

<i>Title:</i> NEON User Guide for Aquatic Benthic Microbe Collection (NEON.DP0.20270)	<i>Date:</i> 12/14/2017
<i>Author:</i> Stephanie Parker	<i>Revision:</i> A

NEON USER GUIDE FOR AQUATIC BENTHIC MICROBE COLLECTION (NEON.DP0.20270)

PREPARED BY	ORGANIZATION	DATE
Stephanie Parker	AQU	12/14/2017
Lee Stanish	FSU	12/14/2017

<i>Title:</i> NEON User Guide for Aquatic Benthic Microbe Collection (NEON.DP0.20270)	<i>Date:</i> 12/14/2017
<i>Author:</i> Stephanie Parker	<i>Revision:</i> A

CHANGE RECORD

REVISION	DATE	DESCRIPTION OF CHANGE
A	11/13/2017	Initial Release

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	2
3 DATA PRODUCT DESCRIPTION	3
3.1 Spatial Sampling Design	3
3.2 Temporal Sampling Design	4
3.3 Variables Reported	4
3.4 Spatial Resolution and Extent	5
3.5 Temporal Resolution and Extent	5
3.6 Associated Data Streams	6
3.7 Product Instances	6
3.8 Data Relationships	6
3.9 Special Considerations	6
4 DATA QUALITY	7
4.1 Data Entry Constraint and Validation	7
4.2 Automated Data Processing Steps	7
4.3 Data Revision	7
4.4 Quality Flagging	7
4.5 Analytical Facility Data Quality	9
5 REFERENCES	9

LIST OF TABLES AND FIGURES

Figure 1 Generic aquatic site layout for benthic microbe sampling in wadeable streams, with benthic microbe sampling locations in red. Benthic samples may be collected anywhere within the 1 km permitted stream reach.	4
Figure 2 Schematic of the applications used by field technicians to enter periphyton and phytoplankton field data	8

<i>Title:</i> NEON User Guide for Aquatic Benthic Microbe Collection (NEON.DP0.20270)	<i>Date:</i> 12/14/2017
<i>Author:</i> Stephanie Parker	<i>Revision:</i> A

1 DESCRIPTION

1.1 Purpose

This document provides an overview of the field data included in several NEON Level 1 data products, the quality controlled product generated from raw Level 0 field data, and associated metadata. In the NEON data products framework, the raw data collected in the field, for example, the type of benthic microbe sample collected, 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 (AD[01]).

1.2 Scope

This document describes the steps needed to generate the L1 data product Periphyton and phytoplankton collection - the collection of periphyton using multiple benthic sampling methods in lakes, non-wadeable streams, and wadeable streams, as well as seston and phytoplankton collected in the water columns of wadeable streams (seston) and lakes or rivers (phytoplankton), and associated metadata from input data. This document also provides details relevant to the publication of the data products via the NEON data portal, with additional detail available in the file, NEON Data Variables for Benthic Microbe Marker Gene Sequences (NEON.DP1.20280) (AD[04]), 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: Aquatic Microbial Sampling (AD[06]). The raw data that are processed in this document are detailed in the file, NEON Raw Data Validation for Aquatic Benthic Microbe Collection, Level 0 (NEON.DP0.20270) (AD[03]), 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., '10033') 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 Level 1, Level 2 and Level 3 Data Products Catalog
AD[03]	NEON.DP0.20280.001_dataValidation.csv	NEON Raw Data Validation for Aquatic Benthic Microbe Collection, Level 0 (NEON.DP0.20270)
AD[04]	NEON.DP1.20280.001_variables.csv	NEON Data Variables for Benthic Microbe Marker Gene Sequences (NEON.DP1.20280)
AD[05]	NEON.DP1.20279.001_variables.csv	NEON Data Variables for Benthic Microbe Metagenomic Sequences (NEON.DP1.20279)
AD[06]	NEON.DP1.20277.001_variables.csv	NEON Data Variables for Benthic Microbe Group Abundances (NEON.DP1.20277)
AD[07]	NEON.DOC.001152	NEON Aquatic Sampling Strategy
AD[08]	NEON.DOC.003044	AOS Protocol and Procedure: Aquatic Microbial Sampling
AD[09]	NEON.DOC.000008	NEON Acronym List
AD[10]	NEON.DOC.000243	NEON Glossary of Terms
AD[11]	OS_Generic_Transitions.pdf	NEON Algorithm Theoretical Basis Document: OS Generic Transitions
AD[12]	Nicl Language.pdf	NEON's Ingest Conversion Language (NICL) specifications

2.2 Acronyms

Acronym	Definition
NAWQA	National Water Quality Assessment (USGS)

<i>Title:</i> NEON User Guide for Aquatic Benthic Microbe Collection (NEON.DP0.20270)	<i>Date:</i> 12/14/2017
<i>Author:</i> Stephanie Parker	<i>Revision:</i> A

3 DATA PRODUCT DESCRIPTION

This data product includes measurements for benthic microbe collection, including benthic microalgae and biofilms, collected from a variety of benthic surfaces including cobble [epilithon], silt [epipelon], sand [epipsammon], woody debris [epixylon], and plant surfaces [epiphyton]). These data are related to the NEON Grand Challenge areas of Biodiversity and Biogeochemistry and include additional data about the microbial communities in streams. Benthic microbe samples are collected along with periphyton (algae) samples three times per year at each NEON wadeable stream site (AD[07]). Sampling dates are based on a combination of variables, including hydrology in streams or ice on/ice off dates in lakes, accumulated degree days (temperature), and riparian greenness (phenology). For additional information, see the NEON Aquatic Sampling Strategy (AD[07]) and AOS Protocol and Procedure: Aquatic Microbial Sampling (AD[08]).

Data are organized into tables for field data collected by NEON technicians and external lab data returned by the external lab (see AD[04, 05, and 06] for data publication). Field data contains metadata on sample time, location, type of habitat and substratum, and the type of sampler used, which determines the benthic area sampled.

3.1 Spatial Sampling Design

Benthic microbe sampling occurs along a 1 km reach of each wadeable stream. Sites are sampled using a percent-based macrohabitat approach (after Moulton et al. 2002), in which the habitats that are sampled depends on the percent cover of habitats present at the respective NEON aquatic site (Figure 1). Habitats sampled are riffles, runs, pools, and step pools. All samples are collected from the surface of the natural substratum present in each macrohabitat. Field protocols differ depending on substratum being sampled. For example, riffles and runs often have cobble/pebble substratum, while pools may have silt or sand substrata. At some sites with sandy or silty bottoms, the majority of the periphyton community may be colonizing the leaves of aquatic plants (epiphytes) or woody debris, thus plant or woody debris substrata are sampled rather than sampling sandy/silty substrata, which can be sparsely populated. Appropriate site-specific sampling procedures are determined prior to sampling following NAWQA protocols (Moulton et al. 2002) and presented in site-specific AOS documents. See sampling protocol AOS Protocol and Procedure: Aquatic Microbial Sampling (AD[08]) for additional details on strategy and SOPs.

For each sampling event at a wadeable stream site, five benthic microbe samples are collected in the dominant habitat type and three samples are collected in the second-most dominant habitat type, for a total of eight samples on a given sampling date at a site. Samples are spread out along the 1 km reach so that ideally no two samples are collected within the same habitat unit.

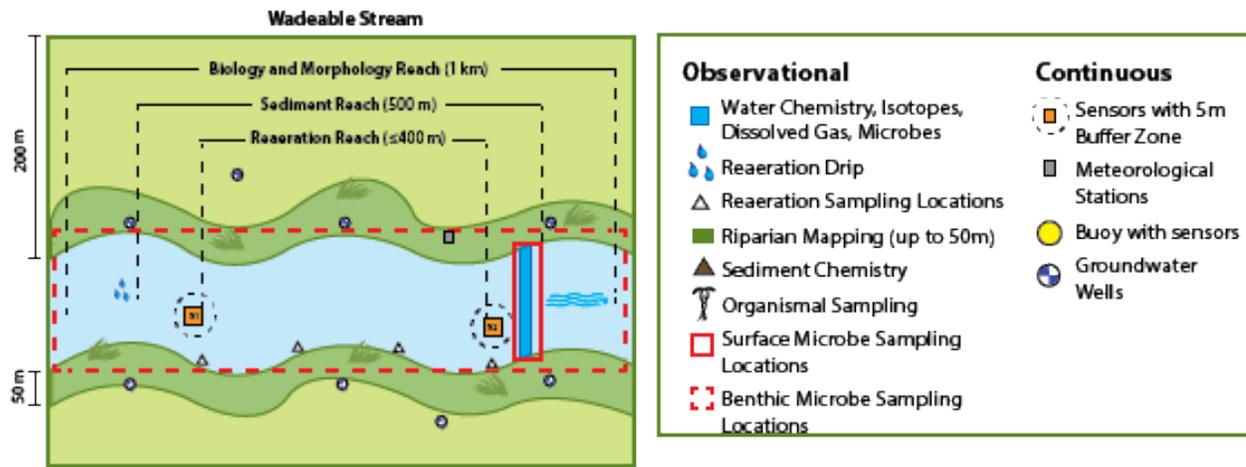


Figure 1: Generic aquatic site layout for benthic microbe sampling in wadeable streams, with benthic microbe sampling locations in red. Benthic samples may be collected anywhere within the 1 km permitted stream reach.

3.2 Temporal Sampling Design

Benthic microbe sampling occurs three times per year in order to capture presence and abundance of multiple species and growth forms of periphyton. Timing of sampling is site-specific and determined based on historical data, including stream discharge, air temperature, and riparian greenness. Specific details on sample dates and strategy are provided in the NEON Aquatic Sample Strategy Document (AD[07]). Sample bout 1 is an early-season date, representing a period of rapid biomass accumulation after winter, typically after ice-off (where applicable) and prior to leaf out. Sample bout 2 targets low flows and high light (mid-summer) at each site. Sample bout 3 represents the late growing season (typically autumn) at each site during leaf-fall. These dates differ on a site-by-site basis but are always based on the same strategy. Sampling should occur at base-flow conditions, and will not occur directly following a flood in the stream (>1.5 x base flow; Biggs et al. 1999) or under ice. A period of 14-days is allowed after a flood event for periphyton/benthic microbes to recolonize before sampling occurs. See NEON Aquatic Sampling Strategy (AD[07]), AOS Protocol and Procedure: Aquatic Microbial Sampling (AD[08]) for additional details.

3.3 Variables Reported

All variables reported from the field or laboratory technician (L0 data) are listed in the file, NEON Raw Data Validation for Aquatic Benthic Microbe Collection, Level 0 (NEON.DP0.20270) (AD[03]). All variables reported in the published data (L1 data) are also provided separately in the files, NEON Data Variables for Benthic Microbe Marker Gene Sequences (NEON.DP1.20280) (AD[04]), NEON Data Variables for Benthic Microbe Metagenomic Sequences (NEON.DP1.20279) (AD[05]), and NEON Data Variables for Benthic Microbe Group

<i>Title:</i> NEON User Guide for Aquatic Benthic Microbe Collection (NEON.DP0.20270)	<i>Date:</i> 12/14/2017
<i>Author:</i> Stephanie Parker	<i>Revision:</i> A

Abundances (NEON.DP1.20277) (AD[06]). See separate data product user guides for those data products.

Field names have been standardized with Darwin Core terms (<http://rs.tdwg.org/dwc/>; accessed 12 August 2017), the Global Biodiversity Information Facility vocabularies (<http://rs.gbif.org/vocabulary/gbif/>; accessed 12 August 2017), the VegCore data dictionary (<https://projects.nceas.ucsb.edu/nceas/projects/bien/wiki/VegCore>; accessed 12 August 2017), where applicable. NEON Aquatic Observation System (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.

3.4 Spatial Resolution and Extent

Wadeable stream Each benthic microbe sample will represent a patch of stream bottom within the 1 km permitted wadeable stream reach. The exact location of each sample will not be tracked as it is intended to represent the overall habitat (locationID = “reach”). Up to two different habitats are sampled at each site to account for the variability or “patchiness” among habitats. Field replicate samples are collected in each habitat, with five samples collected in the dominant habitat and three samples collected in the secondary habitat during each sampling bout (Figure 1).

Overall, this results in a spatial hierarchy of:

habitatType (habitat type sampled) → **locationID** (ID of the sampling location) → **siteID** (ID of NEON site) → **domainID** (ID of a NEON domain).

3.5 Temporal Resolution and Extent

The finest temporal resolution that benthic microbe data are tracked is per sampling day. All 8 samples are collected within a single day at a particular site. A suite of other biological sampling occurs at the site during the same ~30 day bout, and periphyton samples are collected at the same time and location as benthic microbe samples. Three sampling bouts occur per site per year. The finest resolution at which temporal data are reported is at **collectDate**, the date and time of day when the samples were collected in the field.

The NEON Data Portal provides data in monthly files for query and download efficiency. Queries including any part of a month will return data from the entire month. Code to stack files across months is available here: <https://github.com/NEONScience/NEON-utilities>.

<i>Title:</i> NEON User Guide for Aquatic Benthic Microbe Collection (NEON.DP0.20270)	<i>Date:</i> 12/14/2017
<i>Author:</i> Stephanie Parker	<i>Revision:</i> A

3.6 Associated Data Streams

All of the above data products are also loosely related to Aquatic General Field Metadata collected on the same sampling day (NEON.DOC.001646). Data for Aquatic General Field Metadata are available in the NEON data product “Gauge Height” (NEON.DP1.20267). These data products are linked through the **siteID** field and local date in the NEON Data Variables for Periphyton and Phytoplankton Collection (AD[04]).

The benthic microbes data products are also loosely associated with the “Periphyton and Phytoplankton Collection” (NEON.DP1.20166) and “Periphyton and phytoplankton chemical properties” (NEON.DP1.20163), and may be tracked using **parentSampleID** and **sampleID** between data products.

3.7 Product Instances

At each aquatic site, there are up to 24 samples collected per year (8 samples per bout) at wadeable stream sites. Each sample generates data for marker gene sequences (AD[01]) and group abundance data (AD[06]). Data from mid-summer sampling bouts are also analyzed for metagenomics (AD[05]).

3.8 Data Relationships

For each record that is collected in the field (`alg_fieldData`), a number of child records may be created that exist in other L1 data products. In the event that sampling is impractical (e.g., the location is dry, ice covered, etc.), there will be no child records. If sampling is practical, there may be from 1 to 3 child records from a single **sampleID** generated in the field. Child records will be found in the data products Benthic Microbe Marker Gene Sequences (AD[04]), Benthic Microbe Metagenomic Sequences (AD[05]), and Benthic Microbe Group Abundances (AD[06]).

`amb_fieldParent.csv` - > One record is created for each sample collected in the field, creating a parent **sampleID** which is linked to all subsequent tables via **geneticSampleID**. This table also indicates the field conditions, including **habitatType**, **algalSampleType**, **substratum-SizeClass**, and sample depth if applicable (e.g., lake and non-wadeable sites).

3.9 Special Considerations

The benthic microbe samples are collected from a known area of stream bottom. If desired, data users may want to correct data for the **aquMicrobeScrubArea** sampled. Furthermore, epilithon and epixylon are collected as a scrub in the field, resulting in a **fieldSampleVolume** of water. This sample (“slurry”) of scrubbed material is then filtered onto a Sterivex filter resulting in the **geneticFilteredSampleVolume**.

<i>Title:</i> NEON User Guide for Aquatic Benthic Microbe Collection (NEON.DP0.20270)	<i>Date:</i> 12/14/2017
<i>Author:</i> Stephanie Parker	<i>Revision:</i> A

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 or web user interface (UI). 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. The field data entry workflow for collecting aquatic periphyton and phytoplankton data is diagrammed in 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 document NEON Raw Data Validation for Aquatic Benthic Microbe Collection, Level 0 (NEON.DP0.20270) (AD[03]), and provided with every download of this data product. Contained within this file is a field named ‘entryValidation-RulesForm’, 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[12]).

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[11]).

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 table below for an explanation of **dataQF** codes specific to this product.

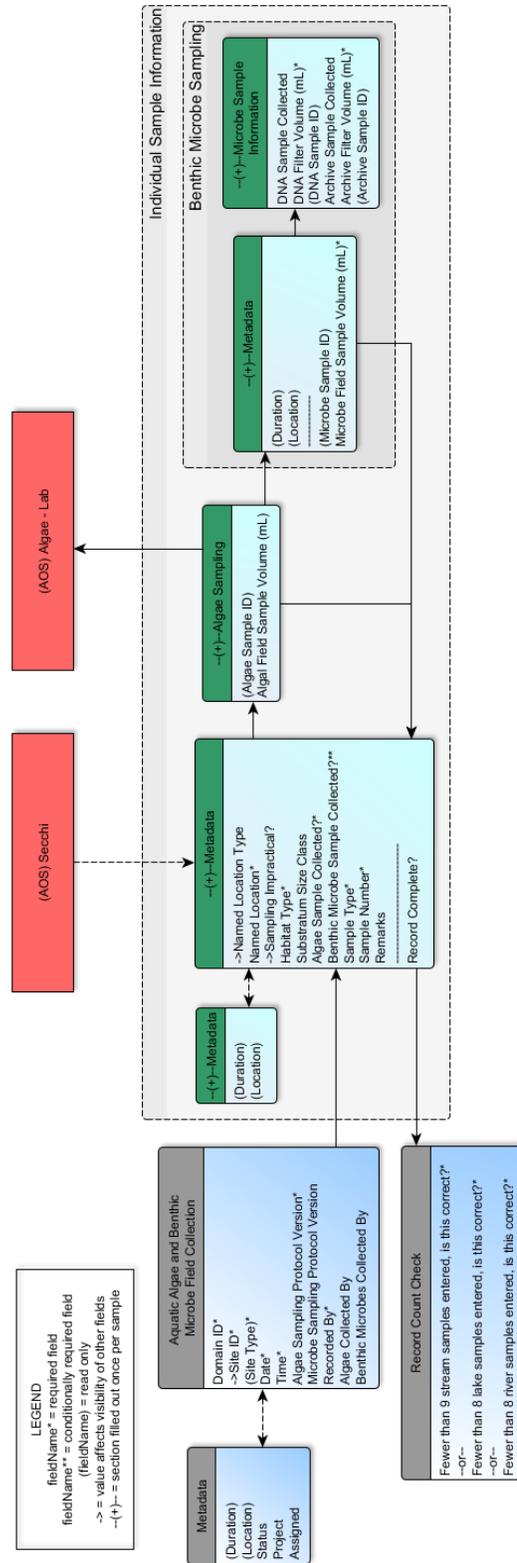


Figure 2: Schematic of the applications used by field technicians to enter periphyton and phytoplankton field data

<i>Title:</i> NEON User Guide for Aquatic Benthic Microbe Collection (NEON.DP0.20270)	<i>Date:</i> 12/14/2017
<i>Author:</i> Stephanie Parker	<i>Revision:</i> A

fieldName	value	definition
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

4.5 Analytical Facility Data Quality

Data analyses conducted on aquatic microbe external lab data are captured in the data product tables for Benthic Microbe Marker Gene Sequences (NEON.DP1.20280), Benthic Microbe Metagenomic Sequences (NEON.DP1.20279), and Benthic Microbe Group Abundances (NEON.DP1.20277), and in the data product user guides for marker genes, metagenomic sequences, and group abundances.

5 REFERENCES

Biggs, B. J. F., R. A. Smith, and M. J. Duncan. 1999. Velocity and sediment disturbance of periphyton in headwater streams: biomass and metabolism. *Journal of the North American Benthological Society* 18: 222-241.

Moulton, S. R., II, J. G. Kennen, R. M. Goldstein, and J. A. Hambrook. 2002. Revised protocols for sampling algal, invertebrate, and fish communities as part of the National Water-Quality Assessment Program. Open-File Report 02-150. U.S. Geological Survey, Reston, VA.