



<i>Title:</i> NEON User Guide to Chemical properties of surface water (DP1.20093.001) and Chemical properties of groundwater (DP1.20092.001)	<i>Date:</i> 08/10/2022
<i>Author:</i> Keli Goodman	<i>Revision:</i> D.1

# NEON USER GUIDE TO CHEMICAL PROPERTIES OF SURFACE WATER (DP1.20093.001) AND CHEMICAL PROPERTIES OF GROUNDWATER (DP1.20092.001)

<b>PREPARED BY</b>	<b>ORGANIZATION</b>
Keli Goodman	AOS
Caren Scott	AOS
Samantha Weintraub	TOS
Zachary Nickerson	AOS
Nora Catolico	AOS



<i>Title:</i> NEON User Guide to Chemical properties of surface water (DP1.20093.001) and Chemical properties of groundwater (DP1.20092.001)	<i>Date:</i> 08/10/2022
<i>Author:</i> Keli Goodman	<i>Revision:</i> D.1

## 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
C	07/16/2020	Included general statement about usage of neonUtilities R package and statement about possible location changes. Updates to quality flags, special considerations, and Figure 1.
D	03/02/2022	Included minor updates for sampling design and analysis changes. Updated section 4.3 Data Revision with latest information regarding data release.
D1	08/10/2022	Updates for external lab quality flags.



## TABLE OF CONTENTS

<b>1</b>	<b>DESCRIPTION</b>	<b>1</b>
1.1	Purpose . . . . .	1
1.2	Scope . . . . .	1
<b>2</b>	<b>RELATED DOCUMENTS AND ACRONYMS</b>	<b>2</b>
2.1	Associated Documents . . . . .	2
2.2	Acronyms . . . . .	3
<b>3</b>	<b>DATA PRODUCT DESCRIPTION</b>	<b>4</b>
3.1	Spatial Sampling Design . . . . .	4
3.2	Temporal Sampling Design . . . . .	6
3.3	Sampling Design Changes . . . . .	6
3.4	Laboratory Quality Assurance and Uncertainty . . . . .	6
3.4.1	Laboratory analyses dependent on sampleCondition . . . . .	7
3.5	Variables Reported . . . . .	7
3.6	Spatial Resolution and Extent . . . . .	7
3.7	Temporal Resolution and Extent . . . . .	8
3.8	Associated Data Streams . . . . .	8
3.9	Product Instances . . . . .	8
3.10	Data Relationships . . . . .	9
3.11	Special Considerations . . . . .	10
<b>4</b>	<b>DATA QUALITY</b>	<b>11</b>
4.1	Data Entry Constraint and Validation . . . . .	11
4.2	Automated Data Processing Steps . . . . .	12
4.3	Data Revision . . . . .	12
4.4	Quality Flagging . . . . .	12
4.5	Note about blank or NA values in analyteConcentration . . . . .	13
4.6	Note for Alkalinity and Acid Neutralizing Capacity . . . . .	14
<b>5</b>	<b>REFERENCES</b>	<b>14</b>

## LIST OF TABLES AND FIGURES

Table 1	Descriptions of the dataQF codes for quality flagging . . . . .	13
Figure 1	Generic layout of surface and groundwater sampling locations in wadeable streams, non-wadeable streams, and lake sites. See AD[06] for further details. . . . .	5
Figure 2	Schematic of the applications used by field technicians to enter water chemistry field data . . . . .	11



<i>Title:</i> NEON User Guide to Chemical properties of surface water (DP1.20093.001) and Chemical properties of groundwater (DP1.20092.001)	<i>Date:</i> 08/10/2022
<i>Author:</i> Keli Goodman	<i>Revision:</i> D.1

# 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 (DP1.20093.001) (AD[07]) and NEON Data Variables for Chemical properties of groundwater (DP1.20092.001) (AD[08]), 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 (DP0.20090.001) (AD[03]), NEON Raw Data Validation for Chemical properties of surface water, Level 0 (DP0.20093.001) (AD[04]), NEON Raw Data Validation for Water Chemistry External Lab Data (DP0.20286.001) (AD[05]), and NEON Raw Data Validation for Water Chemistry External Lab Summary Data (DP0.20287.001) (AD[06]), 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.002652	NEON Data Products Catalog
AD[03]	Available with data download	Validation csv
AD[04]	Available with data download	Validation csv
AD[05]	Available with data download	Validation csv
AD[06]	Available with data download	Validation csv
AD[07]	Available with data download	Variables csv
AD[08]	Available with data download	Variables csv
AD[09]	Available with data download	Categorical Codes csv
AD[10]	NEON.DOC.001152	NEON Aquatic Sampling Strategy
AD[11]	NEON.DOC.002905	AOS Protocol and Procedure: Water Chemistry Sampling in Surface Waters and Groundwater
AD[12]	NEON.DOC.001886	AOS Protocol and Procedure: ASI - Stable Isotope Sampling in Surface and Ground Waters
AD[13]	NEON.DOC.000008	NEON Acronym List
AD[14]	NEON.DOC.000243	NEON Glossary of Terms
AD[15]	NEON.DOC.004825	NEON Algorithm Theoretical Basis Document: OS Generic Transitions
AD[16]	Available on NEON data portal	NEON Ingest Conversion Language Function Library
AD[17]	Available on NEON data portal	NEON Ingest Conversion Language



<i>Title:</i> NEON User Guide to Chemical properties of surface water (DP1.20093.001) and Chemical properties of groundwater (DP1.20092.001)	<i>Date:</i> 08/10/2022
<i>Author:</i> Keli Goodman	<i>Revision:</i> D.1

## 2.2 Acronyms

<b>Acronym</b>	<b>Definition</b>
ALK	Alkalinity
ANC	Acid Neutralizing Capacity
GWC	Ground Water Chemistry
SWC	Surface Water Chemistry
PCN	Particulate Carbon and Nitrogen
mg	Milligram
mL	Milliliter
nm	Nanometer

### 3 DATA PRODUCT DESCRIPTION

The Chemical properties of surface water (DP1.20093.001) and Chemical properties of groundwater (DP1.10033.001) 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[11]). These procedures implement the guidelines and requirements described in the NEON Aquatic Sampling Strategy (AD[10]). 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., specific 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 rivers, a grab sample is collected from the thalweg at 0.5 m depth - unless the river is stratified, in which case a 0.5 m epilimnion sample plus an integrated hypolimnion sample are collected (Figure 1b). In streams and rivers, samples are collected immediately downstream of the most downstream sensor set, S2 or buoy, so that sensor measurements can be validated with water chemistry samples.

In lakes, at the buoy location, samples are taken from variable depths dependent on the degree of lake stratification (Figure 1c and 1d). In lakes with a true inflow and outflow, samples are collected near lake buoy, inflow, and outflow sensor infrastructure (Figure 1d).

Most aquatic sites have up to eight shallow (<100 ft depth) groundwater observation wells on the perimeter of sampled waterbodies (Figure 1). Groundwater chemistry samples are collected from a subset of three to four wells at each site. At wadeable stream sites, the sampling wells are selected in an attempt to cover all of the following categories: upstream, downstream, right bank, and left bank. Preference is also given to wells that are closer to the surface water chemistry sampling locations. For lakes, four sampling wells are selected with two on the side of the water table gradient surrounding the lake. 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).

As much as possible, sampling occurs in the same locations over the lifetime of the Observatory. However, over time some sampling locations may become impossible to sample, due to disturbance or other local changes. When this occurs, the location and its location ID are retired. A location may also shift to slightly



different coordinates. Refer to the locations endpoint of the NEON API for details about locations that have been moved or retired: <https://data.neonscience.org/data-api/endpoints/locations/>

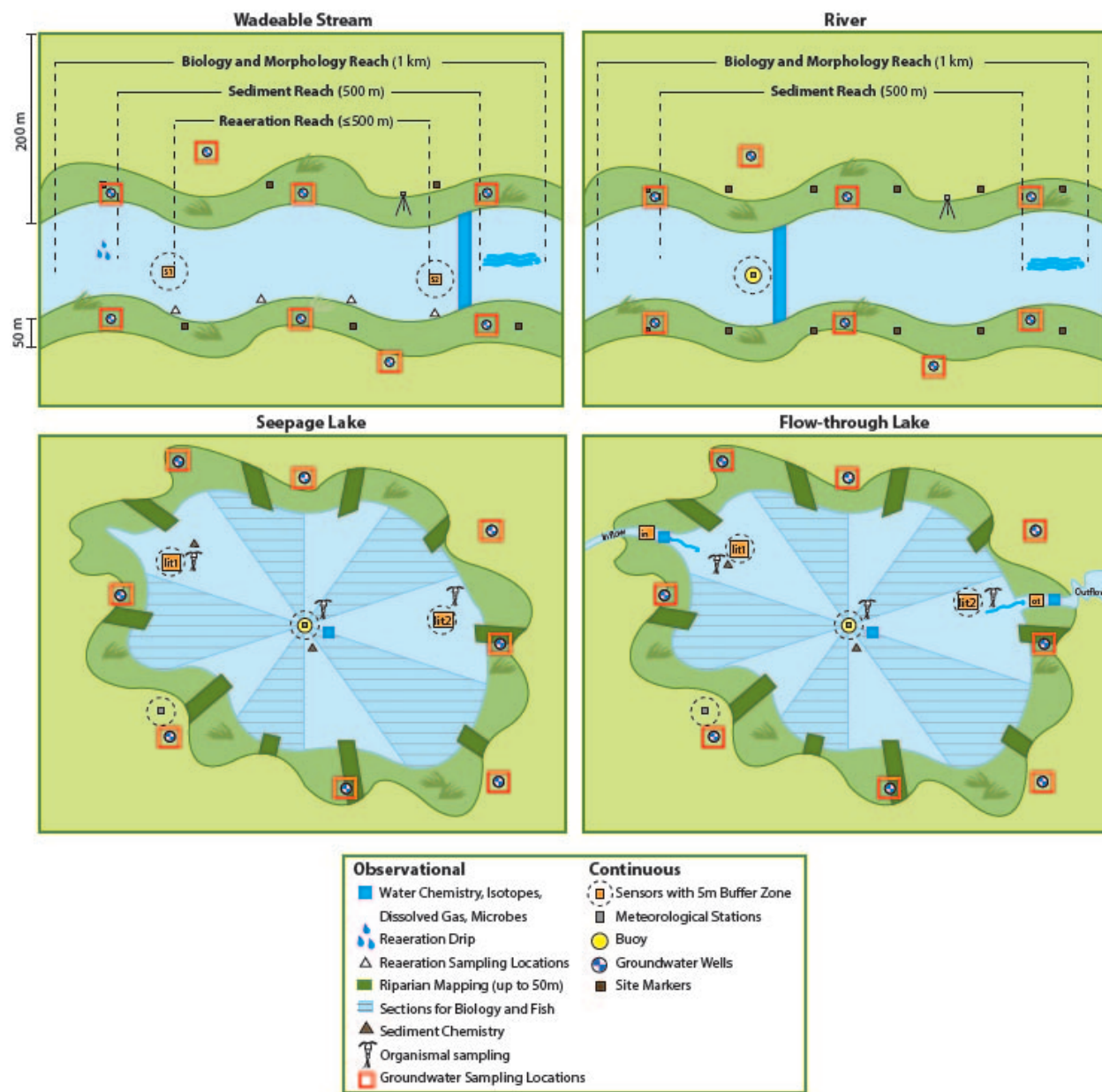


Figure 1: Generic layout of surface and groundwater sampling locations in wadeable streams, non-wadeable streams, and lake sites. See AD[06] for further details.



### 3.2 Temporal Sampling Design

For streams, water chemistry sampling occurs up to 26 times per year, which is monthly with 14 flow-weighted samples guided by historical stream discharge data for each NEON site. 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. 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.

Groundwater samples are filtered in-line at the time of sampling when possible. All water samples are filtered as soon as possible following collection (preferably within 3 hours) and held at cold temperatures ( $\sim 4^{\circ}\text{C} \pm 2^{\circ}\text{C}$ ) 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 shipping or NEON lab processing is 72 hours.

### 3.3 Sampling Design Changes

- Seepage lakes are no longer sampled at their littoral1 and littoral2 locations, implemented 2018
- ANC samples are only taken monthly at all aquatic sites, implemented 2018
- New collection methodology for low-yield wells, implemented 2018
- New collection methodology for permafrost sites, implemented 2019
- Started measuring groundwater field pH measurements, implemented 2019
- Stopped reporting total suspended solids dry mass (total suspended solids normalized to filtered volume remains the same), implemented 2021
- Stopped collecting and analyzing particulate carbon and nitrogen filters in this data product, implemented 2020
- Started publishing particulate carbon and nitrogen concentrations collected as part of the AOS Protocol and Procedure: ASI - Stable Isotope Sampling in Surface and Ground Waters (AD[12]), implemented 2022

### 3.4 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 precision 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 acquired or a change is detected in analytical precision). Additionally, NEON's Calibration/Validation depart-

ment has regular procedures for auditing the quality assurance of external laboratories and their reports are available to data users.

### 3.4.1 Laboratory analyses dependent on sampleCondition

When cold chain is broken and/or hold time is exceeded during storage, the laboratory will report data for only anions (Br, Cl, F, SO<sub>4</sub>) and cations (Ca, Mg, Na, K, Mn, Fe, Si) for an affected sample/sub-sample.

## 3.5 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 (DP0.20090.001) (AD[03]), NEON Raw Data Validation for Chemical properties of surface water, Level 0 (DP0.20093.001) (AD[04]), NEON Raw Data Validation for Water Chemistry External Lab Data (DP0.20286.001) (AD[05]), and NEON Raw Data Validation for Water Chemistry External Lab Summary Data (DP0.20287.001) (AD[06]). 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 (DP1.20093.001) (AD[07]) and NEON Data Variables for Chemical properties of groundwater (DP1.20092.001) (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.

## 3.6 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 'c0', and 'c1', 'c2', and 'c3', as needed (if center is stratified), with 'c1' being the top layer. As of July 2019, Toolik Lake (TOOK), also collects surface water samples at the inflow and outflow, where 'in' and 'ot' designations are used to indicate inflow and outflow sampling locations.

Prior to January 2018, samples were also collected at two littoral locations designated as 'in' and 'ot'. 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 sampleID will be designated as 're' and 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, the station ID is 'w1' through 'w8' with the number indicating the designated well number. 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>

### 3.7 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 stream and river sites, 12 per lake, and 2 per groundwater well subset per site. At stream and river sites, alkalinity is measured 26 times, and ANC is measured monthly at the NEON Support office.

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>

### 3.8 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.

### 3.9 Product Instances

The NEON Observatory contains 34 aquatic sites, consisting of 24 wadeable streams, 3 non-wadeable streams (rivers), and 7 lakes. Up to 8 groundwater wells are installed at 30 of the 34 sites.

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 wadeable streams, there will be up to 26 unique sample records per site per year, non-wadeable streams will yield up to 26 records per



Title: NEON User Guide to Chemical properties of surface water (DP1.20093.001) and Chemical properties of groundwater (DP1.20092.001)	Date: 08/10/2022
Author: Keli Goodman	Revision: D.1

year, lakes will yield 12-36 records per year, depending on stratification, and groundwater will produce 6-8 records per year. Observatory-wide, this will yield a total of 1026-1194 water chemistry records per year.

NOTE: Replicate samples may be taken of a small percentage of 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.

### 3.10 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 blank, 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 aberrations 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.

swc\_asiPOMFieldData\_pub.csv - > One record expected per parentSampleID, generates isotopePOMSampleID, isotopePOMRep2SampleID, and corresponding field data that corresponds to filters collected for particulate carbon and nitrogen analysis.

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

(Deprecated 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)

**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.

Data downloaded from the NEON Data Portal are provided in separate data files for each site and month requested. The neonUtilities R package contains functions to merge these files across sites and months into a single file for each table described above. The neonUtilities package is available from the Comprehensive R Archive Network (CRAN; <https://cran.r-project.org/web/packages/neonUtilities/index.html>) and can be installed using the install.packages() function in R. For instructions on using neonUtilities to merge NEON data files, see the Download and Explore NEON Data tutorial on the NEON website: <https://www.neonscience.org/download-explore-neon-data>

### 3.11 Special Considerations

Particulate Carbon and Nitrogen returned with laboratoryName = "EcoCore\_CSU" are in mg, and should be divided by the sampleVolumeFiltered in the swc\_fieldData\_pub to calculate mg/mL and converted as needed.

Samples analyzed for UV Absorbance (254 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.

When comparing grab samples to sensor data, please note that water chemistry grab samples are collected at the downstream sensor set (sensor set 2) in streams and at the buoy location in rivers and lakes. Given storage times before analysis and shipping logistics, we suggest that users interested in pH and specific conductance use the continuous sensor data or the initialSamplepH in the swc\_domainLabData expanded package and specificConductance in the swc\_fieldSuperParent tables. We encourage the use of the continuous sensor data whenever possible.

External laboratory summary tables should be referenced for mdls. Negative values and values < mdls, should be treated appropriately by the user.

For groundwater chemistry samples, please note the sampleHeadspace fields in the gwc\_fieldData table. Occasionally headspace is unavoidable when only a low volume is available from groundwater sampling. This atmospheric exposure may influence some of the analyte results in the filtered sample. Similarly, the BubbleFree fields will indicate if air was introduced to a sample during collection with the needle method. Note that aquatic sites in Alaska are directed to leave headspace in the filtered groundwater sample bottle in order to prevent bottle breakage due to possible freezing in transport. These sites collect an additional subsample without headspace for Dissolved Inorganic Carbon and pH laboratory analyses.

## 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 (DP0.20090.001), NEON Raw Data Validation for Chemical properties of surface water, Level 0 (DP0.20093.001), NEON Raw Data Validation for Water Chemistry External Lab Data (DP0.20286.001), and NEON Raw Data Validation for Water Chemistry External Lab Summary Data (DP0.20287.001), 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[16]).

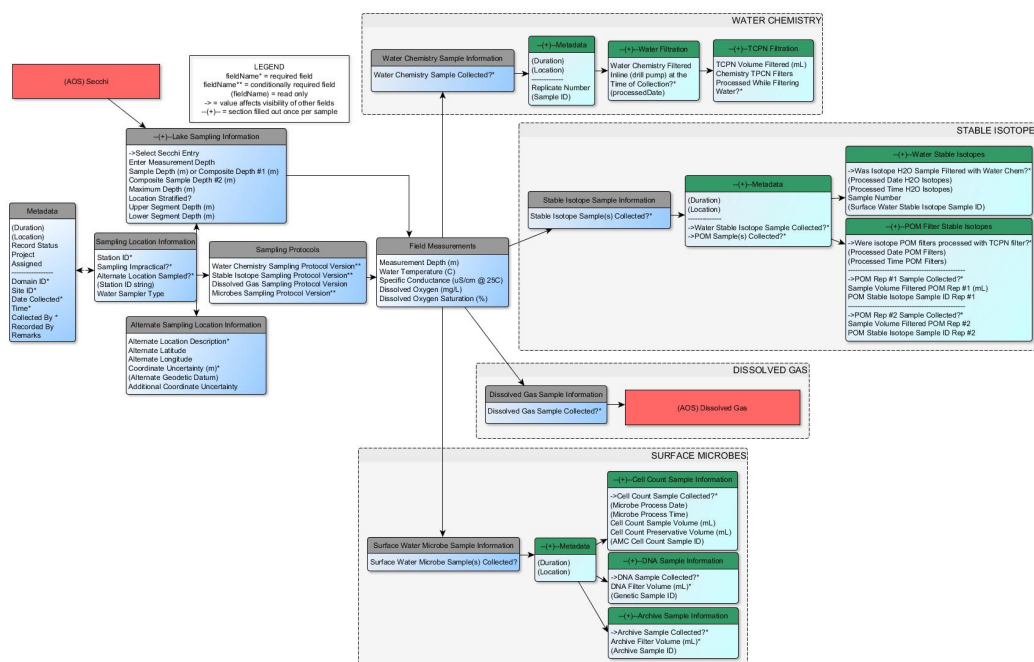


Figure 2: Schematic of the applications used by field technicians to enter water chemistry field data

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.

## 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[15]).

## 4.3 Data Revision

All data are provisional until a numbered version is released. Annually, NEON releases a static version of all or almost all data products, annotated with digital object identifiers (DOIs). The first data Release was made in 2021. 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 Issue Log section of the data product landing page 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.

Table 1: Descriptions of the dataQF codes for quality flagging

fieldName	value	definition
dataQF	Preliminary method: UV absorbance not water blank sub- tracted	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
dataQF	ccvAboveCriteria	The recovery of the continuing calibration verification standard was above 110 percent.
shipmentTimeRange	more than 72 hours after collection	Indicates time delay between sample collection and shipment
gwwDICBubbleFree	N	When a syringe is used for groundwater sampling at permafrost sites, this indicates that air may have been introduced into the DIC groundwater subsample
gwwALKBubbleFree	N	When a syringe is used for groundwater sampling at permafrost sites, this indicates that air may have been introduced into the ALK groundwater subsample
gwwFiltSampleHeadspace	Y	Indicates headspace the filtered groundwater subsample
gwwALKSampleHeadspace	Y	Indicates headspace the ALK groundwater subsample
belowDetectionQF	BDL	Indicates a value above 0 was reported for analyteConcentration, but it was below the method detection limit (MDL).
belowDetectionQF	ND	Indicates no value (i.e., blank or NA value) was reported for analyteConcentration due to a negative reading from the laboratory instrument.

Records of land management activities, disturbances, and other incidents of ecological note that may have a potential impact are found in the Site Management and Event Reporting data product (DP1.10111.001)

#### 4.5 Note about blank or NA values in analyteConcentration

Values can be blank (or NA when downloaded using neonUtilities) for analyteConcentration for a few different reasons. Users are encouraged to look at entries in sampleCondition and belowDetectionQF for details on the reason for a particular record. Records with entries other than “GOOD” or “OK” in sample-



Condition likely indicate that data for a given analyte were not collected and users cannot infer a value for the analyte to include in their analysis. For records with “ND” values for belowDetectionQF, users can infer that the analyte concentration is very low and may want to include the records in their analysis.

#### 4.6 Note for Alkalinity and Acid Neutralizing Capacity

“Hydroxide, carbonate, and bicarbonate concentrations cannot be negative. Alkalinity and acid neutralizing capacity (ANC), however, may be negative because they are a measure of the sample’s H+ deficiency relative to a solution of carbon dioxide in water. A negative alkalinity or ANC simply indicates the presence of some amount of mineral acidity. Report negative values of alkalinity and ANC with the same number of significant figures you would use if the values were positive.” quoted from: <https://or.water.usgs.gov/alk/reporting.html>

Also see AOS Commissioning test report: Surface water chemistry data quality (NEON.DOC.004844) and USGS TWRI Book 9 Chapter A6 for more information about reproducibility and uncertainty associated with these measurements.

## 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).