIA PDA & WEBUI

12.1 All Data: Run the following steps for all fields in the plant diversity data ingest workbook

1. Provide plot map following Figure 1, prompting technicians to enter data for all 20 subplots per plot
2. Warn technicians submitting data from a plot with data entries from fewer than all 20 subplots.
3. Constrain entered values to the correct **dataType**
4. Constrain entered values to conditions specified by **entryValidationRules**
5. Generate a unique ID (**uid**) for each record
6. Follow guidelines for fields in which no data have been entered, as specified in **noDataOutcomePDA** and **noDataOutcomeUI**:
    a. If **noDataOutcome** = fail:
        i. do not let user finalize record until a value is provided
        ii. warning message text, ‘Please enter a value for [fieldname] to continue’, unless an alternative is provided in **warningText**
    b. Else if **noDataOutcome** = warn:
        i. warn user that a value is missing prior to finalizing record, but allow selection of ‘OK’ to continue without a value
        ii. warning message text, ‘Please confirm that there is no for [fieldname] to continue’, unless an alternative is provided in **warningText**
    c. Else if **noDataOutcome** = pass:
        i. allow user to finalize record with no values in this field
7. Follow guidelines for default values, as specified in **defaultValuePDA** and **defaultValueUI**, and **case**
8. Anonymize all technicianIDs (i.e., fields ending in ‘By’; e.g., recordedBy, identifiedBy)

12.2 Plant Diversity Table 1: div_1m2Data_in

1. Provide filtered drop-down menus for the location fields in the *div_dataingest_2014* workbook, as follows:
    a. domainID and siteID, using the TOS subplot-level spatial data lookup table (AD[13])
    b. plotID, using the TOS subplot-level spatial data lookup table (AD[13]) and where **subType =** ‘Base Plot’
    c. subplotID = { '31.1.1', '31.4.1', '32.2.1', '32.4.1', '40.1.1', '40.3.1', '41.1.1', '41.4.1'}

2. Provide filtered drop-down menu and/or typeaheads for controlled lists for all fields except taxonIDRemarks, morphospeciesIDRemarks, remarks, morphospeciesID, percentCover, heightPlantSpecies, identificationReferences, measuredBy, recordedBy.
3. recordedBy and measuredBy – Maximo user list for FOPs or 'Other'
    a. PDA solution: just type in (i.e., typeahead field) rather than select other
    b. UI solution: TBD
    c. PDA will send values from recordedBy as entered on the PDA to both 'recordedBy' and 'enteredBy' in PDR
4. Provide radio-button for divDataType, and allow user to enter either the set of plantSpecies or otherVariables fields as appropriate
5. If 'plantSpecies' is selected:
    a. Provide radio-button for targetTaxaPresent
        i. If 'Y' is selected, continue to (b)
    b. identificationQualifier: lookup table is a list of valid entries, with associated descriptions
        i. show description but send identificationQualifier to Maximo/PDR
    c. Populate taxonID – Use lookup table to create typeahead fields:
        i. lookup table: NEON.DOC.###### NEON Taxonomic Name List of Plants
        ii. field is a typeahead in which the user can type in taxonID or scientificName
        iii. filtering rules: select (unique(taxonID, scientificName)) WHERE
            A. lookuptable.domainID == current domainID AND
            B. D##NativeStatusCode != 'A'
            C. If OTHE (Other – plant) is selected, prompt user to enter additional information in the remarks
        iv. Only taxonID and acceptedTaxonID are sent to Maximo/PDR
        v. Prevent duplicate entries for 'taxonID' in the same 'subplotID'
    d. MorphospeciesID handling – first release is free text field
    e. Provide radio button for heightPlantOver300cm
        i. If 'N' is selected, prompt user to enter value into heightPlantSpecies
6. If 'otherVariables' is selected:
    a. Provide radio-button for otherVariablesPresent
    b. If 'Y' is selected in (a), provide drop-down menu of otherVariables and prompt user to enter value into percentCover
    c. Prevent duplicate entries for 'otherVariables' in the same 'subplotID'

12.3 Plant Diversity Table 2: 10m2data100m2data Data

1. Design of the data entry for this table is identical to all of the plant species entry in Table 1, describe above. The primary differences are:

a. The list of values for subplotID is different (see AD[10]):
    i. subplotID = {'31’, '32’, '40', '41', '31.1.10', '31.4.10', '32.2.10', '32.4.10', '40.1.10', '40.3.10', '41.1.10', '41.4.10'}
b. No data on ‘otherVariables’ are collected at these scales
c. No data on percentCover or height of plants are collected at these scales