NEON SCIENCE AVAILABILITY PLAN

PREPARED BY	ORGANIZATION	DATE
Claire Lunch	SCI	01/17/2020
Cody Flagg	SCI	01/17/2020
Cove Sturtevant	SCI	01/17/2020
Christine Laney	SCI	01/17/2020
Tristan Goulden	SCI	01/17/2020
Kaelin Cawley	SCI	01/17/2020
Kate Thibault	SCI	01/17/2020
Greg Wirth	SCI	04/25/2018

APPROVALS	ORGANIZATION	APPROVAL DATE
Chris McKay	PMO	01/27/2020

RELEASED BY	ORGANIZATION	RELEASE DATE
Anne Balsley	CM	01/27/2020

The National Ecological Observatory Network is a project solely funded by the National Science Foundation and managed under cooperative agreement by Battelle. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Change Record

REVISION	DATE	ECO #	DESCRIPTION OF CHANGE
A	08/22/2018	ECO-05618	Initial Release
B	01/27/2020	ECO-06371	Complete rewrite based on feedback from STEAC and lessons learned; updated template

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 DESCRIPTION

The NEON Science Availability Plan summarizes the strategy for quantifying the operational performance of subsystems within the National Ecological Observatory Network (NEON), a project sponsored by the National Science Foundation (NSF) and managed under cooperative agreement by Battelle.

1.1 Purpose

The NEON Science Availability Plan describes how NEON will measure its success in collecting and publishing science data products. The mission of NEON is to collect, process, and share data on a wide range of ecological parameters consistently, both over space (i.e., across NEON terrestrial and aquatic field sites) and across time (i.e., over the planned 30-year lifetime of the Observatory). To determine whether NEON is fulfilling this mission, NEON will evaluate and report what fraction of the intended data is collected and published. This document defines methods to quantify the operational availability of NEON data products as well as thresholds for acceptable performance on each metric.

1.2 Scope

This document outlines methods for quantifying the availability for each of the NEON data-generating subsystems (AIS, AOP, AOS, TIS, and TOS; see definitions in Section 2.4) based on successful publication of data to the NEON Data Portal or designated partner portals (e.g., PhenoCam Network, Ameriflux). Availability is determined from the perspective of the data user, the most visible and impactful viewpoint of Observatory performance. Detailed consideration of the multi-step process of generating raw measurements, processing them into data products, and publishing them to the Portal will be handled in subsystem-specific documentation.

In addition to defining methods to quantify availability, this document also defines generic thresholds for acceptable performance on each metric. For a subset of data products and sites, customized thresholds are specified below in the subsystem-specific sections. These deviations from the generic thresholds are due to unavoidable limitations of the measurement system and/or site (e.g., water temperature measurements at seasonally dry aquatic sites). The thresholds presented here serve as a baseline from which site and product-specific adjustments may be made. Additionally, since the metrics are intended primarily as a tool for assessing areas in need of improvement, even their relative values (as opposed to the absolute values for comparison with the suggested performance thresholds) should prove useful in comparing the consistency of performance across time and among the various sites or subsystems.

This document does not propose a mechanism for completing the measurements and publishing them in an accessible form. Full implementation of the methods described herein, including venue and frequency of reporting, is the responsibility of the NEON Operations effort.

The Plan outlines those aspects of data availability applicable to all of the NEON data-generating subsystems.

- Section 3 discusses the concept of availability in the context of NEON operations.
- Section 4 describes the existing constraints applicable to availability metrics.
- Section 5 presents planned availability metrics for all of the NEON data-generating subsystems.

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 Document

2.2 Reference Documents

RD [01]	NEON.DOC.000008	NEON Acronym List
RD [02]	NEON.DOC.001984	Flight Boundaries Design Plan
RD [03]	NEON.DOC.000264	Tier 4 AOP Requirements Module
RD [04]	NEON.DOC.002652	NEON Data Products Catalog

2.3 External References

ER [01]	Sebastian-Coleman, Laura.2013. Measuring Data Quality for Ongoing Improvement: A Data Quality Assessment Framework, *In* MK Series on Business Intelligence, ISBN 9780123970336, https://doi.org/10.1016/B978-0-12-397033-6.00034-1.
ER [02]	NOAA US Climate Reference Network (CRN) Test and Evaluation Master Plan. January 2003 https://www1.ncdc.noaa.gov/pub/data/uscrn/documentation/program/X034FullDocD0Sig.pdf

2.4 Acronyms

Acronym	Meaning
AOP	Airborne Observation Platform
AIS	Aquatic Instrument System
AOS	Aquatic Observation System
TIS	Terrestrial Instrument System
TOS	Terrestrial Observation System

3 THE CONCEPT OF AVAILABILITY

Performance metrics are a useful tool for characterizing the success of observatory operations. For the NEON observatory, the primary products are data, and success is measured by data provided to, and used by, the scientific community. This document is focused on metrics of data provided to the community. Availability of data is dependent on successful functioning of all other parts of the observatory: data must be collected in the field and/or produced by analytical facilities, then transmitted to a central data management system, processed, and published. Breakdowns at any of these steps result in interruption to data availability. Thus, data availability can be used to determine how effectively the observatory fulfills its goal of providing data to the science community. Data availability metrics can drive decisions about how to optimally allocate operational resources to maximize the scientific output of the observatory. Here we adopt a data-centric approach to determining availability which can be adapted to any of the NEON subsystems.

We will monitor two basic measures of availability (Sebastian-Coleman 2013):

Completeness (technical availability): The quantity of data (e.g., number of records, pixels) published over a period of time, compared to the amount of data expected.

Validity (scientific availability): The proportion of data published over a period of time that has passed all quality checks.

From an operations perspective, completeness is an indicator of how robust internal processes are in collecting and publishing data. Validity metrics indicate whether internal processes are collecting and publishing useful data. Both metrics also provide key insight into Observatory performance from the perspective of the data user. High levels of completeness and validity, as measures of quality, give the user high confidence in the usefulness of the data in research. Table 1 defines the general thresholds for completeness and validity across all data products.

Table 1. General completeness and validity thresholds across all data products.

Data product	Threshold Completeness	Threshold Validity
All, with exceptions	90%	90%

Deviations from these thresholds will be data-product specific and summarized by subsystem in Section 5, which also explores how to derive these metrics for each of the NEON subsystems. It should be noted, however, that the root cause for missing or invalid data may be outside of NEON's control (e.g., extreme climatic events). Accurate interpretation of the metrics therefore relies on an assessment of root causes.

4 CONSTRAINTS

This document adheres to the following constraints on the reporting of NEON's performance on availability metrics.

4.1 Stakeholders

The key stakeholders likely to make use of these availability metrics are both internal and external to NEON. Internal stakeholders include:

- Field Science staff interested in knowing which instruments and protocols pose the greatest risk to data availability;
- Science staff concerned with how best to deploy science resources to improve performance on availability metrics;
- Cyberinfrastructure staff concerned with building, streamlining, and monitoring data processing and publication pipelines.
- Quality Assurance staff focused on the effectiveness of processes and controls to achieve availability metrics and respond to performance trends; and,
- Observatory leaders tasked with allocating observatory resources optimally to produce the greatest return on investment.

External stakeholders include:

- Scientists planning to use NEON data. They want to know what the track record is for the Observatory to deliver high quality data on time to be able to trust that data of interest will be available for future research.
- Educators using data in their curriculum.
- Scientists interested in knowing how well the Observatory is meeting its targets in providing useful scientific data to the research community.
- National Science Foundation and U.S. Congressional staff concerned with ensuring that the Observatory fulfills its obligations and provides acceptable return on investment by meeting operational requirements.
- Site hosts who want to know which data are being collected at their site and how they may aid in research and land management planning.

4.2 Subsystems

NEON shall measure operational availability performance for NEON's data-generating subsystems: AIS, AOP, AOS, TIS, and TOS.

4.3 Components

Three key inputs are required to generate metrics for completeness and validity:

1. The quantity of data intended to be generated for the product over the time period;
2. The quantity of data actually generated for the product over the time period;
3. The quantity of data generated for the product over the time period that has passed all quality checks.

Completeness equals the ratio of the second input to the first, and validity equals the ratio of the third input to the second.

Completeness and validity metrics will be reported at three resolutions within the observatory:

1. Data product by site availability metrics will apply to individual data products generated from raw data collected by NEON staff, airborne sensor payloads, or fixed sensors at a site.
2. Site-wide, subsystem availability metrics will evaluate the availability of multiple data products generated from a given site to highlight variations in site performance.
3. System-wide data product availability metrics will evaluate NEON's performance in generating a given data product, subsystem, or payload across multiple sites to indicate the relative difficulty in obtaining data for different data types.

Table 2 presents these three resolutions along with computational details applicable to all subsystems. Details specific to each subsystem are presented in their dedicated sections below.

Table 2. Resolutions of Data Availability Metrics.

Resolution	Scope	Computational details
Data product by site	Across all instances of a data product within a site	Completeness: Divide the actual quantity of data generated (total number of records, pixels, samples, etc.) by the expected quantity of data Validity: Divide the quantity of data passing all quality checks by the actual quantity of data generated
Site-wide, by subsystem	Per site, across all data products of a subsystem	For each site, average the data product by site metrics across all products of the subsystem
System-wide, per data product, subsystem, or payload	Per data product, subsystem, or payload, across all sites	For each data product, subsystem, or payload, average the data product by site metrics across all sites and relevant data products, if applicable

4.4 Reporting format

Given the effort involved in compiling performance measures in the early operations phase, the expectation is that performance metrics will initially be reported in the form of a written report compiled and shared with stakeholders. As NEON standardizes and improves the efficiency of generating these metrics, they will eventually be reported and shared with stakeholders through the NEON web portal.

5 DATA AVAILABILITY CRITERIA FOR NEON SUBSYSTEMS

NEON data are divided into three categories:

- Section 5.1 - Airborne remote sensing data gathered by NEON's Airborne Operations Platform
- Section 5.2 - Time series data streaming from sensors deployed at NEON sites
- Section 5.3 - Observational data collected by NEON's Field Science staff at NEON sites

These three fundamentally different types of data merit different approaches to characterizing data availability. The subsections below consider metrics appropriate to each of these data types.

5.1 AOP Data Availability Metrics

NEON's Airborne Observation Platform consists of three payloads with similar instrument configurations: each payload includes a LiDAR sensor, an imaging spectrometer, and a digital camera. The intent of the AOP is to periodically observe a region (minimum 100 km^2) that encompasses the NEON aquatic and terrestrial sites. The AOP will collect sites in single or multi-day campaigns at least three times every four years. D20 and D04 are currently exceptions, with collections expected once every five years at a minimum. The AOP team determines which sites are intended to be flown annually based on the foliar sampling schedule and sites successfully flown in previous years. AOP collections are required to be flown during a period of peak vegetative greenness, which has been pre-determined through analysis of historical satellite imagery. Successful collections of AOP data are highly dependent on weather as instruments cannot be operated in severe weather (e.g., active precipitation) and suffer from inaccuracies in the presence of overhead cloud cover. Timing of observations also considers historical weather trends to estimate sufficient time to collect data under nearly cloud-free conditions. Therefore, the goal of the AOP data collection program is to collect data: 1) at all planned sites, 2) covering the planned survey area, 3) with all instruments functional, 4) during peak greenness and 5) and under acceptable cloud-cover conditions. The aim of the AOP data availability metrics described below is to reflect how closely the NEON AOP collections approach this ideal.

Based on the sites identified in the Annual Program Plan, AOP will publicly release the planned flight schedule each year on the NEON website. The plan indicates the set of NEON sites that AOP will attempt to survey, as well as the intended survey dates. These dates are typically linked to other NEON measurements scheduled at a particular site such as foliar sampling. As defined in RD [02], each NEON terrestrial site includes three types of areas:

- A "Priority 1 Flight Box" (P1FB) area encompassing the TOS sampling boundary, NEON tower, and collocated Aquatic instrumentation. This represents the goal of the area to be surveyed. This area is a minimum of 100 km^2 for terrestrial and collocated Aquatic sites and is variable for non-collocated aquatic sites due to variation in the upstream watershed that contains aquatic instrumentation.
- A "Priority 2 Flight Box" encompassing a larger area based on relevant ecological parameters that extends beyond the TOS sampling area; this area is desirable to survey but of lower importance.
- A tower airshed boundary, defined by the predominant wind direction at the site that identifies the area primarily observed by the tower instrumentation. This is a smaller area within the P1FB that is given highest priority.

Since AOP's primary mission is to observe the P1FB / airshed, we consider only this area in the availability threshold exceptions described below (Table 3). The P1FB at each site contains a series of parallel flight lines which the aircraft follows to cover the intended area. Lines are designed with sufficient spacing to allow overlap in the area observed by the aircraft. Due to the reliance of sufficient clear weather to collect data, the AOP is required to survey 80% of the P1FB and 95% of the tower airshed (RD[03], see NEON.AOP.4.1300 and NEON.AOP.4.1301). As such, these thresholds are specific to the AOP collection efforts and represent the threshold for completeness for all AOP data products at sites collected in a given year.

Table 3. Deviations from default data completeness and validity thresholds.

Data product	Threshold completeness	Threshold validity	Explanation
Lidar and camera data products*	80% / 95%	90%	Observatory requirements specify the AOP cover 80% of the P1FB and 95% of the tower airshed. For lidar and camera data products, the validity metric follows the general guidelines.
Spectrometer data products*	80% / 95%	68% / 80%	Observatory requirements specify the AOP cover 80% of the P1FB and 95% of the tower airshed. For the spectrometer data products, AOP uses historical weather data to inform deployment times with an 80% chance of weather-free data at foliar sampling sites and 68% chance of weather free collection days at non-foliar sampling sites.

* See RD[04] for details

5.2 Instrumented Subsystem Data Availability Metrics

NEON's AIS and TIS subsystems (hereafter referred to as the Instrumented Subsystems, or IS) include sensors that generate time-series data on a host of environmental parameters ranging from the temperature of water in streams, rivers, and lakes, to the observed amount of shortwave radiation at various levels on NEON terrestrial towers. Ideally, these data streams would gather useful data continuously at all sites throughout the year. In practice, sensor data streams can be interrupted by planned maintenance or power outages, for example. Even when data are collected, they may be rendered invalid by environmental conditions such as temperatures that lie outside the operating range of the sensor. Regardless of interruptions and periods of invalid data, the raw data should be processed into data products and users should receive a record for every expected measurement, even if the measurement itself is null and only quality information is supplied. Thus, completeness and validity metrics can provide stakeholders with insight into the areas that can most improve NEON's overall data availability and quality through improved maintenance, sensor redesign, and technology improvements.

All TIS and AIS data products consist of time-series data in which the expected quantity is easily calculated by dividing the total time period by the averaging interval (in the same units). For example, the barometric pressure data product is reported as 1- and 30-minute averages. Thus, expected data quantity for the 30-minute average product over a 30-day month is equal to (30 days * 24 hr/day * 60 min/hr) / (30 min averaging interval) = 1440 records. In cases where multiple averaging intervals are published, the availability metrics will typically be computed for the longer averaging interval.

Because of the regularity of IS data products, computation of the completeness and validity metrics for any time period is straightforward following the computations details of Section 4.3. However, in cases where sites and/or data products are inactivated seasonally as part of the measurement protocol, the expected data quantity will be computed using their planned date ranges of operation.

NEON has adopted default data completeness and validity thresholds each at 90% for IS data products at all resolutions described in Section 4.3. The completeness threshold is derived from the commissioning plans of the United States Climate Reference Network (USCRN), which employed a 90% threshold for data completeness for most data types. Deviations from these default thresholds are described with brief explanations in Table 4. Appendix A describes the methods used to derive the validity threshold exceptions. Due to the complexity of surface-atmosphere exchange (SAE) data products (e.g., CO2 flux), Appendix B provides the dependency mapping and corresponding validity threshold computations for these data products.

Table 4. Deviations from default data completeness and validity thresholds[1]

Data product	Site	Threshold complete-ness	Threshold validity	Explanation
Soil heat flux plate (DP1.00040)	All	90%	60%	Auto-calibration routine occurs every 3.25 hours, each occurrence resulting in 0.5-1.5 hours of intentionally flagged data in the 30-minute data product.
Soil CO2 concentration (DP1.00095)	All	90%	73%	Product depends on three independent sensors.
Soil water content and salinity (DP1.00094)	All	90%	81%	Product depends on two independent sensors.
Shortwave radiation (direct and diffuse pyranometer) (DP1.00014)	All	90%	72%	By design, direct radiation is quality flagged for solar zenith angles > 84°, for which the measurements are unreliable. Zenith angle varies with time of day, season, and latitude. However, integrated across an entire year, the fraction of zenith angles exceeding this threshold is similar across NEON sites (54%-61%). Allowed for the maximum expected flagging fraction and accounted for the fact that the direct radiation sub-product accounts for ⅓ of the data product. Threshold validity applies to an operational year.
Atmospheric CO2 & H2O turbulent concentrations (DP1.00034.001, DP1.00035.001)	All	90%	84%	See SAE dependency mapping in Appendix B.
Atmospheric CO2 and H2O storage concentrations (DP1.00099.001, DP1.00100.001)	All	90%	76%	See SAE dependency mapping in Appendix B.
Atmospheric CO2 isotope concentrations (DP1.00036.001)	All	90%	75%	See SAE dependency mapping in Appendix B.

Data product	Site	Threshold complete-ness	Threshold validity	Explanation
Atmospheric H2O isotope concentrations (DP1.00037.001)	All	90%	62%	See SAE dependency mapping in Appendix B.
Temperature rate of change (DP2.000024.001)	All	90%	87%	See SAE dependency mapping in Appendix B.
CO2 & H2O concentration rate of change (DP2.00008.001, DP2.00009.001)	All	90%	70%	See SAE dependency mapping in Appendix B.
Temperature rate of change profile (DP3.00008.001)	All	90%	76%	See SAE dependency mapping in Appendix B.
CO2 and H2O concentration rate of change profile (DP3.00009.001, DP3.00010.001)	All	90%	56%	See SAE dependency mapping in Appendix B.
Momentum flux, flux footprint characteristics (DP4.00007.001, DP4.00201.001)	All	90%	81%	See SAE dependency mapping in Appendix B.
Turbulent subproducts of: Sensible heat flux, Latent heat flux, CO2 flux (DP4.00002.001, DP4.00067.001, DP4.00137.001)	All	90%	68%	See SAE dependency mapping in Appendix B.
Storage subproducts of: Latent heat flux, CO2 flux (DP4.00067.001, DP4.00137.001)	All	90%	35%	See SAE dependency mapping in Appendix B.
NSAE subproduct of: Sensible heat flux (DP4.00002.001)	All	90%	41%	See SAE dependency mapping in Appendix B.

Data product	Site	Threshold complete-ness	Threshold validity	Explanation
NSAE subproduct of: Latent heat flux, CO2 flux (DP4.00067.001, DP4.00137.001)	All	90%	24%	See SAE dependency mapping in Appendix B.
Full data product (proportional representation of subproducts) of: Sensible heat flux (DP4.00002.001)	All	90%	57%	See SAE dependency mapping in Appendix B.
Full data product (proportional representation of subproducts) of: Latent heat flux, CO2 flux (DP4.00067.001, DP4.00137.001)	All	90%	42%	See SAE dependency mapping in Appendix B.
All IS data products	YELL, BLDE	90%	78%	Mandated bear area closure period from March 10 to June 30 prevents site access/maintenance. This time period represents a continuous 31% of an operational year, during which sensors may not be repaired, cleaned, etc. It is assumed that half of the data during this period will be quality flagged, and dates outside the closure period will follow the default threshold. Threshold validity applies to an operational year.
Water Quality (DP1.20288)	All	90%	81%	Product depends on 2 independent sensors.
Windspeed and direction above water on-buoy (DP1.20059)	All	90%	81%	Product depends on 2 independent sensors.
Stream discharge rating curve (DP4.00133)	All	90%	81%	Product depends on 2 L1 data products.

Data product	Site	Threshold complete-ness	Threshold validity	Explanation
Stream discharge (DP4.00130)	All	90%	73%	Product depends on 2 L1 and 1 L4 data product.
Water Quality (DP1.20288), nitrate (DP1.20033) and Surface water elevation, temperature digital, temperature PRT, and conductivity (DP1.20016, DP1.20054, DP1.20053, DP1.20008)	SYCA, KING, MCDI	90%	60%	Streams will likely go dry for ~4-month periods of the year that sensors are deployed. The remaining period follows the default threshold.
Groundwater elevation, temperature, and conductivity (DP1.20100, DP1.20217, DP1.20015)	HOPB, MART, BLDE, COMO, WLOU	90%	60%	Groundwater wells will likely go dry for ~4-month periods of the year that sensors are deployed. The remaining period follows the default threshold.
PAR below water surface (DP1.20261) and Surface water elevation, temperature digital, and conductivity (DP1.20016, DP1.20054, DP1.20008)	CRAM, LIRO, PRPO, PRLA	90%	60%	Lakes seasonally freeze for up to ~4-month periods of the year that sensors are deployed. The remaining period follows the default threshold.
Water Quality (DP1.20288), nitrate (DP1.20033) and Surface water elevation, temperature PRT, and conductivity (DP1.20016, DP1.20053, DP1.20008)	MCDI, KING, BLDE, COMO, WLOU	90%	60%	Streams seasonally freeze for ~4-month periods of the year that sensors are deployed. The remaining period follows the default threshold.

[1]Validity thresholds were derived via the raw-dependencies method unless noted otherwise.

5.3 Observational Subsystems Availability Metrics

NEON's AOS and TOS subsystems (hereafter the "observational subsystems", or OS) are fundamentally unlike the instrumented systems in terms of both their methods for collecting data and the resultant data products. Rather than employing sensors that constantly gather data on a regular frequency, the observational systems rely on NEON's Field Science staff, distributed across the various domains, to perform the act of collecting data at terrestrial (TOS) and aquatic (AOS) sites. The Field Ecologists do this by executing protocols that involve either recording observations directly (e.g., entering stream depth measurements into a tablet), collecting samples for analysis in the domain lab (e.g., identifying ground beetles collected from traps), or sending field-generated samples to an external analytical laboratory (e.g., shipping sediment chemistry samples for physical and chemical analysis). Adding further complication, executing a given protocol at different sites may involve sampling a different number of locations or executing a different number of bouts. A detailed determination of expected data volume for each data product by site by year is beyond the scope of this document; we will use a combination of protocols, site sampling schedules, and product-specific data latency to estimate expected data availability at any given time.

Observational subsystem data are typically published between one and nine months after sample collection; the time between collection and publication is referred to as latency. Expected latency may differ for each data table within a data product. For example, data collected directly in the field, such as phenology observations, are published within one to two months of collection, while data generated by analytical facilities, such as microbial sequencing, require time for shipping of samples, processing and analysis, and submission of the resulting data back to NEON.

For data *collected* within a given time period, completeness and validity are re-calculated over time, as both the numerator (available data) and denominator (expected data) of completeness change over time. In other words, completeness calculated at any given time is the ratio of available data to expected data, and expected data includes only data whose latency period has passed. To ensure clarity, reported completeness metrics for the OS subsystem will be accompanied by metrics indicating the percentage of total data expected by the time of reporting. Each month's report will include completeness metrics for all previous months within the reporting period. Figure 1 shows a hypothetical rollout of expected data availability and corresponding metrics for a single data product.

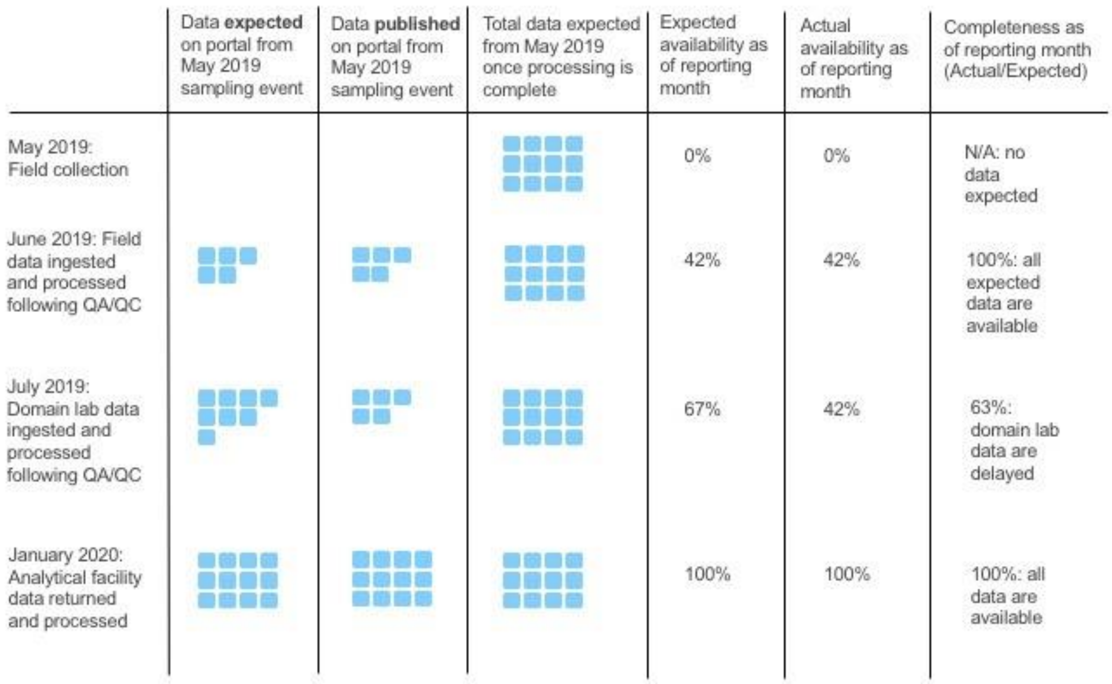

Figure 1. OS Data latency and availability. Completeness metric for May 2019 is re-calculated in each monthly report as expected and actual availability increase over time. For a given data product, for a given time period of data collection, once the longest expected latency period of its data has passed, total completeness and validity for that time period can be calculated. Similarly, for the aggregate availability of the subsystem, once the longest expected latency of any data product has passed, availability of the entire subsystem can be calculated. For example, if the longest latency period of any OS data product is a year, total data availability for the OS system in 2018 can be calculated at the end of 2019.

NEON has adopted default data completeness and validity thresholds each at 90% for OS data products at all resolutions described in Section 4.3. Deviations from these default thresholds are described in Table 5, with explanation.

Table 5. Deviations from default data completeness and validity thresholds.

Data product	Site	Threshold completeness	Threshold validity	Explanation
All in-stream data products and stream discharge	SYCA, KING, MCDI	66%	90%	These sites are intermittently dry for ~4 months; it is expected that one of three biological sampling bouts and 9 of 26 water chemistry bouts will be missed due to lack of water

Data product	Site	Threshold completeness	Threshold validity	Explanation
Chemical properties of groundwater (DP1.20092) and stable isotope concentrations in groundwater (DP1.20276)	OKSR, TOOK, CARI	75%	50%	Often not enough active layer melt water to obtain a sample (~25% of the time no sample can be collected) or complete sample suite at all wells. Data will include only a subset (~50%) of the analytes that were planned.
Chemical properties of groundwater (DP1.20092) and stable isotope concentrations in groundwater (DP1.20276)	MART	75%	50%	Shallow wells with low recharge rate, often not enough water to obtain a sample (~25% of the time no sample can be collected) or complete sample suite at all wells. Data will include only a subset (~50%) of the analytes that were planned.
Chemical properties of groundwater (DP1.20092) and stable isotope concentrations in groundwater (DP1.20276)	BLDE, COMO, WLOU, SYCA, REDB	87.5%	75%	Shallow wells with low recharge rate, often not enough water to obtain a sample (~12.5% of the time no sample can be collected) or complete sample suite at all wells. Data will include only a subset (~75%) of the analytes that were planned.
Chemical properties of groundwater (DP1.20092) and stable isotope concentrations in groundwater (DP1.20276)	HOPB	62.5%	50%	Shallow wells with low recharge rate, often not enough water to obtain a sample (~37.5% of the time no sample can be collected) or complete sample suite at all wells. Data will include only a subset (~50%) of the analytes that were planned.

Data product	Site	Threshold completeness	Threshold validity	Explanation
Fish counts (DP1.20107) and DNA barcoding (DP1.20105)	WLOU, COMO, BLUE, SYCA, REDB, BLDE, TECR, OSKR	50%	90%	Years when snowpack/runoff is above average or precipitation is lower than average sampling may only occur during one bout
TOS products	OSBS, DSNY, DELA, LENO, UNDE, BARR	80%	90%	Seasonal flooding of plots is common

APPENDIX A: METHODS TO DERIVE IS VALIDITY THRESHOLD EXCEPTIONS

Contributors: Cove Sturtevant, Stefan Metzger, Natchaya Durden, Hongyan Luo, David Durden, Chris Florian

The general validity threshold of 90% (Table 1) applies to the majority of NEON IS data products given that they share a similar number of dependencies, maintenance schedules, and sensor failure rates. However, some data products are more complex than others and require greater consideration of their dependencies, maintenance requirements, and operating conditions to derive ambitious, yet realistic thresholds for valid data. For these exception cases, the following presents two systematic approaches to deriving defensible thresholds. Both are equally acceptable, but one may be more appropriate given the degree of independence of the products or components that feed into the target data product. In both methods, if the data product contains multiple sub-products with different dependencies, the validity thresholds for each sub-product are computed separately using the most appropriate method and an average is computed with equal representation of each sub-product.

Raw-dependencies method

The raw-dependencies method is appropriate for data products that are derived from relatively independent sensors or components. The formula for deriving the validity threshold is below. See SAE dependency mappings and associated computations in the blue boxes of Appendix B for example applications of this method.

$$T = \left(1 - \sum P_{i,unique}\right) \prod T_0$$

The components of this equation are described as follows:

- T: Target data product validity threshold (fraction), where validity is defined as the proportion of present records expected to pass all quality checks.
- $\sum P_{i,unique}$: The sum of unique (non-overlapping) exception periods (fraction) for the target data product and all other data products or raw sensors/components from which it is derived. An exception period is the proportion of time that the readings for a given data product, sensor or component are NULL or quality flagged by design (e.g. calibration or validation periods) and not already accounted for in lower-level contributing products, sensors, or components.
- T_0: The raw validity threshold (fraction) for the lowest level (L0) sensor or component from which the target data product is derived. The raw validity threshold is the proportion/probability of normal operating time (not including exception periods) that the sensor or component is expected to report high-quality readings. By default, validity thresholds are set to 0.9 for sensors

(including their accompanying data logger) and 0.95 for supporting components (e.g. pumps), unless justified otherwise.

- $\prod \quad T_0$: The product of raw validity thresholds for all L0 sensors/components which are used to derive the target data product. It represents the independent compound probability of obtaining a valid measurement for the target data product, not including exception periods.

<u>Nearest-ancestors method</u>

In the case of co-depended inputs, also their probabilistic validities co-depend. This is the case for example when calculating time- or space-derivatives, i.e., combining time- or space-shifted inputs. Here, assuming independence (the raw-dependencies method) would duplicate their probabilistic validities and thus unrealistically deteriorate the target validity estimate. For this reason, the nearest-ancestors method represents the correlation structure of the inputs as an empirically derived exponent f. Let's consider the simple case of a time derivative: the data product is derived from two inputs of a single immediate 'ancestor', in which one of the inputs is shifted one step in time. With no exception periods, the probability of a record passing all quality checks (probabilistic validity) is:

$$T = T_{nearest}{}^{f}$$

Here, $T_{nearest}$ is the probabilistic validity of the original time series, or in general the nearest ancestor at any time point. Then, f represents the correlation between the validity of the ancestor at two consecutive time points. When the validities at two consecutive time points are nearly independent, $f \to 2^{-}$ (or $T \to T_{nearest} \times T_{nearest}$, analogous to the raw-dependencies method). Conversely, when the validities are highly correlated, $f \to 1^{+}$ (or $T \to T_{nearest}$). Expanding to N combinations of time-shifted inputs from the same ancestor, $f \to N^{-}$ when their validities approach independence, and $f \to 1^{+}$ when they are highly correlated.

Next, we expand this concept to a target product that is derived from concurrent inputs of two different ancestors with partially correlated probabilistic validities. This would be the case e.g. for two ancestors that are derived from separate scientific instrumentation but a shared pump. Then, f represents the correlation between the validities of the two immediate ancestors, $nearestA$ and $nearestB$:

$$T = (T_{nearestA} \times T_{nearestB})^{f}$$

Here, f follows the same directionality as above but $f \to 1^{-}$ (or $T \to T_{nearestA} \times T_{nearestB}$) when the validity of the inputs approach independence and $f \to \frac{1}{2}^{+}$ (or $T \to T_{nearestA} \to T_{nearestB}$) when they are highly correlated. Expanding further to N different immediate ancestors, again $f \to 1^{-}$ when they are nearly independent and $f \to \frac{1}{N}^{+}$ when they are highly correlated.

Generalizing to any number of immediate ancestors, whether they are different ancestors, a single time- or space-shifted ancestor, or any combination thereof, and incorporating an exception period for the target data product, the validity threshold for the target data product is:

$$T = (1 - P)\left(\prod T_{nearest}\right)^{f}$$

The components of this equation are described as follows:

- T: Target data product validity threshold (fraction), where validity is defined as the proportion of present records expected to pass all quality checks.
- P: Exception period (fraction) for the target data product in which data are NULL or quality flagged by design (e.g. calibration or validation period), and not already accounted for in the nearest ancestors.
- $\prod T_{nearest}$: The product of the validity thresholds for the nearest ancestor data products, sensors, or components from which the target data product is derived.
- f: Empirically-derived exponential adjustment representing the correlation structure of the inputs. The value of f will be larger when the inputs are more independent and smaller when they are more correlated.

An important property of the nearest-ancestors method is that it is infectious. Meaning, once the nearest-ancestors method is used to compute threshold validity for a data product, all descendants of the data product must also use the nearest-ancestors method. This is because the inputs of at least one of its ancestors lack independence. See SAE dependency mappings and associated computations in the yellow boxes of Appendix B for example applications of this method.

APPENDIX B. SAE VALIDITY THRESHOLD COMPUTATIONS

Contributors: Cove Sturtevant, Stefan Metzger, Natchaya Durden, Hongyan Luo, David Durden, Chris Florian

The data product dependency mappings and associated computations for validity thresholds for surface-atmosphere exchange (SAE) data products are given in Figures B1 and B2 below. In both figures the blue boxes follow the raw-dependencies method and the yellow boxes follow the nearest-ancestors method. See Appendix A for method details and rationale for choice of method.

In the case of the nearest-ancestors method (time-shift, space-shift, integrated), we determined the *coefficients* as the median value over 1 month of data at each of a diverse set of five NEON sites (BART, BLAN, CLBJ, DELA and HARV). Determination of the *coefficients* is sensitive to the relative change of data validity from one data product level to the next, and largely independent of absolute data validity. In the special case of independent subproducts we set $f = 1$ (definition for the independent case, see Appendix A).

Gray boxes show raw sensors or components. White boxes show average Level 4 data product validity based on the equal representation of its subproducts. Arrow coloring is for the purpose of differentiation only.

We note that although the SAE validity thresholds below were derived separately, they are in the range of community expectations. A study of daytime data availability (a similar metric to validity) across the FLUXNET network showed turbulent fluxes of sensible heat, latent heat, and CO_2 to be available 68%, 62%, and 30% of the time, respectively (van der Horst et al. 2019). These values compare well to our derived threshold of 68% for all three turbulent fluxes.

References

van der Horst, S.V.J., Pitman, A.J., Kauwe, M.G.D., Ukkola, A., Abramowitz, G., Isaac, P., 2019. How representative are FLUXNET measurements of surface fluxes during temperature extremes? Biogeosciences 16, 1829–1844. https://doi.org/10.5194/bg-16-1829-2019

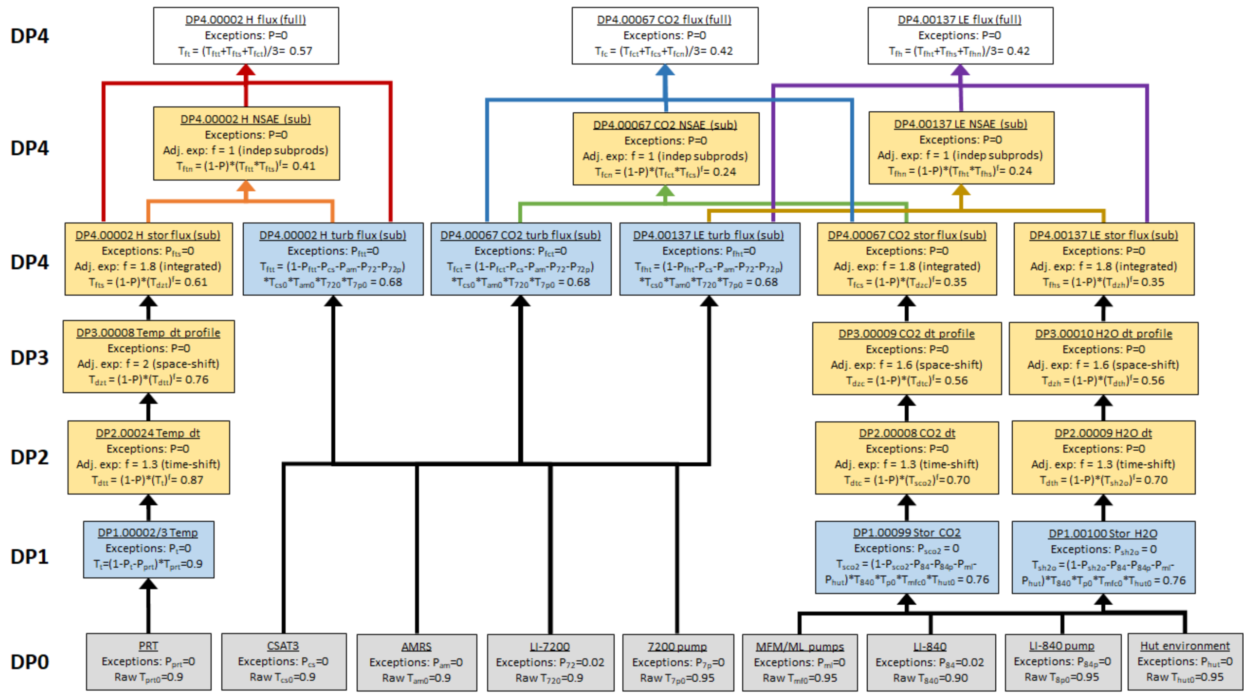

Figure B1. Dependency mapping and validity threshold computations for a sensible heat flux, latent heat flux, and CO_2 flux related data products.

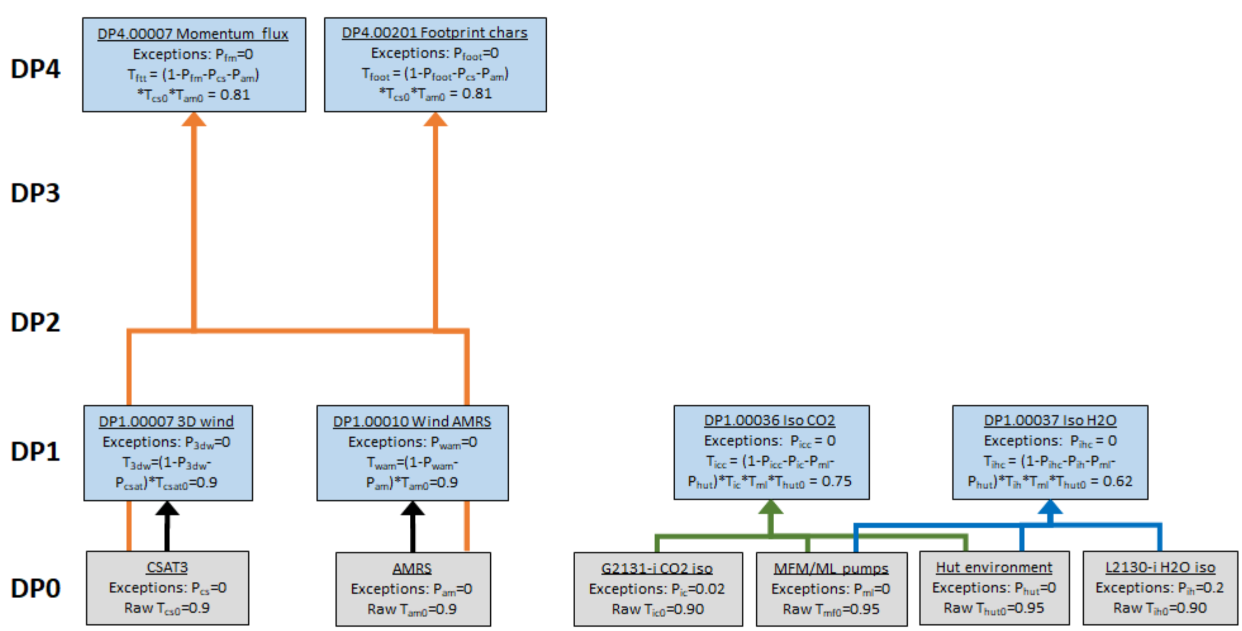

Figure B2. Dependency mapping and validity threshold computations for momentum flux, footprint characteristics, and isotopic CO_2 and H_2O concentration related data products.