NEON ALGORITHM THEORETICAL BASIS DOCUMENT: QA/QC OF AQUATIC GENERAL FIELD METADATA

PREPARED BY	ORGANIZATION	DATE
Keli J. Goodman	AQU	02/01/2014

APPROVALS	ORGANIZATION	APPROVAL DATE
Charlotte Roehm	AQU	05/14/2014
Francois Pradeau	CI	05/14/2014
Mike Stewart	SE	05/25/2014

RELEASED BY	ORGANIZATION	RELEASE DATE
Stephen Craft	SE	05/29/2014

See configuration management system for approval history.

Change Record

REVISION	DATE	ECO #	DESCRIPTION OF CHANGE
A_DRAFT	05/29/2014	ECO-01873	Draft release as baseline

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

1 DESCRIPTION

1.1 Purpose

This document details the algorithms used for creating a subset of observation level metadata. These NEON Level 1 metadata, are derived from certain Level 0, field-collected wadeable stream, lake, and non-wadeable stream observation data In the NEON data products framework, the raw data collected in the field are considered the lowest level (Level 0). Raw data that have been quality checked, via the algorithms detailed herein, or from which simple metrics, such as total species richness, have been calculated are considered Level 1 data products. This document relates only to the generation of the former group of L1 data products, the quality controlled pass-through products, from the Level 0 data products.

This document includes a detailed discussion of measurement theory and implementation, appropriate theoretical background, data product provenance, quality assurance and control methods used, approximations and/or assumptions made, and a detailed exposition of uncertainty resulting in a cumulative reported uncertainty for this product.

1.2 Scope

quality controlled general aquatic field metadata from input data. These metadata include information regarding: samplers and/or data collector identity, time, measurements of stream stage, weather observations, and general site condition observations (Table 1). This document does not provide computational implementation details, except in cases where these stem directly from algorithmic choices explained here. It does, however, provide details relevant to the publication of data products via the NEON data portal (NEON Data Publication Workbook for AQU General Field Metadata: QA/QC of Raw Field Data (AD [11])).

This document describes the algorithms for ingesting and performing automated quality assurance and control procedures on the general aquatic site level metadata collected in the field pertaining to all sampling protocols completed on one date. The raw data processed in this document are detailed in the NEON Raw Data Ingest Workbook for AQU General Metadata (AD [10]).

2 RELATED DOCUMENTS AND ACRONYMS

2.1 Applicable Documents

Applicable documents contain information that shall be applied in the current document. Examples are higher level requirements documents, standards, rules and regulations.

AD[01]	NEON.DOC.000001	NEON Observatory Design
AD[02]	NEON.DOC.005003	NEON Scientific Data Products Catalog
AD[03]	NEON.DOC.005004	NEON Level 1-3 Data Products Catalog
AD[04]	NEON.DOC.005005	NEON Level 0 Data Product Catalog
AD[05]	NEON.DOC.005011	NEON Coordinate Systems Specification
AD[06]	NEON.DOC.001646	NEON General AQU Field Metadata Sheet
AD[07]	NEON.DOC.00XXXX	NEON ATBD Quality Flags and Quality Metrics for OS Data Products
AD[08]	NEON.DOC.001152	NEON Aquatic Sampling Strategy Document
AD[09]	NEON.DOC.004309	NEON Field Site Information
AD[10]	NEON.DOC.001627	NEON Raw Data Ingest Workbook for AQU General Field Metadata
AD [11]	NEON.DOC.001247	NEON Algorithm Theoretical Basis Document: QA/QC Data Validation and Plausibility Testing of TOS and AOS Field and Lab Data

2.2 Reference Documents

Reference documents contain information complementing, explaining, detailing, or otherwise supporting the information included in the current document.

RD [01]	NEON.DOC.000008	NEON Acronym List
RD [02]	NEON.DOC.000243	NEON Glossary of Terms
RD [03]		
RD [04]		

2.3 External References

External references contain information pertinent to this document, but are not NEON configuration-controlled. Examples include manuals, brochures, technical notes, and external websites.

ER [01]	
ER [02]	
ER [03]	

2.4 Acronyms

Acronym	Definition

AIS	Aquatic Instrumentation Systems
AOS	Aquatic Observation Systems
AQU	Aquatic Program at NEON
ATBD	Algorithm Theoretical Basis Document
NEON	National Ecological Observatory Network
NUID	NEON Unique Identifier

3 DATA PRODUCT DESCRIPTION

The AQU general field metadata is a collection of field-level metadata designed to apply to all aquatic data, collected via any field protocol, on any one date. The general field metadata products for aquatic sites consist of the observations returned from the NEON Domain Support Facility.

3.1 Variables Reported

The metadata to be provided to end-users are tabulated below. AQU general field metadata should be reported with all AQU field protocols, per date, per site. Given that metadata products formatted for publication via the NEON data portal will be provided with each protocol on a per date, per site basis, example publication metadata will be in publication-ready spreadsheets within each protocol ATBD.

Table 1. Field metadata associated with any AQU protocol activities (AIS and AOS) at an AQU site at measurement resolution per sample per bout. These metadata should be provided with all Aquatic data products, linked via siteID and date.

	#	METADATA FIELDNAME	DESCRIPTION	DATA TYPE	UNITS
FIELD METADATA	TBD	siteID	4-letter NEON site code	string	N/A
	TBD	date	Year, month, day samples were collected	date	YYYYMMDD
	TBD	aRecordedBy	Employee ID of primary technician	string	N/A
	TBD	bRecordedBy	Employee ID of second technician	string	N/A
	TBD	arriveTime	Time team arrived at AQU site.	time	hh:mm
	TBD	endTime	Time departed from AQU site for day	time	hh:mm
	TBD	initialStage	Stage reading at arriveTime	real	m
	TBD	endStage	Stage reading at endTime	real	m
	TBD	windDescrip	Qualitative description of wind levels	string	N/A
	TBD	cloudCoverPercentage	Sky Condition, % of cloud cover	string	%
	TBD	precipDescrip	Qualitative description of precipitation levels	string	N/A
	TBD	previousRain	Previous Rain event within 48 hours (Yes/No)	string	N/A
	TBD	previousRainEvidence	Evidence of previous rain event	string	N/A
	TBD	airTemperature	Estimated Air Temperature	string	Degrees F
	TBD	waterColorDescrip	Qualitative description of water color and clarity	string	N/A

	#	METADATA FIELDNAME	DESCRIPTION	DATA TYPE	UNITS
	TBD	riparianPhenologyDescrip	Qualitative description of riparian phenology status	String	N/A
	TBD	remarks	General Comments/Flow Conditions	string	N/A
	TBD	algaeRemarks	Post Field Work Observations if applicable: Algae	string	N/A
	TBD	macrophytesRemarks	Post Field Work Observations: Macrophytes	string	N/A
	TBD	leafLitterRemarks	Post Field Work Observations: Leaf Litter	string	N/A
	TBD	woodyDebrisRemarks	Post Field Work Observations: Woody Debris	string	N/A
	TBD	oilsRemarks	Post Field Work Observations: Oils/Surface Films	string	N/A
	TBD	trashRemarks	Post Field Work Observations: Trash	string	N/A

3.2 Temporal Resolution and Extent

The finest temporal resolution that AQU general field metadata will be tracked is at the level of an individual site in a single day.

3.3 Spatial Resolution and Extent

All AQU general metadata will be representative of the site level.

3.4 Associated Data Streams

All of the above metadata are associated with all AQU field protocols conducted at an Aquatic site on a given day, linked through the siteID and date fields.

3.5 Product Instances

All data products associated with STREON sediment basket general field metadata generated are calculated per individual STREON sediment basket.

4 SCIENTIFIC CONTEXT—BACKGROUND

4.1 Theory of Measurement/Observation

General metadata for a given site, on a given day may be applicable to many different protocols performed on that day. Thus, we have standardized general metadata collection into one general field metadata form that covers the qualitative and quantitative description of several variables, which are applicable to a variety of protocols. Additional, protocol specific field metadata can be found within each Aquatic field protocol, if applicable.

4.2 Theory of Algorithm

This document describes the algorithms for assessing the integrity of the L0 metadata stream generated by the AQU general field sampling of NEON Aquatic Field Protocols. The approaches described herein are simple yet necessary components of quality control and quality assurance, including: defining the range of possible values for each data product, specifying the lookup table(s) that contains the accepted values for particular data products, including linking variables, and tracking the individual-level data for consistency and accuracy through time.

4.3 Special Considerations

Note that this ATBD has been developed for consistency with all AQU field protocols.

5 ALGORITHM IMPLEMENTATION

5.1 Algorithm Processing Steps

Run the following processing steps for all data in the aquatic field metadata data ingest workbook.

1. Use afm_fieldSummary spreadsheet (AD[10]) to link controlled vocabulary field names (Column B, 'field') from data entered by technicians (Column C, 'entryData'). Convert Colum C, ' entryData' to Column B, 'field'.
2. Verify that all records in the [aquatic field metadata_dataingest_2014] workbook are complete
    a. Run Validation Test: Complete Records, in [AD(11)], where:
        i. List of data ingest sheets= [*afm_perdate_in*]
        ii. List of database table = [*afm_perdate_db*]
3. Verify that all records of all fields are of the correct data type in the [aquatic field metadata_dataingest_2014] workbook
    a. Run Validation Test: Data Type, in [AD(11)], where:
        i. List of data ingest sheets= [*afm_perdate_in*]
        ii. List of database table = [*afm_perdate_db*]
4. Convert **date** to correct format and verify that this is within acceptable range in the [aquatic field metadata _dataingest_2014] workbook
    a. Run Validation Test: Date, in [AD(11)], where:
        i. List of data ingest sheets= [*afm_perdate_in*]
        ii. List of database table = [*afm_perdate_db*]
5. Verify that technician IDs are valid in any field in which they are entered in the [aquatic field metadata _dataingest_2014] workbook
    a. Run Validation Test: Technician IDs, in [ADX], where:
        i. List of data ingest sheets= [*afm_perdate_in*]
        ii. List of database table = [*afm_perdate_db*]
6. Check for duplicate values in the [aquatic field metadata_dataingest_2014] workbook
    a. Run Validation Test: Duplicate Records, in [AD(11)], where:
        i. List of data ingest sheets= [*afm_perdate_in*]
        ii. List of database table = [*afm_perdate_db*]

iii. List of 'Test Values'= [Date]
iv. List of lists of 'Fieldnames'= [(**date**, **siteID**)]

7. Verify that code values in the [aquatic field metadata_dataingest_2014] workbook are valid, as specified by validation rules
    a. Run Validation Test: Validation Rules, in [AD(11)], where:
        i. List of data ingest sheets= [*afm_perdate_in*]
        ii. List of database table = [*afm_perdate_db*]
8. Verify that **siteID** values are valid for the site in which data were collected in the [aquatic field metadata_dataingest_2014] workbook
    a. Run Validation Test: Location, in [AD(11)], where:
        i. List of data ingest sheets= [*afm_perdate_in*]
        ii. List of database table = [*afm_perdate_db*]
9. Generate a unique ID (**uid**) for each record in the [aquatic field metadata_dataingest_2014] workbook
    a. Run Generate: Unique ID, in [AD(11)]
        i. List of data ingest sheets= [*afm_perdate_in*]
        ii. List of database table = [*afm_perdate_db*]
10. Assign domain and/or site identifications for each record in the [aquatic field metadata_dataingest_2014] workbook
    a. Run Assign: Location IDs, in [AD(11)], where
        i. List of data ingest sheets= [*afm_perdate_in*]
        ii. List of database table = [*afm_perdate_db*]
        iii. LHDD= [LHDD document number]
11. Generate ISO standard time values for records in [aquatic field metadata_dataingest_2014] workbook
    a. Run Generate: ISO Time, in [AD(11)], where:
        i. List of data ingest sheets =[*afm_perdate_in]*
        ii. List of database table = [*afm_perdate_db*]
12. Generate ISO standard date values for records in [aquatic field metadata_dataingest_2014] workbook
    a. Run Generate: ISO Date, in [AD(11)], where:
        i. List of data ingest sheets= [*afm_perdate_in*]
        ii. List of database table = [*afm_perdate_db*]
13. Generate anonymous technician IDs
    a. Run Generate: Technician IDs, in [AD(11)], where:
        i. List of data ingest sheets=[*afm_perdate_in*]
        ii. List of database table = [*afm_perdate_db*]
14. Generate spatial information and uncertainty for records in [aquatic field metadata_dataingest_2014] workbook
    a. Run Generate: Spatial Information and Uncertainty, in [AD(11)], where:
        i. List of data ingest sheets=[*afm_perdate_in*]
        ii. List of database table = [*afm_perdate_db*]
15. Assign data values, from L0 data values, in the [aquatic field metadata_dataingest_2014] workbook
    a. Run Assign: L1 Data from L0 Data, in [AD(11)], where:

i. List of data ingest sheets=[*afm_perdate_in*]
ii. List of database table = [*afm_perdate_db*]
iii. List of non-transferring fieldnames=[]

6 SCIENTIFIC AND EDUCATIONAL APPLICATIONS

NEON aquatic general field metadata will facilitate data sharing, analysis, and interpretation of unfamiliar research data. Furthermore, the NEON aquatic field metadata will be important to maintaining long-term data sets, such as documenting changes in personnel, methods, and site characteristics over time.

7 FUTURE MODIFICATIONS AND PLANS

- Specific instructions for handling and reporting quality flags
- Guidelines for releasing metadata