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

PREPARED BY	**ORGANIZATION**
Keli Goodman	AOS
Caren Scott	AOS
Samantha Weintraub	TOS
Zachary Nickerson	AOS
Nora Catolico	AOS

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.
D.1	08/10/2022	Updates for external lab quality flags.
E	01/03/2023	Updates for external laboratory sample conditions.
F	07/12/2024	Updates to section 3.11 to clarify uncertainty and timeline of pH and specificConductance measurements.
F.1	08/27/2024	Updates to include details of sampling frequency and method changes.
G	04/17/2025	Updates to include details of key groundwater measurements, external lab changes, addition of UV-Vis spectrum table, and new dataQF option. Added information about the new neonUtilities Python package.
H	03/04/2026	Updates describing changes to site-all blanks table and data quality automated data processing steps.

TABLE OF CONTENTS

1 DESCRIPTION **1**

1.1 Purpose . . . 1

1.2 Scope . . . 1

2 RELATED DOCUMENTS AND ACRONYMS **2**

2.1 Associated Documents . . . 2

2.2 Acronyms . . . 3

3 DATA PRODUCT DESCRIPTION **4**

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 . . . 7

3.4.1 Laboratory analyses dependent on sampleCondition . . . 7

3.5 Variables Reported . . . 8

3.5.1 Note on reported nutrient units . . . 9

3.6 Spatial Resolution and Extent . . . 9

3.7 Temporal Resolution and Extent . . . 10

3.8 Associated Data Streams . . . 11

3.9 Product Instances . . . 11

3.10 Data Relationships . . . 11

3.11 Special Considerations . . . 12

3.11.1 Issues with UV-Vis collected between 10/21/2014 and 08/11/2015 . . . 12

3.11.2 Data table format change for data collected before 07/30/2019 . . . 13

3.11.3 Related sensor data recommendations . . . 13

3.11.4 Method detection limit information . . . 13

3.11.5 Groundwater sample headspace information . . . 13

3.11.6 Downloading data for blank samples . . . 14

4 DATA QUALITY **14**
4.1 Data Entry Constraint and Validation 14
4.2 Automated Data Processing Steps 15
4.3 Data Revision 15
4.4 Quality Flagging 16
4.5 Note about blank or NA values in analyteConcentration 17
4.6 Note for Alkalinity and Acid Neutralizing Capacity 17
4.7 Note about applying external lab summary information in analyses 17

5 REFERENCES **17**

LIST OF TABLES AND FIGURES

Table 1 Descriptions of the dataQF codes for quality flagging 16

Figure 1 Generic layout of surface and groundwater sampling locations in wadeable streams, non-wadeable streams (i.e. rivers), and lake sites. See AD[06] for further details. 5
Figure 2 Key groundwater measurements for NEON wells. 10
Figure 3 Schematic of the applications used by field technicians to enter water chemistry field data 15

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 (DP1.20093.001) and Chemical properties of groundwater (DP1.20092.001) - 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[09]). 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]	DP0.20090.001 _dataValidation.csv	NEON Raw Data Validation for Water chemistry, isotopes, dissolved gas, and microbes sampling, Level 0 (DP0.20090.001)
AD[04]	DP0.20093.001 _dataValidation.csv	NEON Raw Data Validation for Chemical properties of surface water, Level 0 (DP0.20093.001)
AD[05]	DP0.20286.001 _dataValidation.csv	NEON Raw Data Validation for Water Chemistry External Lab Data (DP0.20286.001)
AD[06]	DP0.20287.001 _dataValidation.csv	NEON Raw Data Validation for Water Chemistry External Lab Summary Data (DP0.20287.001)
AD[07]	DP1.20093.001 _variables.csv	NEON Data Variables for Chemical properties of surface water (DP1.20093.001)
AD[08]	DP1.20092.001 _variables.csv	NEON Data Variables for Chemical properties of groundwater (DP1.20092.001)
AD[09]	NEON.DOC.002905	AOS Protocol and Procedure: Water Chemistry Sampling in Surface Waters and Groundwater
AD[10]	NEON.DOC.001152	NEON Aquatic Sampling Strategy
AD[11]	NEON.DOC.001886	AOS Protocol and Procedure: ASI - Stable Isotope Sampling in Surface and Ground Waters
AD[12]	NEON.DOC.000008	NEON Acronym List
AD[13]	NEON.DOC.000243	NEON Glossary of Terms
AD[14]	NEON.DOC.004825	NEON Algorithm Theoretical Basis Document: OS Generic Transitions
AD[15]	Available on NEON data portal	NEON Ingest Conversion Language Function Library
AD[16]	Available on NEON data portal	NEON Ingest Conversion Language
AD[17]	NEON.DOC.005424	NEON Algorithm Theoretical Basis Document: OS Data Quality Control

2.2 Acronyms

Acronym	Definition
ALK	Alkalinity
ANC	Acid Neutralizing Capacity
GWC	Ground Water Chemistry
SWC	Surface Water Chemistry
PCN	Particulate carbon and nitrogen subsample
POM	Particulate organic matter subsample
FIL	Filtered subsample
RAW	Unfiltered subsample
NUT	Nutrient subsample
DIC	Dissolved inrorganic carbon subsample
TSS	Total suspended solids
TDS	Total dissolved solids
CO_3	Bicarbonate
HCO_3	Carbonate
Ca	Calcium
Mg	Magnesium
Na	Sodium
K	Potassium
SO_4	Sulfate
Cl	Chloride
Mn	Manganese
Fe	Iron
Br	Bromide
F	Flouride
Si	Silicon
UV-Vis	Ultraviolet-visible
TOC	Total organic carbon
TN	Total nitrogen
TDN	Total dissolved nitrogen
DOC	Dissolved organic carbon
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[09]). 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 (specific conductance, pH, 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, UV-Visible absorbance, and carbon concentrations.

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 1, upper left panel). 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 1, upper right panel). 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 seepage lakes, samples are taken from the buoy location (Figure 1, lower left panel). In lakes with a true inflow and outflow, samples are collected near lake buoy, inflow, and outflow sensor infrastructure (Figure 1, lower right panel).

Most aquatic sites have up to eight shallow (<100 ft depth) groundwater observation wells on the perimeter of sampled waterbodies (Figure 1, red boxes). 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 up gradient and 2 down gradient of the lake water table. 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/

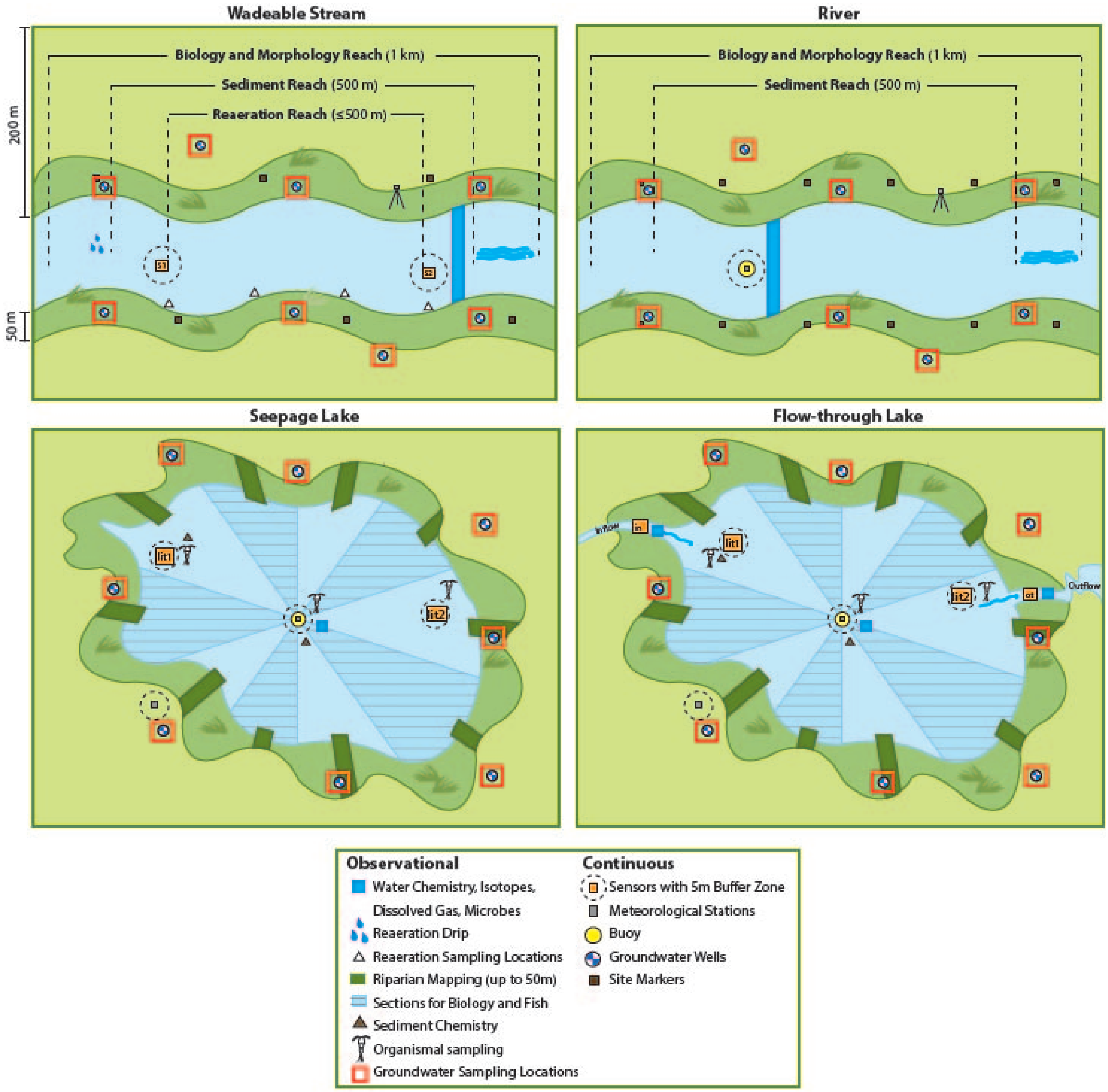

Figure 1: Generic layout of surface and groundwater sampling locations in wadeable streams, non-wadeable streams (i.e. rivers), and lake sites. See AD[06] for further details.

3.2 Temporal Sampling Design

For streams, water chemistry sampling occurs up to 26 times per year, which is monthly plus 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 +/- 2 days 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 3-4 hours) and held at cold temperatures (~ 4°C +/- 2°C) until they are either shipped to a water chemistry lab, frozen for nutrient analysis, 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. Nutrient samples, designated with a .NUT sampleID, are frozen as soon as possible after collection and shipped quarterly to the external laboratory.

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
- Reduce replicate sampling frequency in lakes to one time per year and add additional field blanks, implemented 2019
- Stopped collecting and analyzing particulate carbon and nitrogen filters in this data product, implemented 2020
- Stopped reporting total suspended solids dry mass (total suspended solids normalized to filtered volume remains the same), implemented 2021
- Discontinued particulate carbon and nitrogen analysis in groundwater wells, implemented 2021
- Implemented new laboratory method to realize a lower method detection limit for nutrient analyses, implemented 2021
- 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[11]), implemented 2022
- Discontinued external lab pH, HCO_3, and CO_3 measurements. Beginning with RELEASE-2025, all historical external lab pH, HCO_3 and CO_3 samples will be deleted from future releases. Historical samples included in older releases should not be used. HCO_3 and CO_3 can alternatively be estimated

using either external lab dissolved inorganic carbon (DIC) or domain lab alkalinity (ALK) and the corresponding field sensor pH or domain lab pH, if sensor data is not available, which are reported in these data products, implemented August 2024.
- The external lab now reports UV-visible absorbance spectra in addition to individual wavelength values. See the wc_externalLabAbsorbanceScan table for spectral data, implemented November 2023.

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). For more details, see the data quality note in section 4.7 about interpreting summary file information. 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.

3.4.1 Laboratory analyses dependent on sampleCondition

If the cold chain is broken during shipment or storage prior to analysis:

- If a sample is received $\leq 6\ ^\circ$C there are no flags and all analytes are analyzed
- If a sample is received > 6 and $\leq 10\ ^\circ$C a sampleCondition of "Sample Received Warm But Analyzable" is populated and all analytes are analyzed and reported
- If a sample is received $> 10\ ^\circ$C a sampleCondition of "Sample Received Warm But Analyzable" is populated and analysis is performed for all PCN analytes, TSS only for the RAW sample, and Ca, Mg, Na, K, SO_4, Cl, Mn, Fe, Br, F, Si, TDS, and UV-Vis absorbance for the FIL sample
- If a sample is received $> 10\ ^\circ$C a sampleCondition of "Sample Received Warm But Not Analyzable" is populated and analysis is not performed for DIC for the DIC sample, TOC and TN for the RAW sample, and TDN, DIC, and DOC for the FIL sample

If analytical hold times are exceeded during shipment or storage:

- If a sampleCondition of "Hold Time Exceeded" is populated no analyses are performed on the DIC or RAW samples and Ca, Mg, Na, K, SO_4, Cl, Mn, Fe, Br, F, Si, and UV absorbance are performed on the FIL sample. The remarks field may also contain text indicating an approximate hold time while kept chilled or frozen prior to analysis.

If both hold time is exceeded and cold chain is broken during storage:

- If the cold chain is broken during storage and the hold time is exceeded a sampleCondition including “Cold Chain Broken During Storage” and “Hold Time Exceeded” will be populated and only Ca, Mg, Na, K, SO_4, Cl, Mn, Fe, Br, F, and Si will be analyzed on the FIL sample.

If there is insufficient volume to conduct all analyses:

- Sample collection is prioritized as follows for surface water samples and groundwater samples collected in Domains 01 through D17
  1) particulates (POM, surface water only)
  2) filtered (FIL)
  3) filtered nutrient (FIL.NUT)
  4) alkalinity (ALK)
  5) unfiltered (RAW)
  6) unfiltered nutrient (RAW.NUT)
- Sample collection is prioritized as follows for the needle method, which is used for groundwater samples collected in Tundra and Taiga Domains only (D18/19):
  1) alkalinity (ALK)
  2) field pH reading
  3) bubble-free dissolved inorganic carbon (DIC)
  4) filtered (FIL)
  5) filtered nutrient (FIL.NUT)
  6) unfiltered (RAW)
  7) unfiltered nutrient (RAW.NUT)

If a sample cannot be analyzed for a particular analyte given any of the conditions above, the external laboratory should report a blank analyteConcentration and a sampleCondition explaining the reason for returning a blank record.

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 Geoid12A geoid model for its vertical reference surface. 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.

3.5.1 Note on reported nutrient units

NEON nutrient data is reported as the concentration of the principal element, either nitrogen or phosphorus. However, results from nutrient analyses can be reported in two ways – as the whole compound or as the principal element in the compound. For example, nitrate may be reported as nitrate (NO3) or nitrate as nitrogen (NO3-N). When comparing data from different sources (within NEON or across multiple data producers outside of NEON), it is important to compare like with like and convert the results, if needed.

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 related to the NEON Aquatic Observation System sampling locations can be found on the NEON science webpage at https://www.neonscience.org/data-samples/data/spatial-data-maps. 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. NEON wells are drilled to refusal and users can find key measurements (Figure 2) such as the total well depth (static), well casing to water table depth (at the time of sampling), and water column height (at the time of sampling) in the xxx_fieldSuperParent table. Two sensor-based data products Elevation of Groundwater (DP1.20100.001) and Elevation of Surface Water (DP1.20016.001) provide continuous data on the actual elevation of the water table at locations across each site.

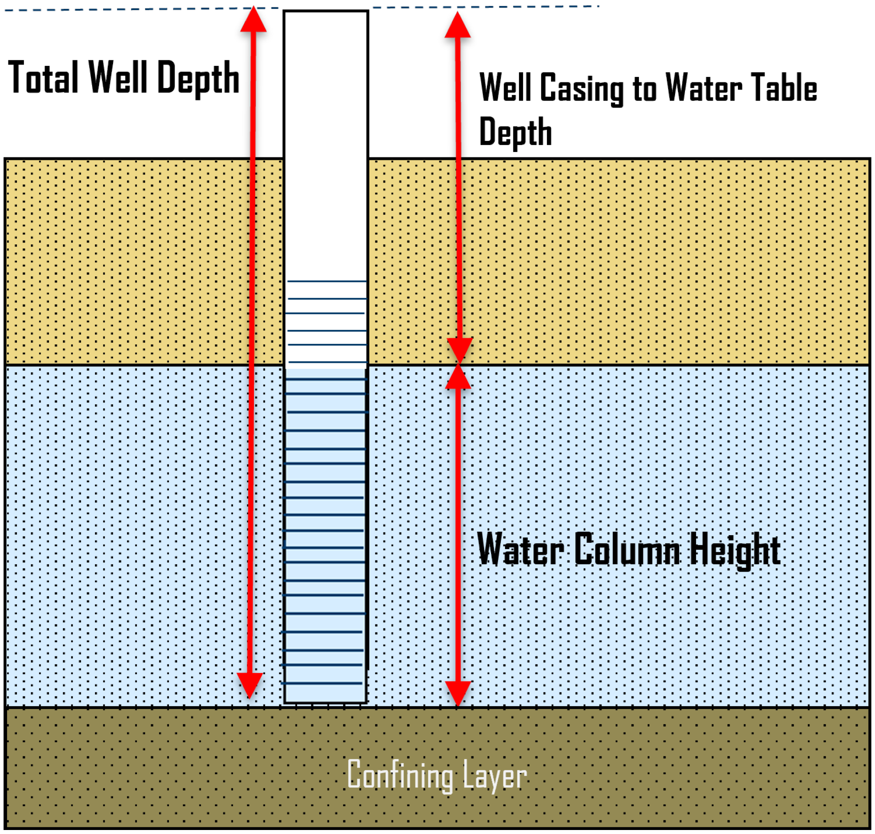

Figure 2: Key groundwater measurements for NEON wells.

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. At lake sites, alkalinity and ANC are both 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 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 an analytical 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 analytical 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_externalLabAbsorbanceScan_pub.csv or gwc_externalLabAbsorbanceScan_pub.csv - > One record expected per sampleID x wavelength x spectrumReplicateID, two spectrumReplicateIDs expected per sampleID, i.e., duplicate spectra are reported per sample across a wavelength range of 220 - 600 (381 records per spectrum).

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 package in R and the neonutilities package in Python contain functions to merge these files across sites and months into a single file for each table. The neonUtilities R 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. The neonutilities package in Python is available on the Python Package Index (PyPi; https://pypi.org/project/neonutilities/) and can be installed using pip. For instructions on using the package in either language 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

3.11.1 Issues with UV-Vis collected between 10/21/2014 and 08/11/2015

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.

3.11.2 Data table format change for data collected before 07/30/2019

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.

3.11.3 Related sensor data recommendations

Given the potential for additional uncertainty arising from sample collection/processing, shipping, and storage time prior to analysis, we strongly recommend using the in-situ Water quality (DP1.20288.001) sensor data whenever possible, particularly for specificConductance and pH. Alternatively, users may use the field measurements of specificConductance and pH included in the swc_fieldSuperParent table, which are measured at the time of sample collection. Alkalinity and ANC titrations are typically performed within 24 hours of sample collection and the initialSamplepH found in the swc_domainLabData expanded package is another option for users when other data is not available. Measurements in the swc_externalLabDataByAnalyte table have the longest time between sample collection and measurement and may not reflect conditions in the stream as closely as the above mentioned measurements. When comparing water chemistry grab samples to sensor-based water quality data (DP1.20288.001), note that water chemistry samples are collected at the downstream sensor set (horizontalPosition 102 or 112) in streams and at the buoy location (horizontalPosition 103) in rivers and lakes. Because of lower uncertainty, we encourage the use of continuous sensor data whenever possible.

3.11.4 Method detection limit information

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

3.11.5 Groundwater sample headspace information

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 laboratory analyses.

3.11.6 Downloading data for blank samples

The expanded package contains the externalLabBlanksByAnalyte table. This table will contain all blanks collected for a give site regardless of the date when they were collected and the dates chosen for the data download. For example, when downloading current data the download package will contain blanks collected in previous years along with more recently collected blank data.

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 3). 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[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.

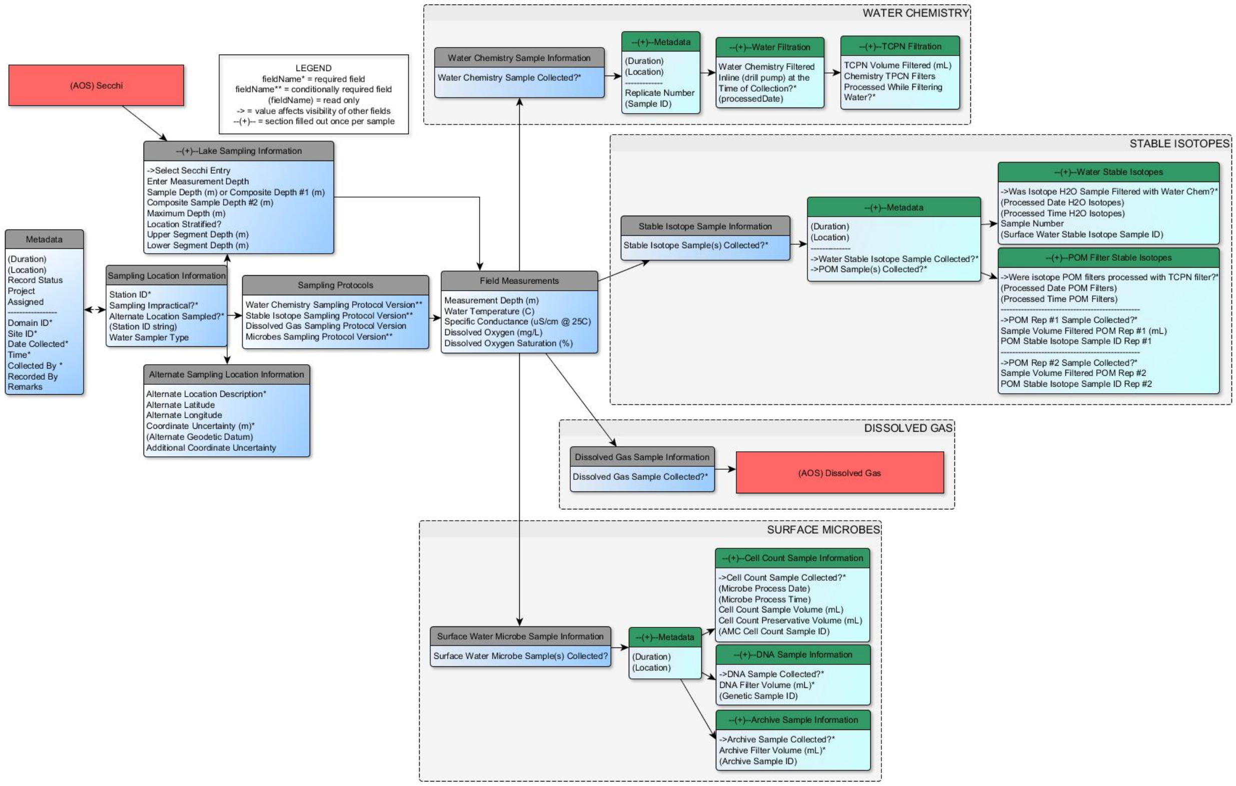

Figure 3: Schematic of the applications used by field technicians to enter water chemistry field data

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[15]).

Published data are reviewed for completeness, timeliness, and validity using an internal set of tests and metrics, as detailed in the NEON Algorithm Theoretical Basis Document: OS Data Quality Control (AD[17]). These quality tests are used to guide process improvements, audits of analytical facilities, and data updates, but do not generate quality flags in published data.

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 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
dataQF	ccvAboveCriteria	The recovery of the continuing calibration verification standard was above 110 percent.
dataQF	likelySampleContamination	An erroneously high value was measured.
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 sampleCondition 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.

4.7 Note about applying external lab summary information in analyses

Many NEON sites are oligotrophic which makes analyzing nutrient concentrations challenging. As NEON has collected more data, we have worked with external labs to improve the quality of our nutrient data by improving the sensitivity of the methods over time. Thus, the detection limit, precision, and uncertainty of the methods have changed over time and users should refer to the lab summary file to ensure they are using the correct values. In cases where there are values from different labs for the same sample and analyte, users should use the data from the lab with the lower detection limit and uncertainty. Data analyzed using lower detection limits are likely to be more accurate for samples with lower nutrient concentrations.

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).