NEON USER GUIDE TO CHEMICAL PROPERTIES OF SURFACE WATER (NEON.DP1.20093) AND CHEMICAL PROPERTIES OF GROUNDWATER (NEON.DP1.20092)

PREPARED BY	ORGANIZATION	DATE
Keli Goodman	AOS	03/30/2020
Caren Scott	AOS	03/30/2020
Samantha Weintraub	TOS	03/30/2020
Zachary Nickerson	AOS	03/30/2020

CHANGE RECORD

REVISION	DATE	DESCRIPTION OF CHANGE
A	05/24/2017	Initial Release
B	04/01/2020	External lab data format change from wide-format to long-format

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

1 DESCRIPTION

1.1 Purpose

This document provides an overview of the data included in this NEON Level 1 data product, the quality controlled product generated from raw Level 0 data, and associated metadata. In the NEON data products framework, the raw data collected in the field, for example the specific conductance of water, are considered the lowest level (Level 0). Raw data that have been quality checked via the steps detailed herein, as well as simple metrics that emerge from the raw data are considered Level 1 data products.

The text herein provides a discussion of measurement theory and implementation, data product provenance, quality assurance and control methods used, and approximations and/or assumptions made during L1 data creation.

1.2 Scope

This document describes the steps needed to generate the L1 data products Chemical properties of surface water and Chemical properties of groundwater - the chemistry of surface and groundwaters based on internal and external laboratory analyses as well as associated metadata from field collections. This document also provides details relevant to the publication of the data products via the NEON data portal, with additional detail available in the files NEON Data Variables for Chemical properties of surface water (NEON.DP1.20093) (AD[08]) and NEON Data Variables for Chemical properties of groundwater (NEON.DP1.20092) (AD[09]), provided in the download package for this data product.

This document describes the process for ingesting and performing automated quality assurance and control procedures on the data collected in the field pertaining to AOS Protocol and Procedure: Water Chemistry Sampling in Surface Waters and Groundwater (AD[10]). The raw data that are processed in this document are detailed in the files NEON Raw Data Validation for Water chemistry, isotopes, dissolved gas, and microbes sampling, Level 0 (NEON.DP0.20090) (AD[04]), NEON Raw Data Validation for Chemical properties of surface water, Level 0 (NEON.DP0.20093) (AD[05]), NEON Raw Data Validation for Water Chemistry External Lab Data (NEON.DP0.20286) (AD[06]), and NEON Raw Data Validation for Water Chemistry External Lab Summary Data (NEON.DP0.20287) (AD[07]), provided in the download package for this data product. Please note that raw data products (denoted by ‘DP0’) may not always have the same numbers (e.g., ‘20093’) as the corresponding L1 data product.

2 RELATED DOCUMENTS AND ACRONYMS

2.1 Associated Documents

AD[01]	NEON.DOC.000001	NEON Observatory Design (NOD) Requirements
AD[02]	NEON.DOC.001152	NEON Aquatic Sampling Strategy
AD[03]	NEON.DOC.002652	NEON Level 1, Level 2 and Level 3 Data Products Catalog
AD[04]	NEON.DP0.20090.001 _dataValidation.csv	NEON Raw Data Validation for Water chemistry, isotopes, dissolved gas, and microbes sampling, Level 0 (NEON.DP0.20090)
AD[05]	NEON.DP0.20093.001 _dataValidation.csv	NEON Raw Data Validation for Chemical properties of surface water, Level 0 (NEON.DP0.20093)
AD[06]	NEON.DP0.20286.001 _dataValidation.csv	NEON Raw Data Validation for Water Chemistry External Lab Data (NEON.DP0.20286)
AD[07]	NEON.DP0.20287.001 _dataValidation.csv	NEON Raw Data Validation for Water Chemistry External Lab Summary Data (NEON.DP0.20287)
AD[08]	NEON.DP1.20093.001 _variables.csv	NEON Data Variables for Chemical properties of surface water (NEON.DP1.20093)
AD[09]	NEON.DP1.20092.001 _variables.csv	NEON Data Variables for Chemical properties of groundwater (NEON.DP1.20092)
AD[10]	NEON.DOC.002905	AOS Protocol and Procedure: Water Chemistry Sampling in Surface Waters and Groundwater
AD[11]	NEON.DOC.000008	NEON Acronym List
AD[12]	NEON.DOC.000243	NEON Glossary of Terms
AD[13]	OS_Generic _Transi-tions.pdf	NEON Algorithm Theoretical Basis Document: OS Generic Transitions
AD[14]		NEON's Ingest Conversion Language (NICL) specifications

2.2 Acronyms

Acronym	Definition
ALK	Alkalinity
ANC	Acid Neutralizing Capacity
GWC	Ground Water Chemistry
SWC	Surface Water Chemistry

3 DATA PRODUCT DESCRIPTION

The Chemical properties of surface water (NEON.DP1.20093) and Chemical properties of groundwater (NEON.DP1.10033) data products provide chemistry data for surface and groundwater samples collected using AOS Protocol and Procedure: Water Chemistry Sampling in Surface Waters and Groundwater (AD[10]). These procedures implement the guidelines and requirements described in the NEON Aquatic Sampling Strategy (AD[02]). All data are reported at the spatial resolution of a single water sample, collected from a unique stationID within a sampled water body. The temporal resolution is that of a single collection date.

Field sampling strategies are specific to the type of waterbody and are described further below. Chemical data are produced by NEON technicians (alkalinity and acid neutralizing capacity (ANC)) as well as external laboratories and include measurements of total, dissolved, and particulate nutrients, as well as anions, cations and general chemistry (i.e., conductivity and pH).

Surface and groundwater chemistry data allow researchers to assess aquatic biogeochemical cycles and dominant driver(s) of nutrient fluxes within aquatic ecosystems and watersheds. Measuring long-term trends in surface and groundwater chemistry is part of the overall NEON biogeochemistry goal to understand how major nutrient and carbon fluxes within and across air, land and water systems change over 30 years.

3.1 Spatial Sampling Design

In wadeable streams, dip sampling in the thalweg is used to obtain samples, assuming the stream channel is completely mixed (Figure 1a). In non-wadeable streams, dip sampling in the thalweg at 0.5 m depth is used - unless the steam is stratified, in which case a 0.5 m epilimnion sample plus an integrated hypolimnion sample are taken (Figure 1b). For all stream types, samples are collected as close as possible and downstream of sensor sets so that sensor measurements can be validated with water chemistry samples.

In lakes, up to three locations are designated for sampling: one in the deepest part of the lake, one near the most prominent intlet, and one near the outlet. Samples are taken from variable depths dependent on the dregree of lake stratification and are collocated near lake buoy, inlet, and outlet sensor infrastructure (Figure 1c).

Groundwater chemistry samples are collected from up to eight 2” diameter, shallow (<100 ft depth) groundwater observation wells on the perimeter of sampled waterbodies (Figure 1). A subset of wells are selected for sampling on a site-by-site basis during each sampling event. Periodic changes to the selected subset of wells may occur during the life of the Observatory and are guided by various parameters, including changes in hydrologic conditions (dry wells, changes in hydrologic flow paths) and status of infrastructure (damaged wells).

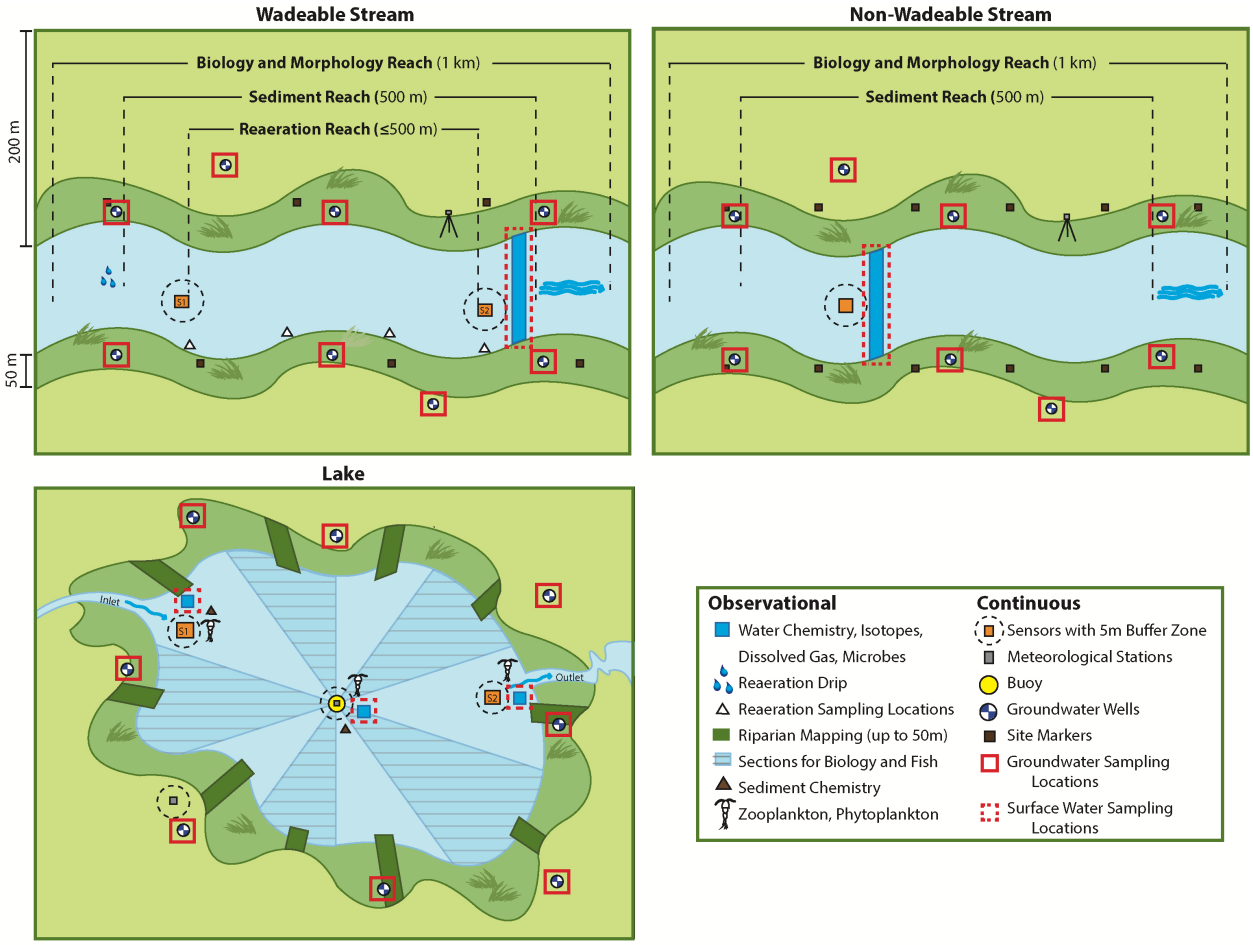

Figure 1: Generic layout of surface and groundwater sampling locations in wadeable streams, non-wadeable streams, and lake sites. See AD[10] for further details.

3.2 Temporal Sampling Design

For streams, water chemistry sampling occurs up to 26 times per year, approximately every other week but guided by historical stream discharge data for each NEON site, with ANC measured only monthly at most stream sites. For example, wadeable streams with little or no flow during the summer dry-season or that are completely frozen during the winter are sampled more intensively during wet periods or snowmelt. When applicable, stream chemistry samples are collected to coincide with NEON atmospheric wet chemistry sampling (Tuesdays) as well as Observatory-wide chemistry sampling efforts.

Lake water chemistry will be collected 12 times per year - approximately monthly and during shoulder seasons to capture ice-on/ice-off and lake turnover events. When applicable, lake chemistry samples are also preferentially collected to coincide with other aqueous chemistry sampling efforts. Groundwater chemistry samples are collected up to twice per year, roughly during early spring and late fall, based on historic cumulative discharge.

Samples will be collected within +/- 1 day of surface water chemistry sampling events where possible.

All water samples are filtered as soon as possible following collection (preferably within 3 hours) and held at cold temperatures (~ 4°C - 2°C) until they are either shipped to a water chemistry lab or processed by NEON technicians for alkalinity and ANC. These tasks should occur within 24 hours. The maximum allowable time period between sample collection and NEON lab processing is 72 hours.

3.3 Laboratory Quality Assurance and Uncertainty

Domain support facility analyses of alkalinity and ANC follow widely adopted methods for measurement of those analytes, namely the US Geological Survey National Field Manual for the Collection of Water-Quality Data, and all NEON technicians conducting this work receive proper training. For external laboratory analyses, facilities have been chosen for their use of analytical methods widely adopted by the aquatic chemistry community. Labs report the method detection limit, along with long-term analytical precison and uncertainty of standards analyzed as unknowns, for each analyte in a summary file. This allows users to interpret and model the chemistry data in the context of its uncertainty range. Contracted external facilities upload a summary file (swc_externalLabSummaryData) when they begin work for NEON, then again once per year or whenever their information changes (for example, a new instrument is aquired or a change is detected in analytical precision). Additionally, NEON's Calibration/Validation department has regular procedures for auditing the quality assurance of external laboratories and their reports are available to data users.

3.4 Variables Reported

All variables reported from the field technician or laboratory (L0 data) are listed in the files, NEON Raw Data Validation for Water chemistry, isotopes, dissolved gas, and microbes sampling, Level 0 (NEON.DP0.20090) (AD[04]), NEON Raw Data Validation for Chemical properties of surface water, Level 0 (NEON.DP0.20093) (AD[05]), NEON Raw Data Validation for Water Chemistry External Lab Data (NEON.DP0.20286) (AD[06]), and NEON Raw Data Validation for Water Chemistry External Lab Summary Data (NEON.DP0.20287) (AD[07]). All variables reported in the published data (L1 data) are also provided separately in the file, NEON Data Variables for Chemical properties of surface water (NEON.DP1.20093) (AD[08]) and NEON Data Variables for Chemical properties of surface water (NEON.DP1.20093) (AD[08]).

Field names have been standardized with Darwin Core terms (http://rs.tdwg.org/dwc/; accessed 16 February 2014), the Global Biodiversity Information Facility vocabularies (http://rs.gbif.org/vocabulary/gbif/; accessed 16 February 2014), the VegCore data dictionary (https://projects.nceas.ucsb.edu/nceas/projects/bien/wiki/VegCore; accessed 16 February 2014), where applicable. NEON AOS spatial data employs the World Geodetic System 1984 (WGS84) for its fundamental reference datum and Earth Gravitational Model 96 (EGM96) for its reference gravitational ellipsoid. Latitudes and longitudes are denoted in decimal notation to six decimal places, with longitudes indicated as negative west of the Greenwich meridian.

Some variables described in this document may be for NEON internal use only and will not appear in downloaded data.

3.5 Spatial Resolution and Extent

The finest resolution at which water chemistry spatial data are reported is a single sample collected from a unique stationID. Overall, this results in a spatial hierarchy of:

sampleID (unique ID given to the individual water sample) → **stationID** (ID of the sampling location) → **siteID** (ID of NEON site) → **domainID** (ID of a NEON domain).

StationID in wadeable and non-wadeable (i.e. rivers) streams is indicated as 'ss', and stationIDs for lakes are designated as 'in', 'ot', 'c0', and 'c1','c2', and 'c3', as needed (if center is stratified), with 'c1' being the top layer.

The basic spatial data included in the data downloaded include the latitude, longitude, and elevation of the stationID where sampling occurred, plus associated uncertainty due to GPS error. Shapefiles of all NEON Aquatic Observation System sampling locations can be found in the Document Library: http://data.neonscience.org/documents. If for some reason samples cannot be taken at the designated stations, the coordinates for the center of the site will be returned. In this case, the actual coordinates of the sampling location are recorded in the **altLocation**, **altLatitude** and **altLongitude** fields.

For groundwater, users can find the depth of the well as the 'z offset' in the full spatial data which can be accessed via:

1. The def.extr.geo.os.R function from the geoNEON package, available here: https://github.com/NEONScience/NEON-geolocation
2. The NEON API: http://data.neonscience.org/data-api

3.6 Temporal Resolution and Extent

The finest resolution at which water chemistry temporal data are reported is the **collectDate**, a single date on which water chemistry samples were collected. The total number of sampling events per year is expected to be 26 per wadeable and non-wadeable stream, 12 per lake, and 2 per groundwater well subset per site. At most wadeable stream sites, ANC will only be measured 12 times, but alkalinity will still be measured 26 times.

The NEON Data Portal currently provides data in monthly files for query and download efficiency. Queries including any part of a month will return data from the entire month. All queries, regardless of the date range specified, will include a copy of swc_externalLabSummaryData or gwc_externalLabSummaryData which provides summary information from the external chemistry lab about the method detection limits, the equipment used, and precision and accuracy. Code to stack files across months is available here: https://github.com/NEONScience/NEON-utilities

3.7 Associated Data Streams

The super parent sample of the field data from the Chemical properties of surface water and Chemical properties of groundwater data products is shared with several other related data products. **siteID**, **stationID** and **collectDate** (or **parentSampleID** of the xxx_fieldSuperParent table in each data product) are the linking variables that tie samples and their associated metadata to those found in the Dissolved gases in surface water (DP1.20097.001),

Stable isotope concentrations in surface waters (DP1.20206.001), Stable isotope concentrations in Groundwater (DP1.20276.001) and Surface water microbe cell count (DP1.20138.001) data products.

3.8 Product Instances

The NEON Observatory contains 34 aquatic sites, consisting of 24 wadeable streams, 3 non-wadeable streams (rivers), and 7 lakes.

Surface and groundwater chemistry sampling yields one unique sample per stationID per sampling event. There will be up to 26 sampling events per year in wadeable and non-wadeable streams, up to 12 sampling events in lakes, and up to 2 sampling events for ground water. Thus in wadable streams, there will be up to 26 unique sample records per site per year, non-wadeable streams will yield up to 26 records per year, lakes will yield 36-60 records per year, depending on stratification, and groundwater will produce 8 records per year. Observatory-wide, this will yield a total of 1226-1394 water chemistry records per year.

NOTE: Replicate samples may be taken of a small percentage of our samples. If replicate samples are taken, there will be one unique sample per **replicateNumber** per stationID per sampling event, and the sample ID(s) of the replicate sample(s) will have the **replicateNumber** appended to the end.

3.9 Data Relationships

The protocol dictates that each siteID x stationID combination is sampled at least once per event (one record expected per parentSampleID in swc_fieldSuperParent). A record from swc_fieldSuperParent may have zero or one child records in swc_fieldData, depending on whether a water sample was collected. In the event that a water sample cannot be taken, a record will still be created in swc_fieldSuperParent, and swc_fieldSuperParent.**samplingImpractical** will be something other than NULL, but there will be no corresponding record in swc_fieldData. Each record from swc_fieldData is expected to have up to three child records in swc_domainLabData (one for ALK, one for ANC if measured, and one for a replicate ALK if measured), and each record from swc_fieldData is also expected to have one record for each chemical analyzed in swc_externalLabDataByAnalyte. However, duplicates and/or missing data may exist where protocol and/or data entry abberations have occurred; *users should check data carefully for anomalies before joining tables*.

swc_fieldSuperParent_pub.csv or gwc_fieldSuperParent.csv - > One record expected per parentSampleID.

swc_fieldData_pub.csv or gwc_fieldData_pub.csv - > One record expected per parentSampleID per collectDate (day of year, local time), generates a single sampleID of filtered water to be used for external laboratory analyses

swc_domainLabData_pub.csv or gwc_domainLabData_pub.csv - > Up to three records expected per sampleID, generates domainSampleIDs associated with either alkalinity or ANC analyses and replicate alkalinity analyses.

swc_externalLabDataByAnalyte_pub.csv or gwc_externalLabDataByAnalyte_pub.csv - > One record expected per sampleID x analyte combination, associated with external laboratory chemical analyses.

(Depricated as of 04/01/2020) swc_externalLabData_pub.csv or gwc_externalLabData_pub.csv - >
One record expected per sampleID, associated with external laboratory chemical analyses. Users

that downloaded external lab data before 04/01/2020 should note that all data is now downloaded as swc_externalLabDataByAnalyte_pub.csv or gwc_externalLabDataByAnalyte_pub.csv) swc_externalLabSummaryData_pub.csv or gwc_externalLabSummaryData_pub.csv - > One record expected per laboratoryName x analyte x method x labSpecificStartDate combination. Can use corresponding variables in externalLabData tables to associate sample data with relevant uncertainty values and method detection limits

sampleIDs and **sampleBarcodes** will be generated for each sampling event. After shipment to external labs and domain lab processing are complete, any physical sample that remains will be discarded.

3.10 Special Considerations

Samples analyzed for UV Absorbance (250 and 280 nm) collected between 10/21/2014 and 08/11/2015 were analyzed using a method in which absorbance was not water blank subtracted. Analytes reported under this method are denoted with a data quality flag.

External lab data for all samples collected before 07/30/2019 were originally uploaded to the NEON database in wide-format (i.e., one record per sample, one column per analyte). These data were converted to long-format (one record per sample x analyte) by NEON HQ Scientists on 04/01/2020. The external lab began uploading data in long-format for all samples collected after 07/30/2019. Therefore, only water chemistry data downloaded after 04/01/2020 will ensure all data is in long-format.

4 DATA QUALITY

4.1 Data Entry Constraint and Validation

Many quality control measures are implemented at the point of data entry within a mobile data entry application (field data) and web user interface (UI, lab data). For example, data formats are constrained and data values controlled through the provision of dropdown options, which reduces the number of processing steps necessary to prepare the raw data for publication (Figure 2). An additional set of constraints are implemented during the process of ingest into the NEON database. The product-specific data constraint and validation requirements built into data entry applications and database ingest are described in the documents NEON Raw Data Validation for Water chemistry, isotopes, dissolved gas, and microbes sampling, Level 0 (NEON.DP0.20090), NEON Raw Data Validation for Chemical properties of surface water, Level 0 (NEON.DP0.20093), NEON Raw Data Validation for Water Chemistry External Lab Data (NEON.DP0.20286), and NEON Raw Data Validation for Water Chemistry External Lab Summary Data (NEON.DP0.20287), provided with every download of this data product. Contained within this file is a field named 'entryValidationRulesForm', which describes syntactically the validation rules for each field built into the data entry application. Data entry constraints are described in Nicl syntax in the validation file provided with every data download, and the Nicl language is described in NEON's Ingest Conversion Language (NICL) specifications ([AD[14]).

Data collected prior to 2017 were processed using a paper-based workflow that did not implement the full suite of quality control features associated with the interactive digital workflow.

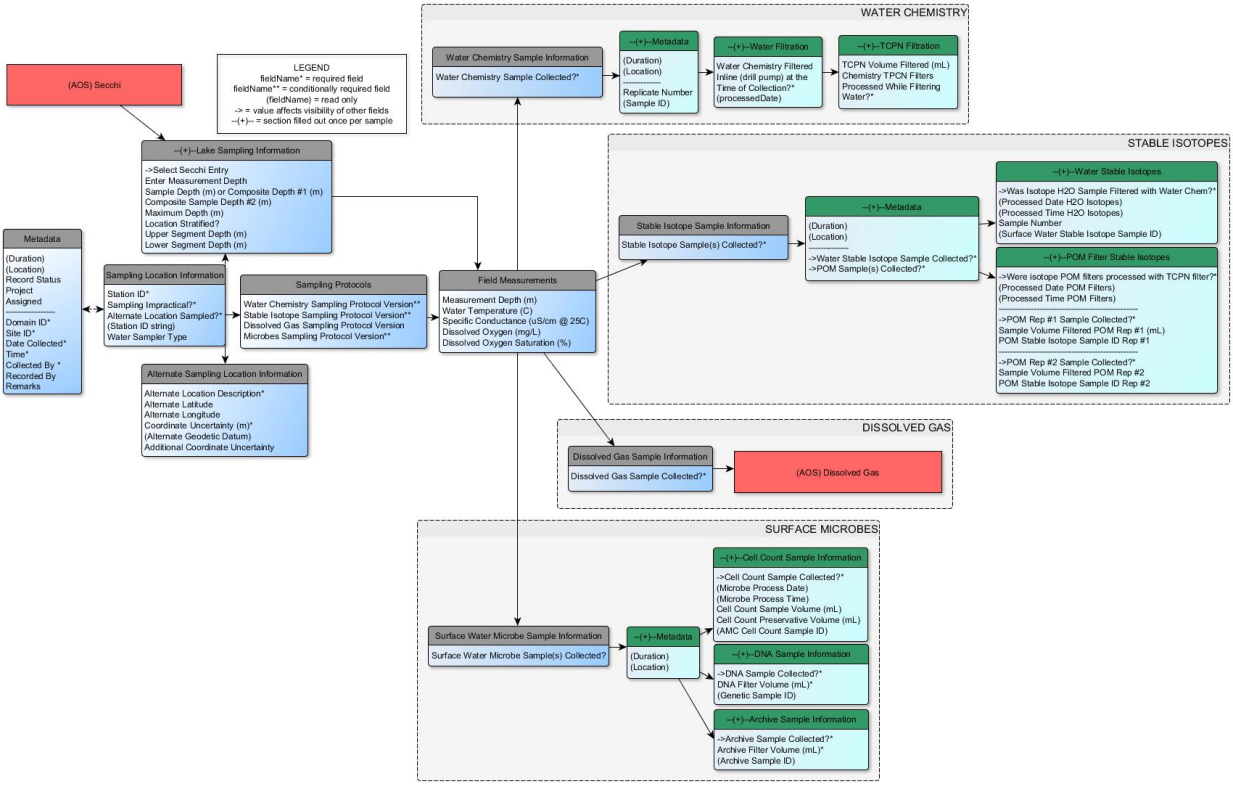

Figure 2: Schematic of the applications used by field technicians to enter water chemistry field data

4.2 Automated Data Processing Steps

Following data entry into a mobile application or web user interface, the steps used to process the data through to publication on the NEON Data Portal are detailed in the NEON Algorithm Theoretical Basis Document: OS Generic Transitions (AD[13]).

4.3 Data Revision

All data are provisional until a numbered version is released; the first release of a static version of NEON data, annotated with a globally unique identifier, is planned to take place in 2020. During the provisional period, QA/QC is an active process, as opposed to a discrete activity performed once, and records are updated on a rolling basis as a result of scheduled tests or feedback from data users. The Change Log section of the data product readme, provided with every data download, contains a history of major known errors and revisions.

4.4 Quality Flagging

The **dataQF** field in each data record is a quality flag for known errors applying to the record. Please see the below for an explanation of **dataQF** codes specific to this product.

fieldName	value	definition
dataQF	Preliminary method: UV absorbance not water blank subtracted	Water processed using an initial method in which blanks were not subtracted
dataQF	legacyData	Data recorded using a paper-based workflow that did not implement the full suite of quality control features associated with the interactive digital workflow
dataQF	formatChange	Data originally entered into the NEON database in wide-format converted to long-format to allow for more flexibility and compatibility in water chemistry analysis

5 REFERENCES

United States Geological Survey, National field manual for the collection of water-quality data: U.S. Geological Survey Techniques of Water Resources Investigations, Book 9, Chapter A4, Version 2.0, 9/2006.

U.S. Geological Survey Techniques of Water Resources Investigations, Book 9, Chapter A6., sec 6.6, Version 2.0, 9/2006 Chapter A6.6, Version 4.0, 9/2012.

Environmental Protection Agency Report: Low-Flow (Minimal Drawdown) Ground-Water Sampling Procedures (Puls, R.W., and Barcelona, M.J., 1996, Report EPA/540/S-95/504) and the United States Geological Survey, National Field Manual for the Collection of Water-Quality Data (U.S. Geological Survey TWRI Book 9, Chapter A4, Version 2.0, 9/2006).