AOS/TOS PROTOCOL AND PROCEDURE: MANUAL DATA TRANSCRIPTION

PREPARED BY	ORGANIZATION	DATE
Cody Flagg	FSU	10/26/2015
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
Andrea Thorpe	SCI	04/01/2016
Vlad Aleksiev	PSE	03/14/2016

RELEASED BY	ORGANIZATION	RELEASE DATE
Judy Salazar	CM	04/04/2016

See configuration management system for approval history.

Change Record

REVISION	DATE	ECO #	DESCRIPTION OF CHANGE
A	05/30/2014	ECO-01835	Initial release
B	01/05/2015	ECO-02633	Added information regarding recent developments of web interfaces and mobile applications for data entry
C	04/04/2016	ECO-03640	• Updated formatting to latest Protocol/SOP template (previous version did not have SOP sections) • Added more text to Background section • Added clarifying steps for data entry QA procedures • Updated acronyms in table 2 • Section 8.2 – updates to best practices section

Title: TOS Protocol and Procedure: Manual Data Transcription		*Date*: 04/04/2016
NEON Doc. #: NEON.DOC.001271	*Author*: C. Flagg	*Revision:* C

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

Title: TOS Protocol and Procedure: Manual Data Transcription		Date: 04/04/2016
NEON Doc. #: NEON.DOC.001271	Author: C. Flagg	Revision: C

1   OVERVIEW

1.1   Background

This document is a cross-protocol set of instructions for data transcription and storage when field data are recorded with paper datasheets. 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 quality assurance methods that will facilitate identification of errors prior to the automated portion of the QA/QC algorithms.

Where mobile data recorder (MDR) 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 (DSF) to ultimately be ingested by NEON Cyber Infrastructure (CYI). Manual data entry is a key step in the data collection process as it is the first point at which errors can be identified and corrected. Manually entering data also poses the risk of personnel introducing errors into a dataset. In almost all cases, it is more efficient and cheaper to detect and correct errors before data are submitted to automated processing steps (Chapman 2005); thus it is vital that all personnel involved in the data collection process familiarize themselves with both the core concepts of data quality and details of data entry.

The aim of quality control during data entry is to prevent the introduction of errors throughout all stages of data collection, entry, and storage. Early detection and correction of errors is a critical preventative step in obtaining high quality data. It is important that personnel understand the data quality process and their role in it at NEON. The data quality process consists of six sequential phases (Figure 1): 1) establishing an overall data collection design with concrete requirements, 2) writing standard operating procedures and organizing effective training for data acquisition, 3) defining rules and methods for controlling data quality, 4) developing processes and algorithms for assessing and assuring data quality, 5) building database systems for data archival, and 6) assembling and delivering useful data products with correct meta-data. Field Operations (FOPS) personnel are directly involved in the data acquisition (phase 2), data quality control (phase 3), and data quality assurance aspects of the data quality process through such activities as training seasonal technicians, rigorously recording field observations, correcting errors on field data sheets, accurately entering data, and systematically reviewing data prior to submission. Data that are submitted with errors ultimately degrades both the effort exerted in the

field and the ability of the Observatory to deliver high quality data products to the scientific community and public.
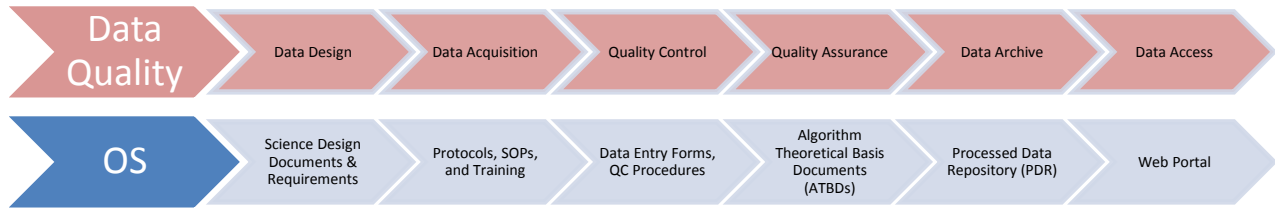
Figure 1 Conceptual model of the data quality process. The top half represents the general data quality process for many science projects; the bottom half represents the specific aspects of the data quality process as implemented by OS at NEON (adapted from J.W. Brunt 2000).

1.2 Scope

This document provides a change-controlled version of Observatory protocols and procedures. 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.

1.3 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.

N/A

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. 002652	NEON Level 1, Level 2 and Level 3 Data Products Catalog
RD [05]	NEON.DOC.014051	Field Audit Plan

2.3   Acronyms

ATBD	Algorithm Theoretical Basis Document
AOS	Aquatic Observation System
DE	Data entry
DEF	Data entry form
DQC	Data quality checking
FOPS	Field Operations
MDR	Mobile data recorder
MS	Microsoft
QA	Quality assurance
QC	Quality control
QC10	Quality control 10%
TOS	Terrestrial Observation System

2.4   Definitions

Common terms used throughout this document are defined here, in alphabetical order.

Data Entry Form (DEF): interface used for entering data, encompasses webUI, Access database, or MDR data entry systems.

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).

Local User Interface (LUI) – A database software system that contains data tables and data entry forms for controlled data transcription for FOPs-collected data, temporary data storage, and associated lookup tables. This software includes Microsoft Access and/or the NEON Intranet databases.

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

NEON Intranet: a generic reference to the NEON Intranet pages that house links to training materials, instructions, protocols, datasheets, and other documents that are shared among FOPS, TOS, and AOS.

WebUI: web user interface; a protocol-specific user interface accessible within the NEON network with internet web browser software; 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).

Mobile Data Recorder (MDR): handheld, field-portable equipment 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   METHOD

The strategy for manual data entry during NEON operations consists of four general steps. 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 annotations should be made on the raw datasheets in red ink, with associated date of corrections and initials of technician. This step is crucial in rectifying introduced errors; for example, incorrectly identified plant species/sub-species may be quickly spotted by a botanist intimately familiar with a site's taxa, but are extremely difficult to detect with automated processing software. After this initial review, the field datasheets should be scanned, and the resulting files saved in the designated network location. The second step is entering the data into the data entry forms (DEF), either databases or the web user interfaces (webUIs), where available (Table 1). In the third step, technicians shall perform a manual quality assurance check on ten percent of the entered data, at a minimum, to ensure that high quality data are being submitted and best practices are being followed; the person checking the data should be different than the person who entered the data While the volume of data collected by NEON Field Operations prohibits manual checks of the majority of the data records, manual data checking can detect systematic errors that are not easily captured by the ATBDs The ten percent threshold reflects the trade-off between using available labor to collect more data and proofing all of the data that are collected, and can increase if this threshold is exceeded upon review Data quality checks are recorded in a central spreadsheet on the NEON Intranet. The fourth and final part of the manual data entry process is properly saving, submitting, and storing files as defined by the protocol's DEF interface (i.e. webUI, Access database etc.)

Data entry forms (DEF) contain controlled values and vocabularies; the tables containing controlled values are referred to as lookup tables and constrain values that can appear in certain fields. Before entering data into any DEF, be sure to review the required data and the constrained values associated with those fields 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.

Table 1. Data entry formats and options by NEON protocol, as of February 2016.

Sub-system	Short module description	Protocol document number	Datasheet document number	Data entry file format
TOS	Vegetation structure	000987	001573	accdb
TOS	Herbaceous biomass	014037	001574	webUI
TOS	Leaf area index	014039	001575	accdb
TOS	Soil measurements	014048	001577	accdb
TOS	Plant phenology	014040	001578	webUI
TOS	Plant diversity	014042	001579	webUI
TOS	Ground beetle sampling	000451	001580	webUI
TOS	Mosquito sampling	014049	001581	accdb
TOS	Tick sampling	014045	001583	accdb
TOS	Small mammal sampling	000481	001585	webUI
TOS	Rodent- borne Pathogen sampling	014044	001402	accdb
TOS	Litter and fine woody debris	001710	002132	accdb
TOS	Bryophyte Productivity	001709	002136	accdb
TOS	Coarse downed wood	001711	002121	accdb
TOS	Aboveground Plant Biomass in Agricultural Systems	003280		accdb
TOS	Cactus Biomass and Handling			accdb
AOS	Aquatic Field Metadata	001646	001646	accdb
AOS	Aquatic Microbial Sampling	003044	003041	accdb
AOS	Periphyton, Seston, and Phytoplankton Sampling	003045	003042	accdb
AOS	Aquatic Plant, Bryophyte, Lichen, and Macroalgae Sampling	003039	003040	accdb
AOS	Aquatic Macroinvertebrate Sampling	003046	003043	accdb

AOS	Zooplankton Sampling in Lakes	001194	002302	accdb
AOS	Surface and Groundwater Chemistry	002905	002906	webUI
AOS	Surface Water Dissolved Gas	001199	002384	accdb
AOS	Aquatic Stable Isotopes	001886	002211	accdb
AOS	Stream Reaeration	000693	002382	accdb
AOS	Stream Discharge	001085		webUI
AOS	Sediment Chemistry Sampling in Lakes and Non-Wadeable Streams	001191	002435	accdb
AOS	Sediment Chemistry Sampling in Streams	001193	002419	accdb
AOS	Riparian Habitat Assessment in Lakes and Non-Wadeable Streams	001195	002764	accdb
AOS	Riparian Habitat Assessment in Wadeable Streams	001196	003156	accdb
AOS	Bathymetry and Morphology of Lakes and Non-Wadeable Streams	001197	003104	accdb
AOS	Fish Sampling in Wadeable Streams	001295	003107	accdb
AOS	Fish Sampling in Lakes	001296	003106	accdb

Once data are entered into the appropriate data entry form, the data should be exported to a csv file. Domain technicians shall save entered data in an initial staging area on the network, hereafter referred to as the "FOPS dropbox" (hyperlink available on the The NEON Intranet page), for subsequent retrieval by Science and CYI teams().

AOS data shall be entered into the appropriate data entry forms (webUI, accdb, etc.) as appropriate. Data shall be entered into the webUI or MDR, if available. If a webUI or MDR is not yet available, data should be entered into the AQU Database. In a very limited number of protocols, no date entry form is available in the AQU Database and data should be uploaded to the AOS dropbox until the AQU database is updated to include these data entry forms. ([\\10.100.128.37\dropbox](\\10.100.128.37\dropbox) AOS).

4 SCHEDULE

4.1 Data Entry

Transcribe data into data entry forms as soon as possible after field data collection, as technician recall will help to improve data quality as datasheets are reviewed. As a best practice, the data should be entered within 14 days of collection or the end of a sampling bout whenever applicable. The following tasks should be prioritized, in descending order, if field staff are pressed for time and cannot complete data entry within 14 days of collection:

- Field datasheets should be checked by lead technicians on a daily basis.

- Scan and store field datasheets on a weekly or bout-level interval.

- Manually enter data starting with the earliest collection date.

- Prioritize manual data entry for protocols that have a webUI; data quality review must occur during the same webUI entry 'session'.

- Every data entry bout should include a data quality review.

4.2 Data Quality Checking (DQC)

Data quality checking should occur during a field season as data are entered. Time must be devoted to this activity, erroneous data submission essentially negates field and lab efforts if the entered data are of poor quality**.** Erroneous data that are submitted through the webUI currently ***cannot be corrected after submission.***

Data quality checking should occur on a minimum of 10% of the records per data entry bout. The definition of a data entry bout varies by the user interface:

- **webUI**: 10% of data records need to be reviewed prior to final submission, as records **cannot** be corrected once they are in the database. Thus 10% of records from a datasheet or set of related datasheets should be quality checked prior to submission.
- **LUI:** 10% of data records should be reviewed by the end of the data entry session related to the sampling bout. Since records entered into a LUI can be edited after submission, data quality checking can occur at any time throughout the data entry bout (not just before submission).

5 PERSONNEL

As a best practice, the technicians entering data should be those that collected the data for the specific protocol in the field. When this is not possible technicians entering data should familiarize themselves with a protocol's datasheets for which they are entering data. Lead technicians should take a guiding role in directing data entry, data quality checking (DQC) tasks, and recording/reporting information from DQC tasks.

5.1 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.. The technician-in-training should then repeat data transcription and have the entered data reviewed again until the 5% threshold is met.

6 STANDARD DATA TRANSCRIPTION PROCEDURES

The tasks associated with entering and reviewing entered data is broken up into a series of separate SOPs. The overall workflow, from start to finish, of the data entry process is visualized in Figure 2.

- **SOP A: Checking and Scanning Field Data Sheets.** Tasks for reviewing, annotating, and correcting data field data sheets.

- **SOP B: Data Entry Procedures.** General guidelines for entering field data into webUI and/or Access database systems. Interface specific instructions can be found on the the NEON Intranet page in the training materials folder.

- **SOP C: Data Quality Checking.** Procedures for checking entered data quality through careful comparisons with field datasheets.

- **SOP D: Best Practices for Data Entry and Data Quality Management**. Suggested best practices and strategies for managing data entry throughout a field season including assigning tasks, tracking progress, and prioritization.
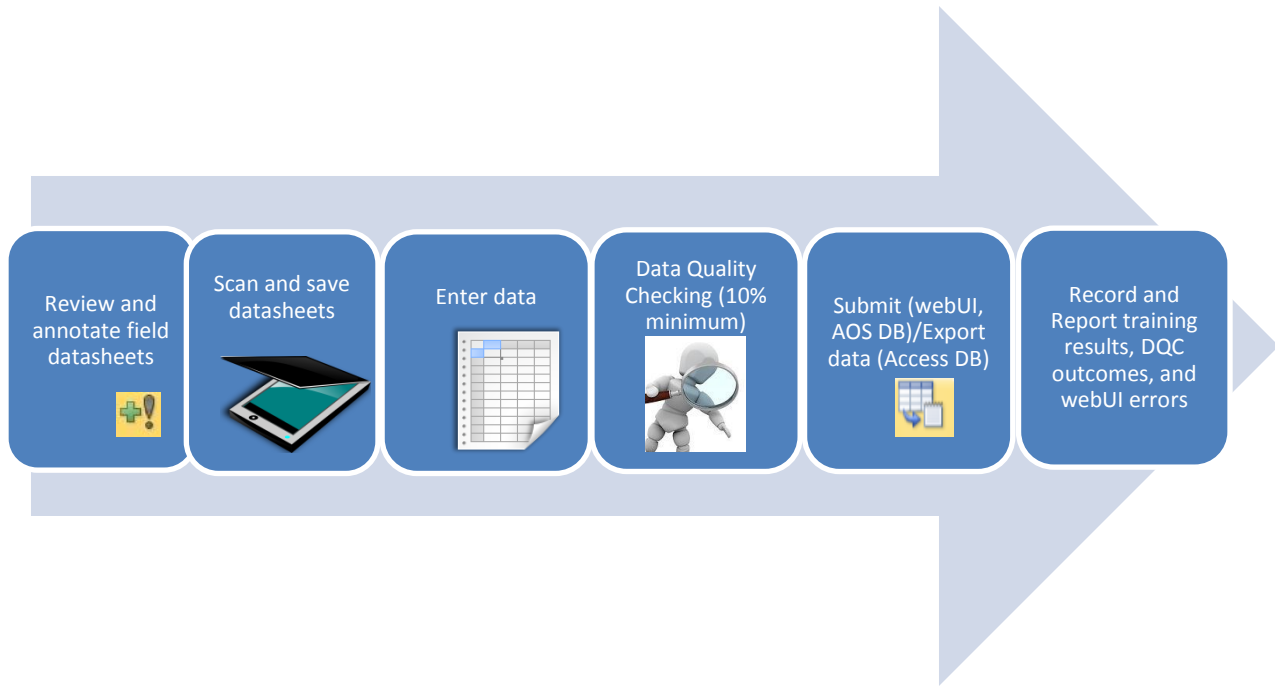

Figure 2 Generalized workflow process from field data collection to data entry when entering data from paper datasheets.

SOP A   Checking and Scanning Field Data Sheets

Field datasheets should first be reviewed to check that all fields are completely filled out, that those fields contain the correct type of data, and that the data are correct. Shorthand notes should be fully annotated on the datasheet; this is to ensure that all information on a datasheet can be read by a separate person without having to reference non-standardized acronyms or abbreviations. Ideally, field datasheet review should occur as close to field collection as possible (i.e. 14 days), and the review carried out by technicians that are trained in the protocol. Field datasheets with clarifications and additions shall be scanned and archived for future reference in a location designated by the domain manager.

A.1   Review Field Datasheets

1. Lead technicians for a particular protocol should review field datasheets at the end of the sampling day or sampling bout, whichever is more reasonable with respect to logistics and scheduling

   a) Review the field datasheets for <u>completeness</u>. Check BOTH sides of the datasheets for extra taxa or notes. Be sure to clarify and annotate any shorthand notes *on the datasheet*; don't assume the shorthand will be meaningful to staff in the future. Finally, check that all required fields/columns have been filled out.

   b) Review the datasheets for correctness. First check that the appropriate data type has been written into the datasheet's fields. For example date fields should not include characters and numeric fields (e.g. "dryMass", "time", "numVials" etc.) should generally not include characters.

      i. While not always feasible, protocol lead technicians should review the correctness of field data as close to the sampling date as possible. Timely datasheet review ensures that errors are (1) detected early in the data entry process and (2) that seasonal staff are available to answer any follow-up questions.

      ii. Take more care with checking fields that mix characters and numbers (e.g. "plotID", "sampleID") as they are more difficult to quickly scan and evaluate.

2. Mark any corrections, clarifications, or additions to the datasheet with red ink and include the associated date of corrections and initials of the reviewing technician.
3. Completed datasheets should be archived in a location designated by the domain manager.
4. See Appendix XX for strategies on managing the organization and tracking of paper datasheets throughout the entire data entry process.

A.2   Scan Field Datasheets

1. Scan reviewed and annotated datasheets.

2. Save scanned datasheets in the folder designated on the NEON Intranet using the following naming convention: "moduleAbbreviation_fds_SITE_YYYYMMDDx"(**Table 3**), 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.

(1) plotID (i.e. siteID_plotNumber e.g. "HARV_003") can also be appended to the filename, after the siteID, to avoid confusion if multiple field crews have been working at the same site and module for a given date.

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 3** for module abbreviations. These are also used in the file naming conventions described below.

SOP B   Data Entry Procedures

This SOP outlines general guidelines and best practices for using data entry forms (DEF) in webUIs and Access databases. Specific step-by-step instructions for each field protocol are provided as training materials on the the NEON Intranet page.

B.1   Transcribing Data: Local User Interfaces (LUI i.e. Access databases)

1. Make sure you have the current version of the database interface.

2. Select the appropriate data entry form (DEF) for the current datasheets.

3. Enter data using the DEFs by following the step-by-step instructions laid out in the data entry training materials, located on the the NEON Intranet page.

4. Verify that all blank cells are intentionally blank before submitting data.

5. Replace commas with semicolons.

6. Do not use special characters (e.g. ', ", /, <, . , $, %, #, ^, &, *, ! etc.) in data fields, and try to avoid their use in the remarks field.

B.2   **Transcribing Data: Web User Interfaces (WebUI)**

1. Be sure you are connected to the NEON network (VPN access is fine). Currently supported web browsers are Mozilla Firefox and Microsoft Internet Explorer; using an unsupported browser may result in lost data. Supported browsers may change in the future.

2. Navigate and login to the appropriate data entry portal; links to data entry portals can be found on the the NEON Intranet page:

   a. **Training Portal**

      i) The training portal provides data entry applications for the purposes of training only. You are free to use this portal to enter test data for training purposes, as these data are not stored in a NEON data repository.

      ii) New technicians should practice data entry on this portal. Entered training data can be exported as a .CSV and reviewed by experienced technicians.

   b. **Production Portal**

      i) The production portal provides data entry applications for actual field data entry. These data will be stored in the NEON data repository and published on the data portal.

      ii) It is nearly impossible to locate and fix incorrectly entered data once they have been submitted to the NEON data repository. Extra time and effort should be devoted to careful data entry and data review when using webUI DEFs.

3. Select the relevant protocol link.

4. Data entry forms are customized for each individual protocol. Detailed training for each protocol interface is available on the the NEON Intranet.

   a. Using your browser's "back" button to navigate between DEF screens is strongly discouraged.

   b. Data entry through a webUI is a two-step process:

      i) 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.

      ii) The final '**Submit**' data button will send the data you have entered to NEON's data repository. You cannot edit or update the data you have submitted to the database. Check your data carefully and ask any questions prior to submitting.

5. Review and correct data upon completing initial entry. Data quality checking (outlined in SOP C below) **must** be carried out before submitting data.

a. Once all of the data have been entered, select "Done entering data, **review and submit the data**"

b. Download the entered data as a .CSV file to review your entries.

c. Review the entered data according to the instructions in **SOP C** below

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.

6. Download and archive a .CSV file of the data if they have been corrected.

a. Be sure to name the file in the appropriate format, described in the training materials

b. The .CSV file should be archived in a local network directory that is designated by the Domain Manager.

c. .CSV files generated from the webUI are useful for tracking a domain's data entry progress

7. '**Submit**' data to the CYI data repository.

8. Record relevant information in the "*FOPS Data Entry Reporting*" spreadsheet on the the NEON Intranet page after finishing data entry on *a daily basis*.

a. Staying up to date on when data entry has been completed is an extremely important step. CYI uses this information to (1) identify *what* data in the NEON database needs to be processed and (2) *when* it should be processed.

b. Relevant information to report includes: protocol, domainID, siteID, and data entry date

9. If you realize that an erroneous dataset has been submitted to the data repository, record the details in the "*Known Data Entry Errors*" form (linked on the on the NEON Intranet or WebUI page).

.CSV

SOP C   Data Quality Checking

C.1   Data Quality Checking (DQC) Procedures

It is imperative that field staff routinely carry out data quality checking (DQC). All data entered from field datasheets shall be quality checked by selecting at least 10% of the data records or 10 records, whichever is greater, to recheck against the hand-written field datasheet. A data record is defined here as a single complete row of data, including all data value fields. All discovered errors shall be corrected prior to electronic data submission. Per-protocol error rates found in data transcription shall be

recorded in a spreadsheet located on the the NEON Intranet page. DQC should be routinely carried out through a field season as data are entered to ensure high data quality. Recording and reporting the results of DQC also provides a "snapshot" of data collection and data entry progress before a field season is over. **It is extremely important that DQC procedures are implemented prior to submitting data for webUIs: erroneously entered data cannot currently be corrected in NEON's data repository.**

1. Field technicians shall quality check a minimum of 10% of the entered data records, or 10 records, whichever is higher, **per SOP and per each sampling bout** or data entry bout (clarified below) by comparing entered values to the original paper datasheet. Specifically, 10 records or 10% of records in a sampling or data entry bout should be checked, whichever number of greater.

   a. **Per sampling or data entry bout**: data quality checks should be carried out for each field sampling bout during a field season, rather than postponed until the end of an entire field season. This is done so that errors in data entry and problematic data entry habits can be detected and corrected early in a field season.

   b. **Per SOP**: the number of records reviewed should be considered per protocol SOP rather than all SOP records totaled together.

      i. **Example**: a domain has 40 records from Litterfall SOP C (Field Sampling) and 390 records from Litterfall SOP D (Lab Processing – Dry Mass) for bout 3 of the season. Staff should therefore review 10 records from SOP C (because 10% of 40 is only 4 records) and 39 records from SOP D (because 10% of 390 is 39).

Table 2. Quick reference table for data quality checking more than 50 records

Minimum Records to Check	Auditor Checks Every *N* Rows
10%	10
15%	7
20%	5
25%	4
30-40%	3
50%	2

2. Procedures for data quality checking vary depending on the data entry interface.

C.2 WebUI DQC

1. **For WebUIs**, the exported .CSV file should be reviewed *prior to final submission*, thus the DQC step is embedded within the data entry procedures outlined in SOP B.2.

2. For .CSV files with more than 50 records, check at least 10% of the records by reviewing every *N* rows across multiple datasheets.

   a. Refer to **Table 2 for** details on which records to check.

3. Compare the entered data values of each field/column to the original paper datasheet with one of the following procedures:

   a. **Read Aloud Review** : Data checking conducted by two technicians: one technician should read the values on the original paper sheet aloud, while the other visually checks against the electronically entered values.

   b. **Independent Data Review** : If it is not possible for two technicians to conduct DQC together, then a technician who did not enter the current dataset should visually review the dataset.

   c. **Redundant Data Review** If a separate technician cannot conduct DQC, then the technician who entered the data should also visually re-check the data.

      i. It is strongly recommended that technicians employ the IDR or RAR procedures over RDR.

4. Correct the data value if an error has been found; add a new column to the .CSV named "qc", and note that the record has been quality checked with the following codes:

      i. "**QC0**" if the checked record has no errors

      ii. "**QC1**" if the checked record had errors and the errors were corrected

5. Record the number of records checked (**totalRecordsChecked**), the number of records with errors discovered (**totalErrors**), the date (**checkDate**), the 3-letter protocol abbreviation (**protocolID**), the 3-letter domain ID (**domainID**), the data quality checking procedure used (**dqcMethod** e.g. DDR, IDR, RDR), and the email of the recording technician (**checkedBy**) in the "Error Rate Tracking" form located on the the NEON Intranet page.

   a. Errors should be counted and summarized per record (not per cell). For example, a single record with an incorrect collectDate, an incorrect plotID, and an incorrect tagID would count as *one error*.

   b. Do not count errors that you inadvertently find while checking random records, however these records should still be corrected. We do not include these in totalErrors because it would bias our "random sampling" of records.

6. Calculate the error rate percentage by dividing **totalErrors** by **totalRecords** and determine whether additional DQC for the data table needs to occur**:**

   a. If the error rate is greater than 5%, an additional 10% of data should be randomly checked with the same procedure. If there error rate is again greater than 5% after a second review, the entire datasheet should be quality checked

   b. If the original check reveals an error rate greater than 10%, then a 100% of the data should be reviewed.

7. Continue with the data entry steps outlined in SOP B.2 for submitting webUI data.

8. Field datasheets should be annotated in the "footer" (i.e. the bottom margin) of the sheet with the date, technician's initials, and number of records checked after DQC has occurred.

9. Quality checked datasheets should be then be archived.

C.3   Access Database DQC

1. **For Access databases**, data should be reviewed after a bout of data entry for each data table, and _must_ occur prior to submitting data to the dropbox

2. Determine the number of records to review, if there are more than 50, according to **Table 2**

3. Locate and have at hand the original field datasheets for a specific protocol and SOP

   a. Try to organize the paper datasheets by siteID and plotID, if possible

   b. Check every *N* record, according to **Table 2**, starting from the set of _paper datasheets_ rather than the digital records. Doing this will simplify finding and checking records, as digital records are more easily sorted than the paper records.

      i. **Tip**: limit the number of records you are viewing in the database by filtering to the finest resolution field identifier e.g. plotID, clipID, trapID etc.

4. Open the appropriate data table for each SOP

   a. For example, if you are quality checking the "Litterfall: Trap Deployment" field datasheet, open the "Trap Deployment" table in the Access database.

5. Compare the data values of each digitally entered field/column to the original paper datasheet with one of the following procedures:

   a. **Read Aloud Review (RAR)**: Data checking conducted by two technicians: one technician should read the values on the original paper sheet aloud, while the other visually checks against the electronically entered values.

b. **Independent Data Review (IDR**): If it is not possible for two technicians to conduct DQC together, then a technician who did not enter the current dataset should visually review the dataset.

c. **Redundant Data Review, RDR:** If a separate technician cannot conduct DQC, then the technician who entered the data should also visually re-check the data.

   i. It is strongly recommended that technicians employ the IDR or RAR procedures over RDR.

6. Correct the data value if an error has been found; note in the "qc" column that the record has been quality checked with the following codes:

   i. "**QC0**" if the checked record has no errors

   ii. "**QC1**" if the checked record had errors and the errors were corrected

7. Record the number of records checked (**totalRecordsChecked**), the number of errors found (**totalErrors**), the date (**checkDate**), the 3-letter protocol abbreviation (**protocolID**), the 3-letter domain ID (**domainID**), the data quality checking procedure used (**dqcMethod** e.g. DDR, IDR, RDR), and the email of the recording technician (**checkedBy**) in the "Error Rate Tracking" form located on the the NEON Intranet page.

   a. Errors should be counted and summarized per record (not per cell). For example, a single record with an incorrect collectDate, an incorrect plotID, and an incorrect tagID would count as *one error*.

   b. Do not count errors that you inadvertently find while checking random records, however these records should still be corrected. We do not include these in totalErrors because it would bias our "random sampling" of records.

8. Calculate the error rate percentage by dividing **totalErrors** by **totalRecords** and determine whether additional DQC for the data table needs to occur**:**

   a. If the error rate is greater than 5%, an additional 10% of data should be randomly checked with the same procedure. If there error rate is again greater than 5% after a second review, the entire datasheet should be quality checked

   b. If the original check reveals an error rate greater than 10%, then a 100% of the data should be reviewed.

a)

CYI data repositoryCYI data repository.CSV

CYI data repository.CSV.CSV.CSV.CSV.CSV

Table 3. Module abbreviations used in worksheet and table naming.

Subsystem	Abbreviation	Module
AOS	AFM	Aquatic field metadata
AOS	ALG	Periphyton, seston, and phytoplankton
AOS	AMC	Aquatic Microbes
AOS	APC	Aquatic Plant Point Counts
AOS	APL	Aquatic Plant Clip harvest
AOS	ASI	Aquatic stable isotopes
AOS	BAT	Lake/river bathymetry
AOS	DEP	Depth and secchi profiles
AOS	DSC	Stream discharge
AOS	GWC	Groundwater chemistry
AOS	FSL	Fish lakes
AOS	FSS	Fish streams
AOS	INV	Aquatic macroinvertebrates
AOS	LSC	Lake sediment chemistry
AOS	REA	Stream reaeration
AOS	RPL	Riparian lakes and rivers
AOS	RPS	Riparian streams
AOS	SDG	Stream dissolved gas
AOS	SMR	Stream morphology
AOS	SSC	Stream sediment chemistry
AOS	SWC	Surface water chemistry
AOS	ZOO	Zooplankton
TOS	BET	Beetles
TOS	BGC	Biogeochemistry
TOS	BRD	Birds
TOS	CWD	Coarse woody debris
TOS	DHP	Digital Hemispherical Photos for Leaf Area Index
TOS	DIV	Plant diversity
TOS	HBP	Herbaceous productivity
TOS	LBS	Camera-quadrat for Leaf Area Index by Species
TOS	LTR	Litter and fine woody debris
TOS	MAM	Mammals
TOS	MFB	Mat-forming bryophyte production
TOS	MOS	Mosquitoes
TOS	PHE	Plant phenology
TOS	SLS	Soil sampling
TOS	SME	Soil microbes
TOS	TCK	Ticks
TOS	VST	Vegetation structure

7 REFERENCES

Brunt, J.W. 2000. Data Management Principles, Implementation and Administration. *Ecological Data – Design, Management, and Processing (ed.* W.K. Michener and J.W. Brunt*),* Blackwell Science Ltd, Oxford.

Chapman, A.D. 2005. Principles of Data Quality – Report for the Global Biodiversity Information Facility 2004. Copenhagen: GBIF.

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/

8 APPENDICES

8.1 Appendix A: Best Practices for Data Entry and Data Quality Management

This SOP outlines best practices for different phases of the data entry process. It also provides suggestions for managing and tracking data quality checks and communicating issues that may arise throughout the entire process.

A.1 Best Practices for Data Entry and Data Quality Checking

- Any shorthand notes written onto paper field datasheets, for example a shorter taxonID codes not approved by NEON staff, *must be* annotated on each datasheet that contains the notes. This is to ensure that any staff that revisit these datasheets are fully able to transcribe the datasheet without external reference.

- Set aside a designated time for critical data entry as soon as possible after field data collection.

- Make sure your workspace is free of clutter and other distractions while entering data.

- It is preferable, although not required, to assign data entry to personnel who have experience with collecting field data for a particular protocol. If the personnel have not conducted field sampling for a specific protocol, they should familiarize themselves with the contents of each field in a datasheet.

- 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 data entry form. 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.

- 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.

- Do not alter the Access database forms or tables unless instructed. **Modifying fields, tables, or forms can disrupt the functionality of validation rules, forms, and queries.** These alterations may further result in data that are not compatible with NEON's data ingest and processing system.

- **Never** use special characters (e.g., #, %, ~, *), even in the remarks (often referred to as comments) column.

- 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.

A.2   Managing Data Entry Tasks and Data Quality Reporting

The data entry workflow process can be complex because each protocol has a unique set of fields, datasheets, and data entry interfaces. Below are a few suggestions for managing the process:

- Assign specific data entry roles (e.g. "Data Entry Lead") to certain personnel for tracking and handling tasks for *specific protocols*. This is logically delegated to lead technicians (full-time or seasonal) that are specialized in botany, invertebrate, small mammal sampling etc.
  - Protocol Leads (i.e. small mammals lead tech, botany lead tech, aquatics lead tech etc.) should focus on reviewing and correcting information on **paper datasheets**
  - Data Entry Leads should focus on tracking data entry progress for their area of expertise rather than datasheets for the entire office
  - Data Entry Leads should be assisted by at least one other technician that is also designated as a "Data Entry Lead". This ensures that a lead's absence does not derail the data entry process.
  - The Data Entry Lead should be trained and experienced with the entire data entry process.
- Track the progress of datasheets in an organized area of the domain support facility. A filing system that groups datasheets by their progress in the data entry process can help alleviate tracking the fate of a datasheet without utilizing a separate spreadsheet:
  - Reviewed
  - Scanned
  - Entered
  - Quality Checked

- o   Archive

8.2   Appendix B: NEON Specific Guidelines and Tips for Taxonomy related Data Entry

A.1   Guidelines

- Identification References:

  - o   a free text field to capture the key and/or field guides used to identify species during a bout

  - o   Enter information only if you used a reference to identify a particular species (for all protocols except small mammals).

  - o   Copy and paste the reference title from the protocol-specific list on the NEON Intranet) to ensure consistent formatting. In the case of multiple sources, use semicolons between different sources.

- Entering a taxon:

  - o   **Type** any part of the Scientific Name or taxonID code into the box and a list will appear.

    - **Example**: Triticum macha Dekap. & Menab.(TRMA11/TRAE)

      - Scientific Name = 'Triticum macha'
      - Authorship = 'Dekap & Menab.')
      - taxonID =TRMA11
      - acceptedTaxonID = TRAE

  - o   **Taxon codes**:

    - NEON taxon codes can be found on the NEON Intranet: FSU-FOPS/taxonTables
    - NEON plant taxon codes follow USDA plants

- Entering an unknown species**:**

  - o   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.

  - o   There are two options for most unknowns -- a 'sp.' and 'spp.' variant

  - o   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:
  - o 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.
  - o 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:
  - o 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.
  - o You can increase traceability by noting the original morphospecies id in the remarks field (NOT the morphospecies field).

o   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'.

o   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').

o   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 the NEON Intranet (FSU-FOPs/morphospeciesTracking).

A.2   Tips

- A sample ID "cheat sheet" is useful to have near the computer, since sample IDs across protocols are similar to one another, but may differ in format.

- Do not write anything in the 'remarks' field that you do not want to show up on the internet (profanity, identifying information about technicians, cannabis references). These comments will be included in the data download *exactly as written*.

8.3 Appendix C: Data Entry training checklist

Data Entry Training Checklist

So, you're ready to begin entering _____ data.
<div align="center">(protocol)</div>

Make sure you complete the following checklist before beginning. A permanent staff member or lead technician will need to sign off on your checklist before actual data entry can begin.

- ☐ Read through data entry protocol
- ☐ Read through Training Materials on thethe NEON Intranet page (WebUI only)
    - ☐ Powerpoint
    - ☐ FAQs/ Common Mistake forms (if available)
- ☐ Enter at least one datasheet, plot or subplot of data into the training WebUI portal or Access Database (at least 20 lines of data)
  http://cert-as-web-1.ci.neoninternal.org:8080/
- ☐ Talk with them through
    - ☐ Annotating datasheets
    - ☐ WebUI Portal login
    - ☐ Appropriate remarks/ use of commas and special characters
    - ☐ QA/QC process
    - ☐ .CSV file review/ saving .CSV files
    - ☐ Where entered and checked datasheets get stored
    - ☐ How to report what data has been entered
    - ☐ How to report time spent on data entry

_____   _____   _____   _____
Trainee Signature   Date   Trainer Signature   Date