Site Characterization Data Quality Report

Report prepared by Cody Flagg and Katie Jones

Contents

1 Overview

This data quality report is based on data files received from ARCADIS on 2/3/2015 forthe five sites at which vegetation characterization field work occurred (over the period November 10th - January 21st). The quality checks contained in this report represent the extent of what we are capable of assessing without field checking contractor performance. An additional quality report may still be filed for field work. For the purpose of this summary, we have divided our concerns into two general areas - issues related to **species identification** and issues related to **data entry**:

Species ID - Two issues arose with respect to species identifications:

1. Incomplete identification - At three sites, HARV, SCBI and TALL, only a portion of the dataset includes values in the "`scientificName`" field (29% complete at HARV, 67% complete at SCBI and 16% complete at TALL). Field work at all of these sites was initiated following fall senescence. In recognition of increased difficulty in positively identifying individuals to species, we relaxed the acceptable resolution to genus level at HARV and SCBI (10/30/14 Git issue #3). This was reiterated in our 11/31/14 conference call as species ID should be to "*Finest resolution possible given the conditions*" (see meeting notes posted on GitHub). **In no instance did we agree that identification was not necessary**.

2. Use of incorrect data type in `scientificName` field - NEON uses USDA Plants as the taxonomic authority for naming/coding conventions but the data entry field for species is `scientificName`, which is defined in the Access DB `vstFieldSummary` table as 'Binomial latin name', this is reinforced through the constrained set of values available as a dropdown list in the `scientificName` field in the `vst_perindividual_in` table (See guidance in Data Transcription protocol section 8.3.2.e.iii).

Nonetheless, all `scientificName` entries in the HARV and JERC datasets are codes. For JERC, converting from USDA code to `scientificName` should be simple to update as all codes appear to be valid USDA Plants accepted codes. On the other hand, in the HARV dataset, most of the codes used *are not* USDA codes, we have no way to confidently assign a `scientificName` to most of these records. See the print out of unique values in the `scientificName` field (section 3.1, below) to identify additional errors; **most of these are avoidable by utilizing the drop down list in the Access DB**.

Data entry - Some of these errors may be attributable to data transcription, others may be errors that originate in field data collection. In either case, data with these errors cannot be ingested into the NEON database.

- `Numeric values in remarks field`. **Are these values data?** Remarks are not machine readable, as such, we are unable to make use of essential information if it is included in this field.
- `Invalid pointIDs`, NEON cannot generate georeferenced stem maps if stems are mapped from points other than those identified in the protocol
- `Invalid growthForms`, only growthForms listed in the protocol will be recognized by our ingest algorithms.

- `Invalid Azimuth values`, individuals are plotted as being outside of the designated plot boundaries.

See site specific summaries below and supplemental error reports (in folder "`QA_Files`") for more details.

2 Plot Completion

- Provides a summary for which data are reported at each site ## Plot Count
- Quick check: How many plots were visited?
- HARV, SCBI and JERC have blank entries in field: `plotID`.

Table 1: Plot Count by Site

Site	PlotCount
HARV	18
SCBI	20
JERC	20
OSBS	15
TALL	20

2.1 Missing Plots

- These plots are in the provided NEON Excel files for each site, but were not found in the data files received from Arcadis.
- Git issue #9 documents that NEON Field Ops completed `HARV_036 and HARV_050`, please verify if this is also the case for missing plots at OSBS. SCBI appears to be a data entry error, please fix.

```
## [1] "HARV_050" "HARV_036" "OSBS_025" "OSBS_036" "OSBS_042" "OSBS_043"
## [7] "OSBS_044"
```

2.2 Extra Plots

- This lists plots that were found in the received files from Arcadis that are not included in the lists provided by NEON.
- This appears to be a data entry error.

```
## character(0)
```

3 Duplicate and Unique Tags

- This summarizes the number of duplicate `tagID` entries per site.
- See the supplemental .csv files (File suffix: `_duplicateTags`) for lists of specific duplicate tag IDs by site.
- The first row in this table (with a blank `siteID`) indicates that 8 entries are missing a `plotID`.

Table 2: Duplicate Tags by Site

siteID	Count
	4
HARV	120
JERC	3
OSBS	9
SCBI	1

3.1 Unique species list by site (`scientificName`)

- Some unique records may be attributable to spelling and capitalization errors during data transcription

```
## [1] "HARV"
##
## *
## * Betula
## * Fagus grandifolia
## * Kalmia latifolia
## * Pinus
## * Quercus
## * Tsuga canadensis
##
## <!-- end of list -->
##
## [1] "SCBI"
##
## *
## * Acer
## * Acer negundo
## * Acer rubrum
## * Berberis
## * Berberis thunbergii
## * Carya
```

```
## * Carya tomentosa
## * Celtis occidentalis
## * Cornus florida
## * Diospyros virginiana
## * fraxinus
## * Fraxinus
## * Fraxinus americana
## * Fraxinus americana var. microcarpa
## * Fraxinus pennsylvanica
## * Juglans nigra
## * Lindera benzoin
## * Liquidambar styraciflua
## * Liriodendron tulipifera
## * Platanus occidentalis
## * Prunus serotina
## * Quercus
## * Quercus alba
## * Quercus rubra
## * Rosa multiflora
## * Rubus
## * Rubus phoedalibarda
## * Rubus phoenicolasius
## * Sassafras albidum
## * Sassafras albidum var. molle
## * Symphoricarpos orbiculatus
## * Ulmus
## * Ulmus americana
## * Vitis
##
## <!-- end of list -->
##
## [1] "JERC"
##
## *
## * Acer floridanum
## * Aesculus pavia
## * Bignonia capreolata
## * Bignonia capreolataBignonia capreolata
## * Campsis radicans
## * Carya glabra
## * Celtis laevigata
## * Celtis occidentalis
## * Cercis canadensis
## * Cocculus carolinus
## * Cornus florida
## * Crataegus flava
## * Crataegus flava Aiton
## * Crataegus spathulata
## * Diospyros virginiana
## * Fraxinus americana
## * Gelsemium sempervirens
## * Ilex glabra
## * Ilex vomitoria
## * Juniperus virginiana
```

```
## * Liquidambar styraciflua
## * Malus angustifolia
## * Morus rubra
## * Nyssa sylvatica
## * Parthenocissus quinquefolia
## * Pinus clausa
## * Pinus echinata
## * Pinus palustris
## * Prunus americana
## * Prunus serotina
## * Quercus falcata
## * Quercus geminata
## * Quercus hemisphaerica
## * Quercus hemispherica
## * Quercus incana
## * Quercus laevis
## * Quercus lyrata
## * Quercus margaretta
## * Quercus minima
## * Quercus nigra
## * Quercus virginiana
## * Rhus copallinum
## * Sassafras albidum
## * Sideroxylon languinosum
## * Sideroxylon lanuginosum
## * Smilax bona-nox
## * Smilax rotundifolia
## * Smilax smallii
## * Taxodium distichum
## * Tilia americana
## * Toxicodendron radicans
## * Ulmus alata
## * Unknown
## * Vaccinium arboreum
## * VIBURNUM
## * Viburnum prunifolium
## * Viburnum sp.
## * Vitis rotundifola
## * Vitis rotundifolia
##
## <!-- end of list -->
##
## [1] "OSBS"
##
## *
## * Asimina incana
## * Diospyros virginiana
## * Pinus palustris
## * Quercus geminata
## * Quercus geminata Small
## * Quercus laevis
## * Quercus margarettae
## * Vaccinium arboreum
##
```

```
## <!-- end of list -->
##
## [1] "TALL"
##
## *
## * Acer rubrum
## * Carya
## * Cornus
## * Ilex opaca
## * Liquidambar styraciflua
## * Liriodendron tulipifera
## * Pinus
## * Pinus palustris
## * Pinus taeda
## * Quercus
## * Quercus alba
## * Rhus
## * Symplocos tinctoria
## * Vaccinium arboreum
##
## <!-- end of list -->
```

3.2 Unique growth forms list by site

- **Acceptable values** : lia, sbt, mbt, sms, sis, smt, sap

```
## $HARV
## [1] ""     "mbt" "sap" "sbt" "sis" "smt"
##
## $SCBI
## [1] ""     "lia" "mbt" "sap" "sbt" "sis" "sms" "smt"
##
## $JERC
## [1] "lia" "mbt" "sap" "sbt" "sis" "sms" "smt"
##
## $OSBS
## [1] ""     "mbt" "sap" "sbt" "sis" "sms" "smt"
##
## $TALL
## [1] "lia" "mbt" "sap" "sbt" "sis" "sms" "smt" "snt"
```

4 Point and Azimuth Validation

- This checks that all pointIDs are valid for plot dimensions and azimuth values are within pre-defined ranges for specific pointID.
- See supplemental .csv files (file suffix: `pointID_azimuth_QF`) for specific details regarding where `pointID` and `stemAzimuth` errors occur.

- "`pointIDQF`" and "`azimuthQF`" fields flag errors for each row in the supplemental files; "**1**" = no error, "**-9999**" = potential error.

Table 3: Point/Azimuth Errors by Site

	siteID	pointID.error	azimuth.error
2	HARV	0	3
3	JERC	0	0
4	OSBS	0	9
5	SCBI	0	6
6	TALL	0	8

4.1 Growth Form Measurements by Site

- This table summarizes the types of growth forms that have stem distance and azimuth measurements.
- Note that a blank in the "growthForm" column suggests there is a missing entry.

Table 4: Growth Form Measurements by Site

siteID	growthForm	Count
		5
HARV		1
HARV	mbt	562
HARV	sap	352
HARV	sbt	2213
HARV	sis	8
HARV	smt	1339
JERC	lia	160
JERC	mbt	23
JERC	sap	1096
JERC	sbt	455
JERC	sis	59
JERC	sms	379
JERC	smt	347
OSBS		3
OSBS	mbt	14
OSBS	sap	113
OSBS	sbt	659
OSBS	sis	4
OSBS	sms	373
OSBS	smt	318
SCBI		1
SCBI	lia	63
SCBI	mbt	53
SCBI	sap	488
SCBI	sbt	930
SCBI	sis	512
SCBI	sms	288
SCBI	smt	290
TALL	lia	5
TALL	mbt	21
TALL	sap	544
TALL	sbt	732
TALL	sis	93
TALL	sms	111
TALL	smt	276
TALL	snt	1

5 Missing Scientific Names

- This table summarizes how many scientific names are missing (i.e. blanks, NA, unknown, unk etc.).

Table 5: Missing Scientific Names by Site

siteID	NA	Blank	Unknown	Blank%
HARV	0	3425	0	76.468
SCBI	0	851	0	32.419
JERC	0	1	1	0.040
OSBS	0	2	0	0.135
TALL	0	1499	0	84.072

6 Tagged Plants by Site

- This table summarizes the number of tags per plot per site.

Table 6: Tags Used by Site

siteID	minTags	maxtags	meanTags
HARV	4	380	236
SCBI	54	247	131
JERC	48	331	126
OSBS	1	145	93
TALL	36	151	89