TOS PROTOCOL AND PROCEDURE: MANUAL DATA TRANSCRIPTION

PREPARED BY	ORGANIZATION	DATE
Julia Spencer	FSU	04/25/2014
Katherine M. Thibault	FSU	01/05/2015
Sarah C. Elmendorf	DPS	01/05/2015
Nick Schroeter	FOPs	01/05/2015

APPROVALS	ORGANIZATION	APPROVAL DATE
Hanne Buur	PSE	04/10/2015

RELEASED BY	ORGANIZATION	RELEASE DATE
Judy Salazar	CM	04/10/2015

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	05/30/2014	ECO-01835	Initial release
B	01/13/2015	ECO-02633	Added information regarding recent developments of web interfaces and mobile applications for data entry; added error rate tracking table and query to Access databases; added additional information for AQU protocols; added quick reference

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

1 DESCRIPTION

1.1 Purpose

The primary purpose of this document is to provide a change-controlled version of Observatory protocols and procedures. This document provides the content for training and field-based materials for NEON staff and contractors. Documentation of content changes (i.e. changes in particular tasks or safety practices) will occur via this change-controlled document, not through field manuals or training materials.

1.2 Scope

This document is a cross-protocol set of instructions for data transcription and storage when manual data recording is necessary. Where PDA devices are not available for data collection, it is important to have a standardized process for entering data collected in the field or in the domain support facility (hereafter referred to as 'the lab') to ultimately be ingested by NEON Cyber Infrastructure (CYI).

1.3 Acknowledgments

This document was informed by the Rocky Mountain Bird Observatory 2011 Point Transect Quality Assurance/Quality Control Protocol (Birek et al. 2011).

2 RELATED DOCUMENTS AND ACRONYMS

2.1 Applicable Documents

Applicable documents contain information that shall be applied in the current document. Examples are higher level requirements documents, standards, rules and regulations.

AD [01]	NEON.DOC.001398	NEON Raw Data Ingest Workbook for TOS Digital Hemispherical Photos
AD [02]	NEON.DOC.001399	NEON Raw Data Ingest Workbook for TOS Plant Diversity
AD [03]	NEON.DOC.001400	NEON Raw Data Ingest Workbook for TOS Ground Beetle Abundance and Diversity
AD [04]	NEON.DOC.001401	NEON Raw Data Ingest Workbook for TOS Mosquito Abundance, Diversity and Phenology
AD [05]	NEON.DOC.001402	NEON Raw Data Ingest Workbook for TOS Rodent-borne pathogen sampling
AD [06]	NEON.DOC.001403	NEON Raw Data Ingest Workbook for TOS Terrestrial Biogeochemistry: Chemistry of Soils and Plants
AD [07]	NEON.DOC.001404	NEON Raw Data Ingest Workbook for TOS Terrestrial Biogeochemistry: Stable Isotopes in Soils and Plants
AD [08]	NEON.DOC.001405	NEON Raw Data Ingest Workbook for TOS Soil Microbial Diversity
AD [09]	NEON.DOC.001406	NEON Raw Data Ingest Workbook for TOS Small Mammal Abundance

		and Diversity
AD [10]	NEON.DOC.001407	NEON Raw Data Ingest Workbook for TOS Breeding Bird Abundance and Diversity
AD [11]	NEON.DOC.001408	NEON Raw Data Ingest Workbook for TOS Plant phenology observations
AD [12]	NEON.DOC.001920	NEON Raw Data Ingest Workbook for TOS Herbaceous Plant Biomass
AD [13]	NEON.DOC.001921	NEON Raw Data Ingest Workbook for TOS Coarse Woody Debris: Field Tally Data
AD [14]	NEON.DOC.001922	NEON Raw Data Ingest Workbook for TOS Coarse Woody Debris: Bulk Density Field and Lab Data
AD [15]	NEON.DOC.001923	NEON Raw Data Ingest Workbook for TOS Coarse Woody Debris: Raw Void Volume Images
AD [16]	NEON.DOC.001924	NEON Raw Data Ingest Workbook for TOS Litterfall and Fine Woody Debris
AD [17]	NEON.DOC.001925	NEON Raw Data Ingest Workbook for TOS Belowground Biomass Soil Core
AD [18]	NEON.DOC.001926	NEON Raw Data Ingest Workbook for TOS Leaf Area Index by Species
AD [19]	NEON.DOC.001927	NEON Raw Data Ingest Workbook for TOS Mat-forming Bryophyte Productivity
AD [20]	NEON.DOC.001928	NEON Raw Data Ingest Workbook for TOS Vegetation Structure
AD [21]	NEON.DOC.001929	NEON Raw Data Ingest Workbook for TOS Leaf Area Index

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
RD [04]	NEON.DOC.005003	NEON Scientific Data Products Catalog
RD [05]	NEON.DOC.014051	Field Audit Plan
RD [06]	NEON.DOC.000824	Data and Data Product Quality Assurance and Control Plan

2.3 Acronyms

MS	Microsoft

2.4 Definitions

A **protocol** is a formal summary description of a procedure and its related rationale, and includes information on the knowledge and resources needed to implement the procedure. A **procedure** is a set of prescribed actions that must take place to achieve a certain result, and can also be called a method. It differs from a science design in that science designs provide a more complete description of the rationale for selecting specific protocols. It differs from a training manual in that training manuals

provide materials in support of skills acquisition in the topic areas including information on how to best train staff rather than detailing only the steps of the procedure.

Field datasheet – A pdf document to be printed and filled out by hand in the field or the lab.

Data ingest workbook – A Microsoft (MS) Excel workbook that contains multiple worksheets that include descriptions of the data fields and example worksheets for data entry. These workbooks are intended for (1) CYI reference, in association with the relevant Algorithm Theoretical Basis Documents (ATBDs), (2) creating the associated data ingest databases for FOPs data entry, and (3) data entry by external contractors or, for some modules, by FOPs (Table 1).

Data ingest database – An MS Access database that contains multiple tables for controlled data transcription for FOPs-collected data, temporary data storage, and associated lookup tables.

Lookup table – Contains the valid values for specific fields; these tables are indicated by an 'L' at the beginning of the table name in the data ingest databases; e.g. "L_plant_species_list".

E-datasheet – A worksheet in the data ingest workbook or a table in the data ingest database that is intended for data entry; these are indicated by an '_in' at the end of the worksheet or table name; e.g. "mam_capturedata_in".

WebUI – a protocol-specific user interface accessible within the NEON network; created by NEON CYI to provide controlled data entry for use by NEON FOPs technicians to submit data to the NEON Processed Data Repository (PDR).

PDA – a portable digital assistant that runs protocol-specific applications created by NEON CYI to provide controlled data entry for use by NEON FOPs technicians during the data collection effort in the field and then submit data to the NEON Processed Data Repository (PDR).

3 BACKGROUND AND OBJECTIVES

Although the use of mobile devices to capture data electronically in the field is ultimately planned for NEON sampling efforts, the full suite of needed mobile applications has yet to be developed. Moreover, it is anticipated that, throughout the course of NEON operations, the use of mobile devices will not be possible at all times, due to equipment failure or other unpredictable logistical complications. Consequently, some observational data will require manual data transcription by FOPS personnel from paper field datasheets to be ingested by NEON CYI. Manual transcription introduces a number of additional steps to be completed by Field Operations staff, such as file naming, management, and storage. In addition, the transcription process potentially introduces novel sources of error to the data products. Since NEON will be generating a prohibitively large volume of observational data to allow for manual checks of every record, this document describes a quality assurance method that will facilitate identification of errors prior to the automated portion of the QA/QC algorithms.

4 PROTOCOL

The strategy for manual data entry during NEON operations consists of four general parts. The first component of the process is datasheet review and digitizing (i.e., scanning). The recommended practice is to have one technician who is leading a particular protocol review field datasheets at the end of the sampling day. During this review, needed clarifications or additions to the raw data are often discovered, and appropriate changes and notations should be made on the raw datasheets in red ink, with associated date of corrections and initials of technician. After this initial review, the field datasheets should be scanned, and the resulting files saved in the designated network location. The second is entering the data into the digital data ingest workbooks or databases or the web user interfaces (webUIs), where available (Table 1). Third, technicians shall perform a manual quality assurance check on ten percent of the data entered to ensure that best data entry practices are being followed. The volume of data collected by NEON Field Operations to meet NEON requirements prohibits manual checks of the majority of the data records. The selection of ten percent reflects the trade-off between using available labor to collect more data and proofing all of the data that are collected. Additionally, further quality checking of all ingested data will be automated by CYI (as described in NEON's ATBDs). The fourth and final part of the manual data entry process is saving the e-datasheets with the appropriate naming convention in the designated location for storage.

Data ingest files contain multiple worksheets or tables for data entry. E-datasheets; e.g., AD[01 - 22], and data ingest databases contain controlled values and vocabularies; these tables of controlled values are referred to as lookup tables and constrain values that can appear in certain fields. Before entering data into any e-datasheet, be sure to review the "FieldSummary" worksheet or table to confirm what the intended content and format of each field are, as well as which fields or values are associated with a provided lookup table. Tables with the suffix '_in' in the name should be used for data entry.

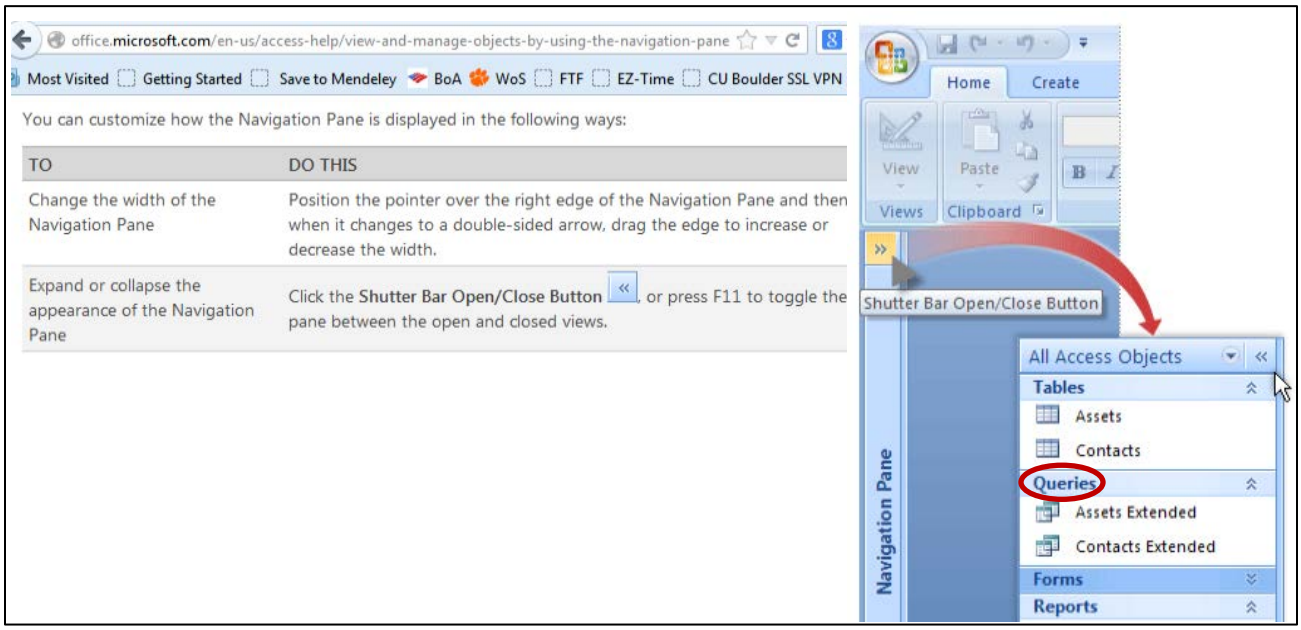

Figure 1. Using the Navigation Pane in MS Access to explore the database tables, queries, and forms. From office.microsoft.com – accessed 3/2014

Table 1. Data entry formats and options by NEON protocol, 2015. Note that file names that include DXX indicate that each domain has a unique version of the file in order to accommodate large species lists or other constraints.

Sub-system	Short module description	Protocol document number	Datasheet document number	Data entry file name(s)	Data entry file format
TOS	Vegetation structure	000987	001573	DXX_vst_dataIngest_2015	accdb
TOS	Herbaceous biomass	014037	001574	NA	webUI
TOS	Leaf area index	014039	001575	dhp_dataIngest_2015	accdb
TOS	Soil measurements	014048	001577	sls_dataIngest_2015	accdb
TOS	Plant phenology	014040	001578	NA	webUI
TOS	Plant diversity	014042	001579	NA	webUI
TOS	Ground beetle sampling	000451	001580	NA	webUI
TOS	Mosquito sampling	014049	001581	mos_dataIngest_2015	accdb
TOS	Tick sampling	014045	001583	tck_dataIngest_2015	accdb
TOS	Small mammal sampling	000481	001585	NA	webUI
TOS	Rodent- borne Pathogen sampling	014044	001402	rpt_dataIngest_2015	accdb
TOS	Litter and fine woody debris	001710	002132	ltr_dataIngest_2015	accdb
TOS	Mat-forming Bryophyte Productivity	001709	002136	bry_dataIngest_2015	accdb

TOS	Coarse downed wood	001711	002121	cdw_dataIngest_2015	accdb
TOS	Aboveground Plant Biomass in Agricultural Systems	001714	TBD	TBD	accdb
TOS	Cactus Biomass and Handling	001715	TBD	TBD	accdb
TOS	Toxicodendron Biomass and Handling	001716	TBD	TBD	accdb
TOS	Ground Beetle and Mosquito Specimen Processing	001100	001582	TBD	accdb
AOS	Macroinvertebrates Sampling in Wadeable Streams	000690	002197	inv_strFieldDataIngest_DXX	csv
AOS	Periphyton and Seston Sampling in Wadeable Streams	000691	002199	alg_strFieldDataIngest_DXX; alg_str_DLabDataIngest_DXX	csv
AOS	Aquatic Plant, Bryophyte, Lichen, and Macroalgae Sampling in Wadeable Streams	000692	002196	apl_strDlabDataIngest_DXX; apl_strTransectEstDataIngest_DXX; apl_strTransectDataIngest_DXX; apl_strQuadratDataIngest_DXX	csv
AOS	Microbes in Lakes and Non-wadeable Streams	001200	002194	amc_lakFieldDataIngest_DXX	csv
AOS	Microbes in Wadeable Streams	001201	002198	amc_strFieldDataIngest_DXX	csv
AOS	Aquatic Plant and Macroalgae Sampling in Lakes and Non-wadeable Streams	001202	002195	apl_lakFieldDataIngest_DXX; apl_lakDLabDataIngest_DXX	csv
AOS	Algae Sampling in Lakes and Non-wadeable Streams	001203	002192	alg_lakFieldDataIngest_DXX; alg_lakDLabDataIngest_DXX	csv
AOS	Macroinvertebrate sampling in Lakes and Non-wadeable Streams	001204	002193	inv_lakFieldDataIngest_DXX	csv
AOS	Zooplankton Sampling in Lakes	001194	002302	zoo_lakFieldDataIngest_DXX	csv
AOS	Secchi Depth and Depth Profile Sampling	NA	002191	dep_lakFieldSecchiIngest_DXX; dep_lakFieldDepthIngest_DXX	csv
AOS	Aquatic Field Metadata	NA	001646	Field Metadata	accdb

AOS	Stream Water Chemistry	000694	002383	Field Water Chem	accdb
AOS	Alkalinity and ANC	000694	002383	Domain Lab Alkalinity	accdb
AOS	Dissolved Gas – streams and Lakes	001199	002384	sdg_dataIngest_DXX	csv
AOS	Isotopes – streams, lakes and groundwater	001886	002211	asi_dataIngest_DXX	csv
AOS	Reaeration	000693	002382	rea_dataIngest_DXX; logger_site_stationID_YYYYMMDD	Xlsx; csv
AOS	Lake and River Water chem	001190	002686	Field Water Chem	accdb
AOS	Groundwater Chem	001219	002690	Field Water Chem; gwc_dataIngest_ DXX (for GWC field metadata only)	Accdb; csv
AOS	Discharge	001085	NA	XXXXMMDD; where XXXX is the 4-letter site code	csv
AOS	Bathymetry and morphology of lakes and non-wadeable streams	001197	NA	bat _lakFieldDataIngest_DXX	csv
AOS	Riparian Mapping - Streams	001196	NA	srm_strFieldDataIngest_DXX	csv
AOS	Riparian Mapping – lakes and non-wadeable streams	001195	NA	lrm_lakFieldDataIngest_DXX	csv
AOS	Fish Sampling – wadeable streams	001295	NA	fsh_strFieldDataIngest_DXX	csv
AOS	Fish Sampling – lakes	001296	NA	fsh_lakFieldDataIngest_DXX	csv
AOS	Sediment Chemistry - wadeable streams	001193	002419	ssc_strFieldDataIngest_DXX	csv
AOS	Sediment Chemistry - lakes and non-wadeable streams	001191	NA	lsc_lakFieldDataIngest_DXX	csv
STR	STREON sediment basket metadata	001946		TBD	TBD
STR	STREON chamber incubation data	001947		TBD	TBD
STR	STREON chamber biological sampling	001949		TBD	TBD

Once data are entered into the appropriate Access table, Excel spreadsheet, or webUI, the e-datasheet should eventually be exported to a csv file. Domain technicians shall save e-datasheets from Access or Excel-based data entry in an initial staging area on the network (\\10.100.128.37\dropbox; hereafter referred to as the FOPS dropbox), for subsequent retrieval by Science and CYI (**Figure 2**). E-datasheets generated via the webUIs need only be stored locally within each domain.

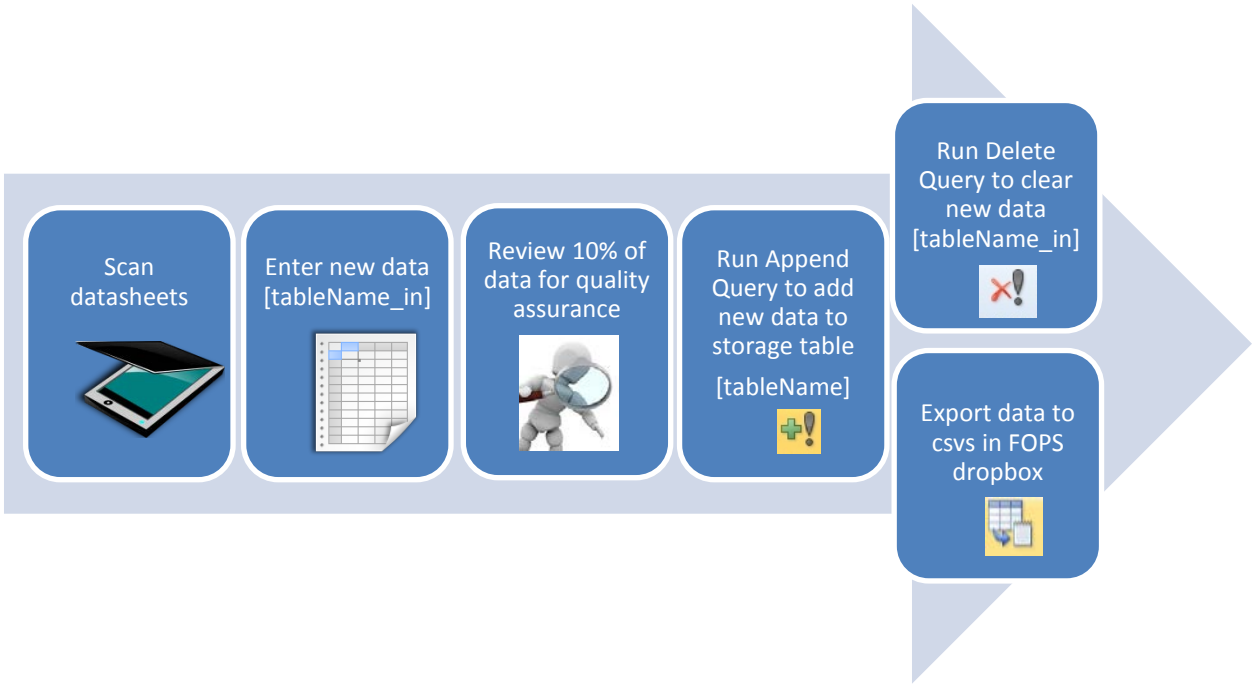

Figure 2. Example workflow for data transcription using Access data ingest databases

5 QUALITY ASSURANCE AND CONTROL

All data entered from the field datasheets into the e-datasheets or webUIs shall be quality checked by selecting 10% of the data to recheck against the hand-written field datasheet. As a best practice, it is recommended, although not required, that double-checking include two technicians working in concert, one to review the e-datasheet while the other reads the correct values aloud directly from the field datasheet. All discovered errors shall be corrected. Per-protocol error rates found in data transcription (via webUI or any other means) shall be reported to NEON staff scientists using a spreadsheet on the NEON intranet (FSU-FOPS/dataEntry2015 folder) on an, annual basis. Additional details for conducting the QA/QC can be found in section 8.6 below.

6 PERSONNEL REQUIREMENTS

As a best practice, the technicians entering data should be those that collected the data for the specific protocol in the field.

7 TRAINING REQUIREMENTS

Technicians shall have basic knowledge of Microsoft Excel and Access and internet browser software and have familiarized themselves with the relevant data ingest sheets. Prior to using a particular webUI, technicians shall enter a plot (or subplot) of data from one bout into the webUI housed on the Training portal (see section 8.5 for details).

8 STANDARD DATA TRANSCRIPTION PROCEDURE

Transcribe data into the e-datasheets or databases as soon as possible after field data collection, as technician recall will help to improve data quality. As a best practice, the data should be entered within 7 days of collection or the end of a sampling bout (where applicable). However, given logistical constraints, the maximum timeline for entering data is within 14 days of collection or the end of a sampling bout (where applicable).

8.1 Scan Field/Lab Datasheets

1) The technician who is leading a particular protocol should review field/lab datasheets at the end of the sampling day. Be sure that datasheets have been filled completely. Check BOTH sides of the datasheets for extra taxa or notes. Be sure to clarify any shorthand used; don't assume the shorthand will be meaningful to all.

2) Any corrections, clarifications, or additions to the raw data should be made on the datasheets in red ink, with associated date of corrections and initials of technician.

3) Datasheets with clarifications and additions shall be scanned for future reference.

   a) Scan datasheets.

   b) Save scanned datasheets in the folder designated by the Field Operations Manager using the following naming convention: "moduleAbbreviation_fds_SITE_YYYYMMDDx", where:

      i) 'fds' refers to field datasheet; lds should be used for lab datasheets

      ii) 'YYYYMMDD' is the most recent date on the datasheet;

      iii) 'x' is an optional character to be used in the event of multiple files (representing a-z). Multiple files can result from multiple datasheets needed to accommodate the quantity of data collected on a given date for a given module and/or a module requiring multiple distinct field (or lab) datasheets. If multiple files are not needed, there is no need to use the x.

      iv) Example: phe_fds_CPER_20130710b (this example is 22 characters long, which is less than the maximum length allowed by some models of scanners). See Table 2 for module abbreviations. These are also used in the file naming conventions described below.

c) If a domain elects to scan datasheets twice for maximum security (i.e., once before data entry and once after), the filenames should follow the convention above, with one of the suffixes '_orig' or '_final' appended to the name.

8.2 Best Practices for Data Entry

Best data entry practices shall be followed for all data transcription activities.

1) Set aside designated time for critical data entry as soon as possible after the field data collection.
2) Make sure your workspace is free of clutter and other distractions while entering data.
3) Prior to entering data, review all field datasheets for completeness, previous annotations, notes on the back, if all sheets are present, and visually scan all sheets to gain familiarity with the data.
4) It is preferable, although not required, to assign data entry to those who collected the field data or to individuals who work on the given protocol.
5) For technicians new to data entry, it is recommended that their work be double checked by a more experienced technician at regular intervals, or after each field datasheet is entered (if time allows).
6) Do not alter original data on the datasheet directly. All corrections and clarifications to a datasheet must be clearly identified and differentiated from original collection data by adding detailed notes in margins, on the backs of datasheets or on a separate and attached sheet. These notes should end up in the remarks field of the associated e-datasheet (but not in the webUI). Any deletion from the original datasheet at this point of the process should be indicated by a single line crossing out the error coupled with the date and initials of the corrector. All marks shall be made with red ink. All corrections must be transcribed; original errors on the datasheet do not need to be retained in an e-datasheet.
7) During transcription, check for invalid entries where a controlled vocabulary (i.e., lookup table) exists, especially where cell values are not formatted to be constrained. Controlled vocabularies can be found in lookup tables accompanying the data ingest workbooks and databases.
8) Do not alter the original data ingest workbooks or lookup tables, as these are change controlled documents. **Do not alter the data ingest databases, except to add data to them, as the functionality of these databases are dependent on the relationships defined in the design process. Deleting tables or fields can disrupt the functionality of validation rules, forms, and queries.** If problems, issues, or omissions with the workbooks and/or tables arise, create a problem ticket, assigned to the Staff Scientist responsible for the given protocol.
9) Make sure that data types are consistent within a field and with the data type specified in the data ingest workbook. For example, if the field contains numeric values, be sure there are no text values in any of the fields. **Never** use special characters (e.g., #, %, ~, *), even in the remarks (often referred to as comments) column.

10) Use semicolons rather than commas to separate items in a list, where relevant (e.g., in the remarks field). This prevents parsing errors once the data are saved into a comma-separated format (.csv). This guideline should only be relevant to the free-form remarks fields, as data fields should only contain one value per cell.

11) The remarks fields are free text fields and will be published on the NEON Data portal. Please use discretion and professionalism when entering comments in these fields as they will be made available to the public.

8.3 Transcribing Data: Preparation

a) Review the "QA/QC" and "Data Handling" sections of the protocol for which you are entering data.

b) Read and become familiar with the "FieldSummary" worksheet/table of the workbook/database or the relevant data ingest workbook, if entering data via a webUI. These provide descriptions and intended formats for each field, and indicate whether the field can or cannot be NULL and which fields have a controlled vocabulary found in a lookup table. Transcribing Data: E-Datasheets and Databases

1) Getting started

a) Open the appropriate data ingest workbook or database. These workbooks/databases are named according to the convention "protocol_dataIngest_20XX".

i) If using a database, start every new data entry session with a blank e-datasheet. That is, the 'in' tables should be clear of the data that were appended to the data storage tables at the end of the previous session.

ii) If using a workbook, add the new data to the previously entered data in the next available rows of each data entry worksheet, to produce a running table – until instructed by the Field Operations Manager or Staff Scientist to export the data to the FOPS dropbox.

b) Save a copy of the data ingest workbook or the database, in the domain's network directory that is designated by the Field Operations Manager (i.e., a directory that is distinct from the network location of the original version of the data ingest workbooks and databases). If any changes to the workbooks or databases are made by TOS staff due to iterations in JIRA with FOPS, those changes will be made in the master database and communicated to FOPS so they can be sure to download the most recent version of the database.

c) Lookup tables are indicated by an 'L' at the start of the worksheet or table name.

d) Choose the appropriate tab, table, or form (Figure 3) for the current datasheet. **All worksheets or tables that are intended for technician data entry are indicated by an '_in' at the end of the worksheet or table name** (Figure 4).

2) Entering data

a) The format for all date fields in e-datasheets shall be YYYY-MM-DD; for example, use 2013-07-10 for July 10th 2013. Note that entering the hyphens is unnecessary in the data ingest databases, as this should happen by default.

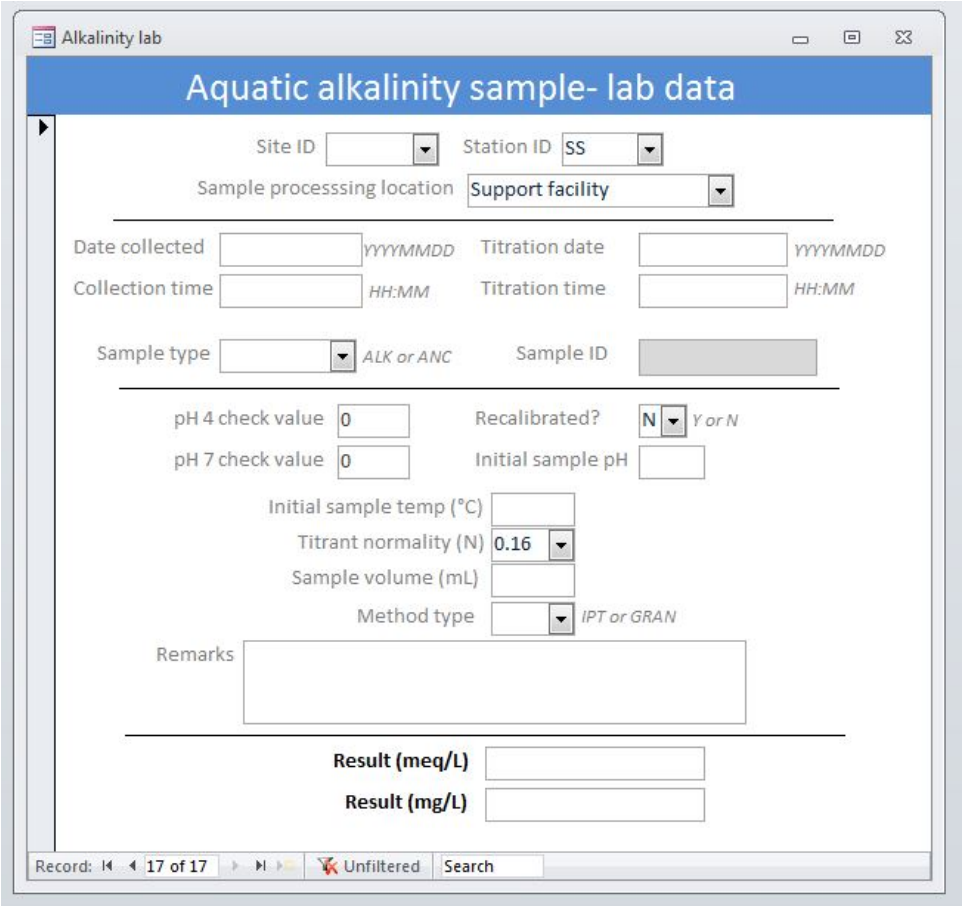

Figure 3. Example of an MS Access database form for data entry.

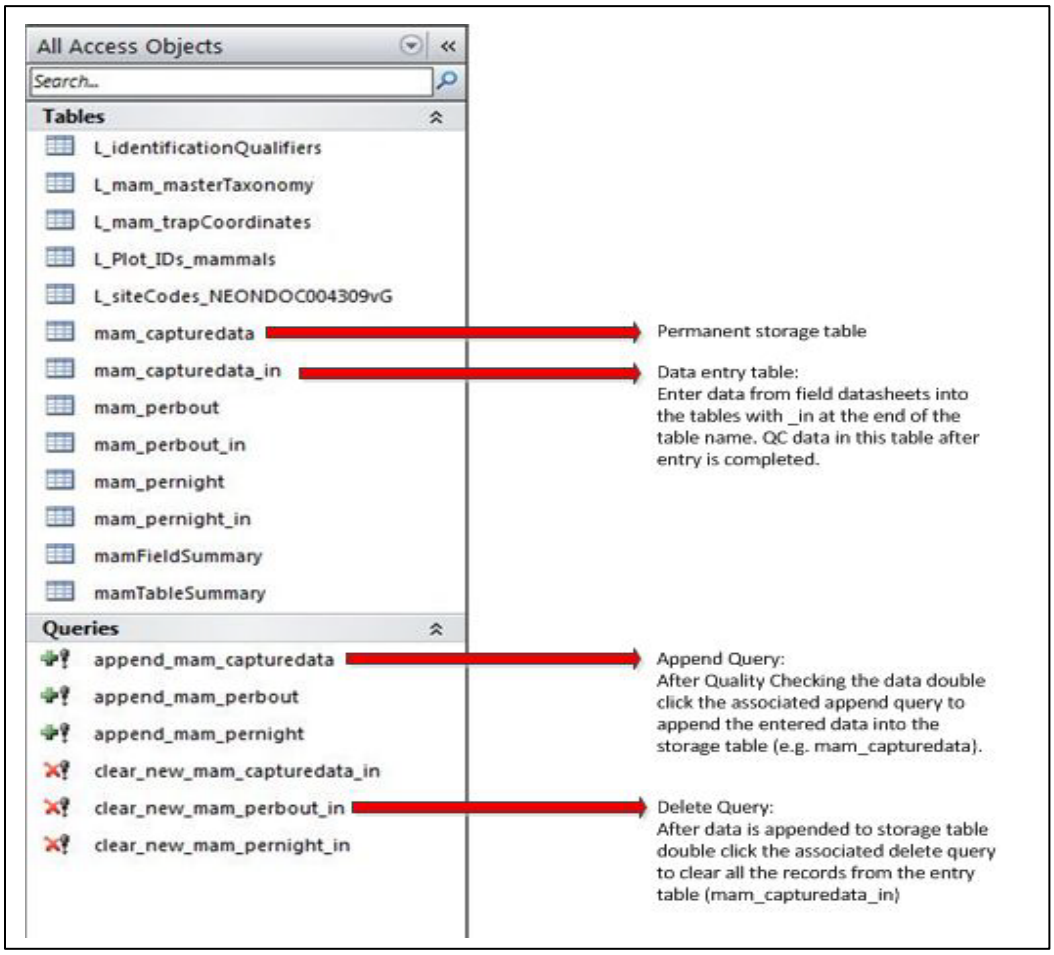

Figure 4. Example data ingest database structure showing the availability of (1) lookup tables (L_tableName); (2) tables intended for data entry (tableName_in); (3) tables intended for longer term data storage (tableName), until exported to the FOPS dropbox for ingest by CYI; and (4) queries for appending and deleting data.

b) Save workbooks frequently during transcription to prevent loss of transcribed data. Note that data entry in Access does not require manual saving, as it is done automatically.
   i) If using Excel workbooks: Whether appending to an existing workbook or creating a new workbook or worksheet, save each data entry bout with the domain number (DXX) and, if useful, date of data entry appended to the original file or worksheet name (e.g., originalName_D10_20130730).

c) For organismal modules that include an abbreviated version of the identification qualifier (i.e., idQ or idqCode) on the paper datasheet, be sure to enter the unabbreviated version of the code into the field (i.e., enter 'cf. species', not just 'CS'). The lookup table defining all of the abbreviations is included in the relevant data ingest databases.

d) If no data exist for a given cell, leave the cell blank.

e) Tips for entering data into MS Access:

i) Ctrl + adds a new record to a table

ii) Ctrl ‘ auto-populates the next record's cell with the value in the cell immediately above

iii) Lookup tables are frequently used to provide lists of possible values to be entered into a field. These values can be either typed into the cell or selected from the drop-down menu that appears when you click in a cell. When the full list is already known (e.g., taxonIDs), only values that match one of the entries in the list can be entered. When the full list is not yet known (e.g., plotIDs), entries are not constrained to only the available values.

(1) Issue a problem ticket if you think a valid option for data entry does not exist in the lookup table.

(2) For Vegetation Structure and Plant Diversity, the domain specific plant species lookup tables have OTHER as an entry option. Use this entry **ONLY** if an identified plant species is not included in the lookup table. In the remarks column for that record include the scientific name and description of the new species found and issue a problem ticket. Please refer to section 8.10 below for additional information.

At the end of each bout of data transcription, verify that all blank cells are intentionally blank, replace commas with semicolons, and remove any special characters. If special characters have been used to convey critical information, please use the remarks field to convey this information, as opposed to leaving a special character in a data field.

8.4 Transcribing Data: WebUIs

1) Getting started

a) Be sure you are connected to the NEON network (VPN access is fine). Use firefox, internet explorer (IE10 or higher), or Chrome as your browser. The use of Safari or lower versions of internet explorer (IE9) is not supported.or internet explorer as your browser.

b) Using your browser’s back button to navigate screens is strongly discouraged. Instead, use the ‘cancel’ button at the bottom of each page.

c) Training portal: http://**cert**-as-web-1.ci.neoninternal.org:8080/web/som-portal/home

i) The certification (cert) portal provides all of the data entry applications for the purposes of training only. You are free to use this portal to enter test data for training purposes, but be aware that these data will not be stored in the NEON CYI database.

d) Production portal: http://**prod**-as-web-1.ci.neoninternal.org:8080/web/somportal/home

i) The production portal provides the data entry applications that must be used for actual data entry. These data will be stored in the NEON CYI database and published on the data portal.

e) Protocol-specific instructions are found on SharePoint – in the Document Collaboration/FSU-FOPs/Web UI Training Materials folder.

f) Logging in: You must use a neoninc.org or field-ops.org email address to create your account. You need to create a separate account for each portal (CERT, PROD). You can use the same username and password on all, but you will need to create the login for each portal separately. Once you have created an account, you can obtain additional permissons (e.g., to enter data) on your account as needed by emailing one of the server administrators listed on the home page of the portal.

2) Entering data: General guidelines

a) Data entry through the webUI is a two-step process. The '**Record**' data button will save the data locally (through your browser). The webUI will allow you to review and edit data you have 'recorded' prior to submission. The '**Submit**' data button will send the data you have entered to NEON's permanent data repository (PDR). You cannot edit or update the data you have submitted to the database. Check your data carefully and ask any questions prior to submitting.

b) You will be automatically logged out after a few minutes of inactivity. Partially entered data that has been 'recorded' will be saved locally but not submitted to the data repository. When you log back in, you can continue where you left off, and the partial plot data previously entered will be available for review.

c) For some webUIs, you cannot move on to the next plot until you have submitted all of the data for a given plot. Other webUIs allow the entry of multiple plots before a final submission. However, it is highly suggested for all webUIs that data are entered in a plot, date, or bout level increment, whichever is appropriate for an individual webUI.

d) The current versions of the webUIs do not support linking to photos. You can ignore any information about "Camera ID" and "Photo ID". You should use one of the remarks fields to record photograph-related data.

3) Entering data: Location, Date, and Personnel Page

a) The data entered on this initial screen will typically be used to autopopulate these fields in all records for a plot. However, some of the fields (e.g., recordedBy) can be updated on a record by record basis, if needed.

b) Plot Selection

i) Filtered drop-down menus to show only sites that are valid for a given domain, and plots that are valid for a given protocol in a given site.

ii) The PlotID may contain some extraneous information at the end of the name (example: STER_007.baseplot.div, for STER_007 -- ignore the '.baseplot.div' component of the name).

iii) If the plot you are looking for does NOT appear on the dropdown, use issue tracking software to notify the relevant staff scientist.

c) Personnel

i) You cannot change Entered By. If a new person takes over data entry, you will need to log out and have them log in. However, they will not be able to view your partially entered plot data. It is encouraged to enter data at the per plot, grid, or sampling event level depending on the protocol.

ii) Email addresses of personnel should be entered into personnel fields (e.g., measuredBy, recordedBy). The current versions of the webUI does not confirm that email addresses are valid, so enter carefully. Your browser should provide a drop-down list once values have been entered.

d) The summary button on this page (e.g., 'Record XX data for specified date and plot') does NOT send the data to the database, just saves the information you have entered locally and allows you to proceed to the next screen.

4) Entering data: Review the data prior to submission

a) Once all of the data for a plot have been entered, select 'Done entering data for this plot, review and submit the data'.

b) Download the entered data as a .csv (button at bottom of page) to review your entries. All webUIs should provide the options to download a csv for review before submission of data to the database.

c) Review the entered data according to the instructions in section 8.6 below, Quality Checking 10% of Data.

d) If you find any errors in the data, select 'Cancel' to return to the plot summary page and fix all errors through the webUI. Be sure to review the data **prior to submission** to confirm that all errors were successfully addressed.

8.5 Quality Checking 10% of Data

1) After all of the data from one sampling period has been entered into the e-datasheets or webUI, field technicians shall quality check at least 10% of the data in each worksheet by comparing it to the original paper datasheet.

2) Records to be reviewed for accuracy should represent a random selection of the data. For a given worksheet or table, check every 10th record. If the worksheet has fewer than 10 rows of data, review at the least the first and last rows. If checking a csv that was generated from a webUI, please check all metadata in the first few rows of the csv before performing the 10% check as outlined in this section.

3) Check selected records against the original paper datasheet.

    a) It is recommended that checks be done by two technicians, one reading the values on the original paper sheet out loud, while the other checks those against the electronically entered values.

    b) For data ingest workbooks and databases only: Keep a record in the remarks column of all errors as you find and fix them. If you check a record and find no errors, enter 'no transcription errors found' into the remarks column to facilitate review and error tracking. Every worksheet/table should have a remarks column; issue a problem ticket if this column is missing.

    c) Calculate an error rate by counting the number of cells checked (number of rows checked x cells per row) and dividing that number into the count of cells with errors found to yield a percentage of error. Access databases include an errorRateTracking table to facilitate these calculations and tracking of error rates (Figure 5).

    d) If a greater than 5% error rate is found during quality checking, an additional 10% of the data in that sheet shall be checked in the same manner. If the combined error rate between the two checks is greater than 5%, all of the data in that sheet needs to be reviewed and fixed.

    e) If the original check reveals an error rate greater than 10%, it shall elicit a 100% review of the full dataset (i.e., all cells in all records checked).

4) Once a calendar year, calculate an error rate by counting the number of cells checked (number of rows checked x cells per row) and dividing that number into the count of cells with errors found to yield a percentage of error. Access databases include an annualErrorRate query that summarizes all data entered into the errorRateTracking table (Figure 5). Report the protocol-specific error rates on an annual basis using a spreadsheet on the NEON Intranet (FSU-FOPS/dataEntry2015 folder).

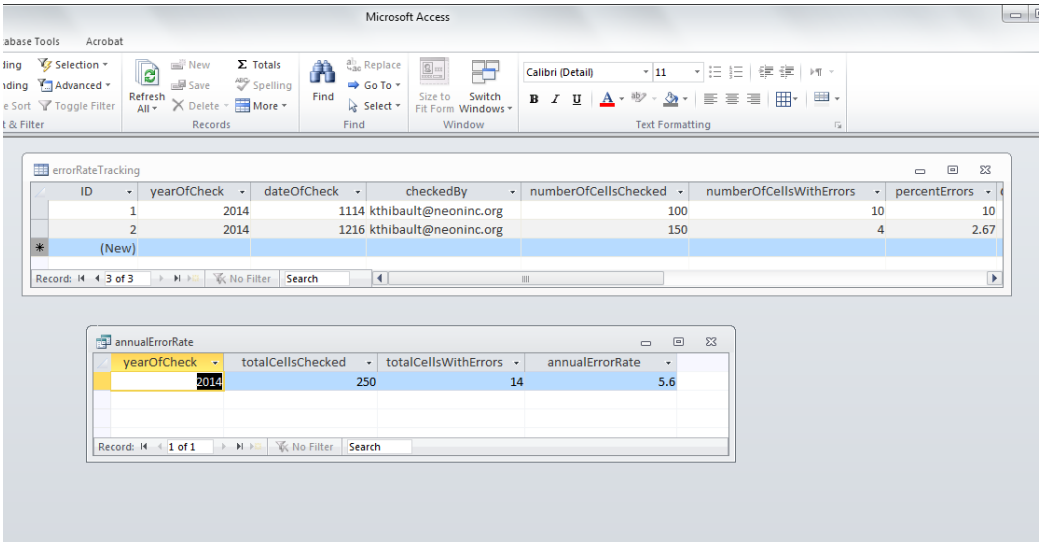

Figure 5. Example of an error rate tracking table and annual error rate query in an MS Access database.

8.6 Transcribing Data: Final Steps

8.6.1 Appending and Clearing Data in a Data Ingest Database

For data ingest databases, queries that append newly entered and (ideally) quality-checked data in the entry tables (tableName_in) to the data storage table (tableName) are provided, using the naming convention 'append_tableName'. Double-clicking on one of these queries will generate a warning message that 'You are about to run an append query that will modify data in your table. Are you sure you want to run this type of action query?' Select yes, if you are ready to append the data.

Once data have been successfully appended from the tableName_in table to the tableName table, delete the data in the tableName_in table. Delete queries are also included in the data ingest databases that will clear all of the data from the tableName_in table; these queries are named using the convention 'clear_new_tableName_in'. A similar warning prompt will appear upon double-clicking, as in the append queries, which will require a selection of 'Yes' to run the query. Once a query of this type is run, the tableName_in table will be empty and ready for the next bout of data transcription.

8.6.2 Submitting the Data on the WebUI

1) After reviewing and correcting 10% of the data, as outlined in section 8.6, confirm that all errors were successfully addressed. Then, select 'Submit' data to the CYI database. You will then see the message: 'This plot data was successfully posted to the database!'

2) Once you hit Submit, you cannot undo this. No errors in the data can be fixed, once the data have been submitted (except for the plant phenology webUI).

a) Please do not hit Submit more than once, or intentionally re-enter a record as a method of data correction. These actions will generate duplicate records that the post-processing algorithms may not handle properly.

b) An email will be sent if the database submission is unsuccessful; this is the only scenario in which the Submit button should be hit again for the same set of records.

3) If you realize after submission that you have made a major error in transcribing the data:

a) You must note it here: https://docs.google.com/forms/d/18KkxkXb-05mmZb97fezXqUcyPpcA465VgdYN2BiiDWw/viewform

b) Although errors cannot be fixed at this time, a log of known errors is required. These errors will hopefully be corrected once such functionality is implemented.

c) Minor errors (such as failing to transcribe remarks, not adding 'identification references') you do not need to report.

4) After submitting the data, record the site and bout data entered by protocol in the provided spreadsheet on the NEON Intranet.

8.7 Saving Entered Data

After completing quality checking, the data are ready to be saved in .csv format, either from Excel, Access, or the webUI. Save the csv files from Access and Excel only in \\10.100.128.37\dropbox, and consult with your Field Operations Manager on where csvs (including those generated from the webUIs) will be stored locally.

8.7.1 File Naming Convention

For e-datasheets and databases, file naming should follow the convention used to name the worksheets and tables, with the addition of domain number (DXX) to the end of the file name (e.g., mam_capturedata_D01).

For webUIs:

- The default naming convention for the csvs downloaded from the webUI is: ProtocolNamePlotData_usernameatneoninc.org_YYYY-MM-DD-HH-MM.csv, where times are recorded in UTC. The phenology webUI is the only exception to this; there is no need to rename the files downloaded from the phenology webUI

- Rename the file to the general format: plotID_yyyymmdd_moduleAbbreviations.csv (example CPER_001_20140829_div.csv), where date is the latest date of the included field data (not when they were entered). Module abbreviations are found in Table 2. Additional specifications can be added to file names (e.g., sorting), if it increases data discovery by FOPS personnel.

8.7.2 Exporting from E-Datasheets and Databases

All data should be added to a running table or spreadsheet, with all previously entered data retained. Export all data to the FOPs dropbox, after each bout of data entry. Exports do not need to occur more often than once a week.

- In Excel, make sure the worksheet you want to export is the active sheet. Go to File – Save As – and select CSV (Comma Delimited) from the dropdown menu.
- In Access, highlight the tableName you want to export in the Navigation Pane. Click on External Data – Export – Text File, and complete the Export Text Wizard.
    - Wizard Window 1: Do NOT specify any export options; just specify a file path and name. Be sure to change the file extension from .txt to .csv
    - Wizard Window 2: Use default selection of 'Delimited'
    - Wizard Window 3: Select comma-delimited and check the 'Include Field Names on First Row' box (Figure 6).
    - Wizard Window 4: Select Finish!

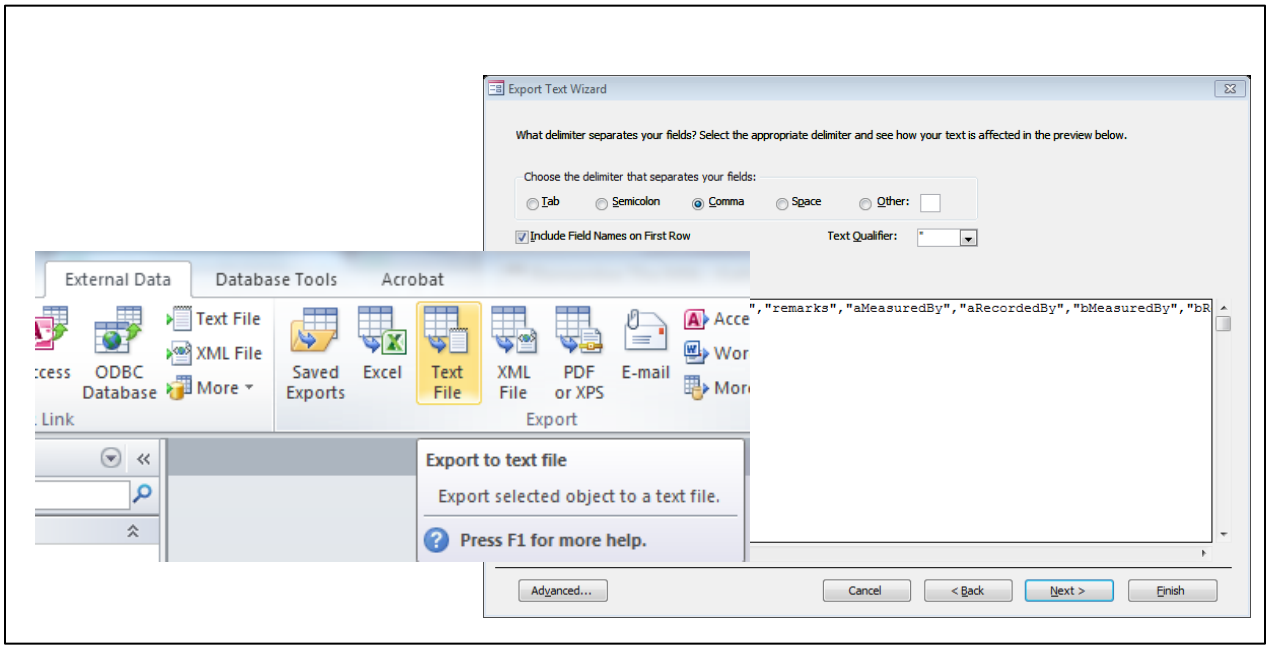

Figure 6. Screenshots of the export process in MS Access.

8.7.3 Downloading from WebUIs

After submission, re-download the provided .csv of the submitted data (replacing the original download with the corrected version, if applicable), rename file as directed above, and save the file to the locations specified by the Field Operations Manager.

8.8 Final Processing and Storage of Original Field and Lab Datasheets

1) Following completion of data transcription, stamp or write date entered and name of transcriber onto each fully transcribed paper datasheet.

2) Completed original datasheets shall be stored in a designated file system at each domain.

3) Original datasheets shall be mailed to the appropriate Assistant Director at NEON headquarters at least once a year for archiving.

Table 2. Module abbreviations used in worksheet and table naming.

Subsystem	Abbreviation	Module
AOS	AFM	aquatic field metadata
AOS	ALG	algae, periphyton, phytoplankton
AOS	AMC	aquatic microbes
AOS	APL	aquatic plants, bryophytes, lichens, and macroalgae
AOS	ASC	aquatic sediment chemistry
AOS	ASI	aquatic stable isotope
AOS	BAT	lake/river bathymetry
TOS	BET	beetles
TOS	BGC	biogeochemistry
TOS	BRD	birds
TOS	BRY	bryophyte production
TOS	CDW	coarse downed wood
AOS	DEP	depth and secchi profiles
AOS	DGA	dissolved gas
TOS	DHP	digital hemispherical photos for leaf area index
TOS	DIV	plant diversity
AOS	DSC	stream discharge
AOS	FSH	fish
AOS	GWC	groundwater water chemistry
TOS	HBP	herbaceous productivity
AOS	INV	aquatic macroinvertebrates
TOS	LPS	camera-quadrat for leaf area index per species
TOS	LTR	litter and fine woody debris
AOS	LWC	lake water chemistry
TOS	MAM	mammal abundance and diversity
TIS	MGP	soil megapit

Subsystem	Abbreviation	Module
TOS	MOS	mosquitoes
TOS	MPT	mosquito pathogens
TOS	PHE	plant phenology
AOS	REA	stream reaeration
AOS	RPL	riparian plants
TOS	RPT	rodent pathogens
TOS	SLS	soil sampling
TOS	SME	soil microbes
AOS	SMR	stream morphology
AOS	SWC	stream water chemistry
TOS	TCK	ticks
TOS	VST	vegetation structure
AOS	ZOO	zooplankton

8.9 NEON Guidelines for Specific Data Fields

- Identification References:
    - a free text field to capture the key and/or field guides used to identify species during a bout
    - Enter information only if you used a reference to identify a particular species (for all protocols except small mammals).
    - Copy and paste the reference title from the protocol-specific list on SharePoint (https://neoninc.sharepoint.com/sites/warehouse/FSUFOPS/Forms/AllItems.aspx) to ensure consistent formatting. In the case of multiple sources, use semicolons between different sources.
- Entering a taxon:
    - **Type** any part of the Scientific Name or taxonID code into the box and a list will appear.
        - **Format** of viewable dropdown on webUI= “Scientific Name Authorship (taxonID/acceptedTaxonID)”.
            - TaxonID can include synonyms. In the case of multiple available synonyms, use the name that matches your field key, if possible.
            - AcceptedTaxonID is the current accepted taxon code for each taxon; it is used to map synonyms to accepted taxonomy, where appropriate.
        - **Example**: Triticum macha Dekap. & Menab.(TRMA11/TRAE)
            - Scientific Name = ‘Triticum macha’
            - Authorship = ‘Dekap & Menab.’)

- taxonID =TRMA11
- acceptedTaxonID = TRAE

- **Taxon codes**:
  - NEON taxon codes can be found on Sharepoint: FSU-FOPS/taxonTables
  - NEON plant taxon codes follow USDA plants

- Entering an unknown species**:**
  - Drop-down menus of taxa provided in data ingest databases and webUIs include **unknown codes** for higher taxa (e.g., family sp., genus spp., etc.), when you cannot resolve the taxonomy to the species-level.
  - There are two options for most unknowns -- a 'sp.' and 'spp.' variant
  - When to use sp. vs spp.:
    - Use **sp.** when you think there is only ONE species of the unknown in the list you are populating (such as plant diversity Plot 38, subplot 31.1.1, for a given sampling bout).
    - Use **spp**. for lumping (i.e., when you suspect there is possibly more than one species present). Note that this option is not available for small mammals, since they are identified at the individual level.
- Entering a taxon that is not included in the list:
  - Drop-down menus of taxa provided in data ingest databases and webUIs are based on species lists populated based on available data on geographic range. They may not be complete.
  - If you have identified a taxon (to any rank), and that species (or genus sp.) does not appear in your dropdown list:
    - Check the NEON taxon table or USDA plants to make sure you are using a known name/code), and it's not a spelling problem.
    - If this doesn't resolve it:
      - Select Other - '[Kingdom/Class] Sp. OTHE/OTHE'
      - Issue a JIRA ticket (to protocol staff scientist) that the list needs to be expanded
      - Type ScientificName and taxonID, whichever (or both) you know, in taxonRemarks (if available) or remarks
      - Use sp. and spp. as they would appear on the dropdown (e.g., 'scientificName = Asteraceae sp.')
      - Using the exact syntax provided here (and correct spelling) is highly desirable: 'scientificName=Xxx xxxx, taxonID = XXXX'

- Please note that the Other option is distinct from an unknown target taxon (e.g., the 2PLANT option for plants you cannot identify to family or lower resolution). Other indicates that you do know what the taxon is, but it is not available in the drop-down list.

- Entering Morphospecies:
  - Morphospecies that have been identified (i.e., taxonomy resolved to the best of your ability) prior to data entry do NOT need to get entered as morphospecies.
  - You can increase traceability by noting the original morphospecies id in the remarks field (NOT the morphospecies field).
  - A taxon code is required for some protocols (e.g., plants). If you can make a reasonable educated guess from the morphospecies name, choose the lowest rank taxon code that is appropriate.
    - For example, if the morphospecies name is 'yellow mustard 7', you could select 'Brassicaceae sp'.
  - If you have absolutely no idea what the genus or family is for the morphospecies, select the most relevant unknown option (e.g., '2Plant -- Unknown Plant').
  - If you enter a morphospecies into the webUI so that it ends up in PDR (and will therefore live on for the life of NEON), enter it into the master tracking list on Sharepoint (FSU-FOPs/morphospeciesTracking).

9 REFERENCES

N.J. Van Lanen, C.M. White, J.A. Fogg, M. F. McLaren. 2012. Integrated Monitoring In Bird Conservation Regions (IMBCR): Data Entry Protocol. Unpublished report. Rocky Mountain Bird Observatory, Brighton, CO, USA. URL: rmbo.org/v3/Portals/0/Documents/Science/Protocols/

APPENDIX A QUICK REFERENCES

Complete all steps within 7 days
No later than **14 days** after data collection/end of bout.

Quick Reference: Data Transcription Workflow

1. Review field datasheet for completeness, previous annotations, notes on the back, if all sheets are present, and visually scan all sheets to gain familiarity with the data.
   - Make needed corrections/additions in red ink.
2. Scan field datasheet. Move scans to N:\ drive: DEPT > FOPS > DXX > Data > Scanned_field_datasheets.
3. Transcribe data into Excel workbook, Access database, or webUI.
   - Review the "QA/QC" and "Data Handling" sections of the protocol.
   - Review the "FieldSummary" worksheet/table of the workbook/database or the relevant data ingest workbook, if entering data via a webUI.
4. Perform quality assurance check

   If no data exist for a given cell, leave the cell blank.
   - Compare 10% of the data to recheck against the hand-written field datasheet.
   - Calculate an error rate (the count of cells with errors found/the number of cells checked (number of rows checked x cells per row)*100 = percentage of error.
   - If error rate is >5%, review 10% more of the data. If the combined error rate between the two checks is > 5%, all of the data should be reviewed.
   - If the original check reveals an error rate > 10%, all of the data should be reviewed.
5. Make any additional corrections to the datasheet in red ink.
6. Stamp field datasheet.
7. Export csv of data to the FOPS dropbox (\\10.100.128.37\dropbox; if Access or Excel-based) or to a local domain folder (for webUIs).
8. Store field datasheet.
9. Once a year, calculate an annual error rate for each protocol and report it via issue tracking software.

MS Access – Quick Tips

- Ctrl + adds a new record to a table
- Ctrl ' auto-populates the next record's cell with the value in the cell immediately above
- Access does not require manual saving, as it is done automatically
- Enter data into the entry tables (tableName_in)
- After QC, run append query (append_tableName) to add the new data to the storage table (tableName)
- Then clear the data from the entry table (tableName_in) using the delete query (clear_new_tableName_in)

WebUI – Quick Tips

- Use IE10+, Firefox, or Chrome.
- Use the Cancel buttons - NOT the browser back button.
- Training portal: http://**cert**-as-web-1.ci.neoninternal.org:8080/web/som-portal/home
- Production portal: http://**prod**-as-web-1.ci.neoninternal.org:8080/web/somportal/home
- Protocol-specific instructions are found on SharePoint – in the Document Collaboration/FSU-FOPs/Web UI Training Materials folder
- Download the entered data as a .csv to review your entries BEFORE you submit the data to the database.
- Once you hit Submit, you cannot undo this. For most webUIs, no errors in the data can be fixed, once the data have been submitted.
- If you realize after submission that you have made a major error in transcribing the data: note it here.

Please use discretion and professionalism when entering comments in the remarks fields, as they will be made available to the public.