Battelle Response to NEON STEAC September 2018 Advisory Report
October 29, 2018

I. Overview

According to its bylaws, the STEAC is "primarily an advisory body to the NEON Project and will provide strategic advice to Battelle, the NEON Principal Investigator (PI), and NEON Project staff on the planning, construction, and operation of the NEON Project and other relevant programs." This response to the STEAC report from September 2018 therefore combines the input of the Observatory Director/Chief Scientist of NEON, Dr. Sharon Collinge, with that of several Battelle staff whose duties impact the execution of the NEON project.

Battelle appreciates the STEAC's recognition of the NEON project's progress on construction and operations since the last STEAC meeting and report in April 2018, as well as the thoughtful comments that the STEAC provided during their September meeting and formally for this report. As representatives of the scientific community committed to the success of NEON, we value the observations and recommendations of the STEAC, and appreciate the sustained commitment to "provide continued advice as the NEON Project transitions into full operations and becomes an open data organization."

Following are responses to the major sections of the Advisory Report.

II. Organizational Structure and Decision-making

Battelle appreciates the recognition of the need for a shift in orientation of the organizational structure and decision-making as the project moves from construction to operations. The STEAC notes the importance of clear lines of communication between NSF, NEON project leadership, and Battelle leadership, and recommends that these relationships be reflected more clearly in the NEON organizational chart. Discussions are currently underway with all three entities (NSF, NEON project leadership, and Battelle leadership) to clearly define and assign communication and decision-making processes as appropriate among these entities. Further refinement of these relationships will promote the smooth operation of the observatory, and will also enable clearer communication to the external community of the structure and decision-making processes of the NEON project. We note that because it is inconsistent with good governance, Battelle's authorities to operate do not authorize subcontractors to discharge personnel decisions, nor obligate Battelle financially. That said, the Chief Scientist for the NEON Project is expected to provide input to the performance evaluations of key project staff, and to the development goals for these staff. The CS is also charged with setting the scientific direction and priorities for the Project.

III. Cyberinfrastructure

The STEAC astutely observes that cyberinfrastructure "is a need that crosses all NEON functions, from data collection, QA/QC, data release to the community, and user feedback." Battelle appreciates the recommendation of the need for a Chief Information Officer (or similar title) that has a holistic view across all aspects of IT for the observatory, and is able to interact with project leadership to set priorities for maintaining, developing, and improving NEON's cyberinfrastructure. Battelle concurs with the STEAC recommendation regarding the value of collaboration with groups outside of NEON that share common issues and generate best practices. Current collaborations between Battelle and outside organizations that are proving particularly productive include DataONE (C Laney), Earth Science Information Partners

(ESIP) (C Laney), EarthCube Council of Data Facilities (CDF) (C Laney is a member of the Executive Committee), and CyVerse, who participated in this year's Data Institute. A new collaboration scheduled to kick off in early November is with the NSF-funded pilot study for a Cyberinfrastructure Center of Excellence. This opportunity relates directly to the STEAC recommendation to connect with existing NSF-funded centers in CISE (Computer Information Science and Engineering), and Battelle plans to explore more such opportunities. Each of these collaborations provides rich opportunities for sharing experiences and improving performance of cyberinfrastructure, and ultimately for more efficient and effective collection, delivery, and usage of NEON data. Regarding the recommendation to establish a CI advisory board and TWG, Battelle looks forward to further discussion with the STEAC to determine the proposed role of these entities and how they would be distinct from one another.

Open-source-driven development: Battelle agrees that all of NEON's code should be made publicly available, within the constraints of cybersecurity concerns, and includes this in the current three-year plan. This is a significant effort, particularly given staff turnover in CI, and will take some time. Battelle is prioritizing the overall improvements to the IS data pipeline, as supported by the recent $3M CI enhancement award, and increasing data availability, over the substantial effort it will take to prepare the code for publication. It should be noted that the data processing details are captured in the data product-specific ATBDs for both IS and AOP. For the OS data pipeline, Battelle is currently prioritizing the preparation and publication of code that increases data usability for end users, although the pipeline code will eventually be published. For AOP, Battelle will prioritize publication of the AOP code over the next year. Staff resources are the limiting factor here, but Battelle is currently hiring new staff to the team, to hopefully free up some resources.

Training for NEON software products: Several training tutorials are available online, and Battelle has received anecdotal feedback that these are very helpful, especially to new users: https://www.neonscience.org/resources. Other opportunities for training include the NEON Faculty Mentoring Network (FMN), which trains faculty to teach undergraduate courses using NEON data, with the outcomes of (1) disseminating and distributing training in NEON data use and (2) adding tutorials to the NEON set, since the faculty develop their own lessons using NEON data, in addition to teaching the lessons. NEON runs the summer Data Institute (one-week long workshop at HQ where participants dive into specific NEON data) and workshops at ESA annually, with at least one ESA workshop every year focused on data access and/or use. NEON also collaborates with Data Carpentry and QUBES to provide educational tools to the NEON data user community. In November 2018, NEON will host a two-day data workshop for 10 graduate students (there were 90 applicants for these 10 slots!) The funding source for this workshop was a one-time opportunity, but based on what is learned from teaching the workshop, and evaluations from the students who attend, Battelle will explore future options for teaching similar material. Alternative funding, virtual workshops and webinars, and host-sponsored remote workshops are all options.

IV. Science Staff Professional Development

NEON Science staff have been incorporated into the "matrix" of scientists originally developed for the long-standing, environmental operations at Battelle. This matrix includes the descriptions of how different staff roles are distinguished by their degree of client interaction, innovation and intellectual property, marketing and business development, direction of other staff, and education and experience. It describes six levels along the career path from Research Associate to Research Leader. NEON project staff have been provided this matrix. The matrix is a standard tool in corporate human resources. It is not as clear as tenure-track requirements often are, which leads to frustration and confusion amongst

the staff. Also, promotion is based on the duties that one performs, not the amount of time one has been performing the duties, as well as the need on a project for someone to perform those duties. The head of Battelle Contract Research, Matt Vaughan, has expressed a desire to improve the performance metrics and evaluation of technical staff at Battelle and has invited NEON's input to this process. NEON Science managers are currently working to define internally some consistent guidelines that can be communicated to staff to increase clarity.

Battelle recognizes that training and professional development opportunities are critical to retaining and developing high quality staff and a successful Observatory. Battelle provides support for professional development, and science staff have taken advantage of these opportunities. The time constraints are, of course, very real, and NEON science leadership are actively working with the Observatory Director to incorporate more science staff time into collaboration with the external community. The Observatory Director is actively promoting increased collaboration opportunities with external communities, including NCAR, NOAA, LTER, CZO, and AmeriFlux.

The integration of all staff into Battelle has greatly improved library access for the science staff, providing access to tools such as Scopus and Web of Science. The Battelle librarian has reached out to NEON staff to understand staff needs, and she is currently working with the list of priority journals NEON provided to evaluate options. She recently provided a presentation on library resources at the latest science team meeting, and progress is expected to continue.

V. Data Availability and Training

Make data review by science and domain staff a top priority.
Battelle Science and domain staff appreciate the recommendation by the STEAC that the staff review of data and testing of data products is critical to ensure high quality data. Currently, a significant portion of science staff time is allocated to data QA/QC. At present, high priority is given to correcting known quality issues of severe magnitude, such as improving the robustness of stream sensor systems and the instrumented systems data pipeline. Targeted data review currently occurs upon field reports of quality concerns, followed by manual flagging if warranted. As major quality issues are resolved, Battelle will continue to develop and expand automated algorithms to identify aberrant data, reducing reliance on manual data review.

Develop realistic, incremental benchmarks for availability of data products
Battelle thanks the STEAC for their recommendation to develop more realistic data availability targets in the initial years of operations, and this will occur following the advice and examples of the TWGs and other relevant networks/experts. Battelle will begin to share more information about data availability to end users through a prominent Data Availability page on the Data Portal over the next few months.

Re-evaluate communication of allowable and unallowable reasons for data loss.
Battelle appreciates the difficulty for the external community to interpret current reporting regarding NEON data loss. Denoting events such as computer crashes, scheduling problems, and power failures as unallowable data losses is not to suggest that Battelle does not expect these events to occur. Rather, the purpose is to distinguish types of data loss that can and should be minimized through improving the robustness of the Observatory to foreseen problems, for example by incorporating a backup power supply in the design of critical systems. Separating the unallowable losses mentioned above from allowable losses that by design are not possible or advised to reduce, such as maintenance and calibration periods, allows NEON to set more realistic targets for data availability as well as identify

areas of improvement. Battelle will work to clarify the language surrounding Table 1 in the NEON Science Availability Plan as well as to re-evaluate our targets to ensure they are realistic.

VI. Field Staff Training

Battelle appreciates the recognition of the critical importance of staff training to NEON data quality, as well as the benefit of training materials for the broader user community.

Regarding the short-term commitment of additional staff to produce high-quality training materials, Battelle has allocated 1.4 FTE in FY19 for curriculum development. This is sufficient staffing to provide annual updating of existing TOS and AOS materials and continued development of IS training materials. In the current project budget, videography is limited and of a production value that is appropriate for internal NEON staff (not external audiences). The results for the newly-implemented field audits will be used to identify if/where additional resources are needed for further development of training materials. With the recent integration into Battelle, the NEON project is also exploring how to integrate NEON training materials into the Battelle learning management system, which may have ramifications on the format/type of training resources we develop.

The NEON Strategic Engagement Plan (current version, September 2018), aspires to complete the following two activities, although these are not currently funded, under Goal 3 (Cultivate Future Users), Objective 3 (Provide NEON Data Collection Training and Resources):

- Develop online NEON data collection protocol training resources that can be used by NEON and the scientific community
- Research and develop a Field Ecology Training Course that can be offered for non-NEON staff using NEON data collection protocols

Battelle agrees that developing materials that are usable by the broader community is desirable and we will strive to include funds to move this forward in our FY2020/21 plans as possible.

Design field audits to provide actionable information.
The design, development, and implementation of audits for NEON falls under the responsibility of the Quality Assurance organization. This ensures objectiveness and the removal of undue bias from the auditing method, audit information, and trend analysis. The Quality Assurance organization is developing a NEON Audit Program to encompass protocol, field, and Domain Support Facility audits for OS, IS, AOP, Infrastructure, Data, and External Laboratory Analysis. The Audit Program will apply International Organization for Standardization (ISO) auditing guidelines and standards to ensure competency of auditors, establish methods, determine schedule, and define scope to conduct audits, including field audits. The intent of the audit program is to assess the entire life cycle of a data stream from requirements through collection to data processing, and meet the following objectives:

- Monitor and verify effective and consistent protocol implementation across Domains
- Verify Protocol and Training Objectives are being met
- Verify NEON requirements for measurement systems are met
- Ensure proficient and consistent provision of comparable data across observatory

The stages of development, implementation, and finalization will take approximately 12-18 months. The external laboratory analysis audit component is fully developed, and the OS field audit is the next

component currently under development. For the field audit development, the QA organization is working with the appropriate stakeholders in Science, Training, Field Science, and supporting areas to:

- Develop checklist and audit reporting templates for use with all protocols
- How to schedule and conduct protocol field audits during sampling season
- Determine sampling size and frequency for protocol audits
- Determine method and database for responding to and tracking findings and associated corrective actions

Prior to the end of the 2018 sampling season, QA plans to conduct a proof of concept with limited scope to attain initial audit information. This will include:

- Develop a draft method/process for conducting audits
- Develop draft audit checklists for a subset of OS protocols based on prioritization by Science and representative of sampling that occurs throughout the year.
- Perform field audits for AOS and TOS protocols at a subset of Domains to test the audit method (verification of consistency across domains).
- Use results of field audits to expand development of an OS field audit program and conduct field audits in FY19.

Battelle currently uses fulltime Field Science staff to train new fulltime staff. The challenge with this is the availability of a trainer when needed, such as when new staff onboard during the height of field season. When trainers are available, this model works well. Initial exploration of a dedicated trainer could be to have one fulltime person (based at HQ) to fill the training gaps, support/train the Field Science trainers, and conduct field audits. In the short term, it will be useful to review the field audit program to identify where the training weaknesses are.

Another model for training that Battelle will pilot in FY19, is allocating approximately 30% of a senior Field Science ecologist's time to providing remote training to colleagues at other domains. This is focusing on the instrument system, in particular, troubleshooting and problem solving. To support this real-time/on-the-job training, this staff member is developing reference materials, which will eventually be integrated into more formal training components. It is a model that perhaps would be useful for other subsystems. The challenge is back-filling the tasks that this domain-based trainer will no longer be able to do.