

Battelle Response to NEON STEAC Spring Meeting 2023 Advisory Report

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 May 15th and 17th 2023 also combines the input of several members of the program team responsible for the execution of the NEON project. Battelle is grateful for the time and expertise these members of the community have provided and appreciates the formal recommendations in their Advisory Report. The following are our responses to the key recommendations.

STEAC Recommendations:

I. NEON Updates: Provisional data

We appreciate the support of the STEAC regarding the recent published misuse of NEON provisional streamflow data. This presented an opportunity for NEON to review and make even more transparent the distinction between provisional and release data, and, with helpful input from the STEAC, we have updated our website accordingly, with emphasis on the Data Product Release and Revisions and Citing NEON pages. We agree with the STEAC that it is important to continue publishing provisional data. Big open environmental data are new to the ecological community, and a collaborative publication focused on ‘best practices for use of environmental data’ led by the community would undoubtedly be helpful.

We thank the STEAC for their interest in our work with ESIL at their kick-off Innovation Summit in May. We will provide a detailed report at an upcoming monthly meeting.

II. Demographics - New data and benchmarking

As the presentation by ESA STEAC member Adrienne Sponberg showed, and as the STEAC notes, the field of ecology does not represent the demographics of the U.S. population in terms of diversity. We agree that NEON has a special opportunity to enhance demographic representativeness through the network, particularly the seasonal staff.

In response to the specific recommendations:

- The diversity of NEON staff is currently being tracked over time.
- We understand that comparability of these data is difficult across organizations due to different ways to categorize and different levels of granularity. Battelle follows U.S. government standards in this regard, and thus the high-level demographics are comparable among entities. We will continue to evolve our metrics accordingly.

- We are currently tracking metrics related to the demographics of the audiences attending our NEON engagement activities, which include data skills workshops, seminar series, and domain-led community events. We track these audience metrics at the primary audience level, not the individual level (e.g., MSI, early career, etc.). The demographics of the broader NEON active-user community are quite difficult to track. For example, as per STEAC recommendation, NEON data users are not required to create a user profile; less than 5% of users create one. Moreover, data users who do create a profile are not required to provide demographic data; we cannot get insights into the diversity of data users. NEON Connect will provide the opportunity for the community to self-identify, e.g., attendees as they register for NEON-sponsored events.
- We make every effort to provide professional development pathways for NEON staff. Because NEON has very few permanent positions relative to the number of seasonal staff, we seek to provide training, certification, and recognition that will enable them to be successful in their careers.

III. NEON Connect, Ambassadors as mentors & resources; creating a community of helpers

We appreciate the positive comments from the STEAC regarding NEONConnect, the NEON Ambassadors program, and the NEON Code Hub, an upcoming data quality survey, and NEON internships. We are glad that STEAC members are willing to help encourage the use of NEON resources.

We value STEAC members' willingness to serve as platform testers for NEONConnect prior to release to the broader science community. Our release communications plan includes highlights of the benefits for registered users; we appreciate the specific recommendations from the STEAC.

As discussed with the STEAC, an external evaluator may be brought in to help benchmark expectations on number and quality of usage. We plan to engage an external evaluator in AY24 and work with them to prioritize the assessments they can support. This CodeHub assessment will be on that list and weighed against competing NEON evaluation priorities and the available budget.

The STEAC encouraged NEON to consider ways to identify existing community derived data products and to build linkages between researchers. The upcoming NEON Ambassador Derived Data Products workshop series is designed to further these outcomes. Moreover, through our Memorandum of Understanding (MOU) with the NSF Environmental Data Initiative (EDI), we have developed an approach. EDI serves as a repository of derived data products, and NEON will consider approaches to linking to these products from our website. We will scope out the new software development that would be needed and then prioritize and plan into the development roadmap, ideally in the next award year (AY24).

IV. NEON Biorepository Sample Use

We thank the STEAC for their discussion of ways to improve NEON Biorepository sample use by encouraging a breadth of NEON users to highlight these resources in outreach. Considering whether providing image data across the breadth of Biorepository specimens (currently not in scope) would increase sample use is an avenue that could be explored e.g., by connecting with iDigBio. To add comprehensive image data to NEON's scope would need such data to be championed by the community.

V. Meet with Domain staff re: engagement

We thank the STEAC for being open to interacting with their local NEON Domains across a variety of opportunities. We agree that interacting with co-located LTERs is a priority, and we have recently engaged more fully with them following a very productive LTER Executive Committee meeting that Dr. Mabee attended. LTER PIs expressed enthusiasm for NEON Domain staff to participate in their Annual Meetings, and Domain staff are following up on this at the co-located LTERs.

VI. Assignable Asset Program - updates and future directions

The STEAC recommendation for a more flexible Assignable Assets cost estimate model such as described would reduce the outsized staff time investment, and it is thus highly desirable for the sustainability of NEON operations. We appreciate the STEAC's vision of diversifying the types of Assignable Asset projects by NSF providing "NEON-enabled" funding opportunities across a broader set of programs. We will continue to work with the NSF to realize efficiencies and communicate STEAC recommendations, while also encouraging the STEAC and the broader community to communicate their ideas directly to the NSF.

VII. Google Cloud planning update

We fully understand that NEON's raw file structure is currently difficult to navigate in the Google Cloud Storage buckets and uses CSV files that are not cloud native. We are ready to move ahead with the STEAC's suggested longer-term solution to create a bucket that uses a cloud-native tabular format, such as parquet or avro, and we have started to actively plan this significant effort. We appreciate the STEAC's continuing support for both making AOP data available via Google Earth Engine and moving towards support for cloud-native, platform-agnostic access to NEON tabular data in buckets. As we are not alone in migrating data and communities to the cloud, we concur that learning from and with other agencies would be valuable. We have initiated engagement with Openscapes, an initiative that is currently partnering with NASA, and are actively seeking out additional groups. We remain open to specific suggestions from the STEAC or contacts from the community, and we will continue to provide updates to the STEAC on our progress in this emerging area.

VIII. OKN Proposal

Because of the limited time to develop a proposal, NEON is not moving ahead this year with a submission to the NSF Prototype Open Knowledge Network (Proto OKN)

solicitation. We remain enthusiastic, however, about leading the development of an environmental node for data discoverability by machines/AI and people, and we are moving ahead in assembling partners in service of a pilot proposal to generate the relevant knowledge graph. We appreciate the contacts that the STEAC provided, and we plan to reach out to them. We noted your suggestion that NEON keep abreast of other efforts to integrate environmental data into AI by private groups.

IX. Dust data product updates

As described in the STEAC report and minutes from previous STEAC meetings, NEON has procured feedback from the community and recommended specific user groups to carefully consider the multiple issues involved in dust data products. Given the outcomes, the STEAC recommended that NEON suspend particulate mass sampling. This leaves open the opportunity for the community to use the Assignable Assets program to conduct sampling for individual research projects. We will advance this proposal to the NSF for consideration.

X. Data Product catalog strategy

Some NEON data products are enhanced and delivered by external partners. These partner-generated derived data products are not consistently treated within the NEON data product catalog and data portal, and we seek a consistent approach to enumerating and representing this diversity. We appreciate the STEAC's attention to this matter, and particularly for their suggestion to develop a communication plan emphasizing that the raw data NEON collects remains unchanged. Further, we like the STEAC recommendation that NEON uses a naming convention that is descriptive and includes tags from the code hub to link the product with relevant code. Note that the latter will require new software development for the data portal, thus it will need to be prioritized and planned into the development roadmap. Also, given that partner organizations will maintain responsibility for communicating any changes in the product, we will adopt the STEAC recommendation that the strategy include methods for indicating major changes in derived data products. Our primary approaches for such communications include the Data Notifications on our website that users can subscribe to and the Issue Log included on each data product landing page. Finally, we will carefully consider the STEAC suggestion that we develop a new Technical Working Group to develop a strategy for derived data products developed by the wider community. We note that the Environmental Data Initiative (EDI) is currently the strategy, and that we work closely with EDI as part of our MOU.