STEAC MEETING REPORT
(05/15/2023 and 05/17/2023)

The STEAC met on May 15th and 17th, 2023, with a quorum of thirteen members attending (Meghan Avolio, Henry Bart, Sarah Bevins, Michael Dietze, Rich Fiorella, Shannon LaDeau, Sparkle Malone, Jackie Matthes, Kim Novick, Steve Petruzza, Sydne Record, Daniel Rubenstein, Shawn Serbin, Sydne Record, and Adrienne Sponberg). Nineteen NEON-Battelle staff attended (Teresa Burlingame, Chris Florian, Nico Franz, Beth Fugate, Kristan Godfrey, Darcy Gora, Bridget Hass, Steve Jacobs, Samantha Kremidas, Mike Kuhlman, Christie Laney, Claire Lunch, Paula Mabee, Becky Miller, Chris McKay, Bonnie Meinke, Mike SanClements, Kate Thibault, and Rommel Zulueta).

The meeting was virtual, and the following topics were discussed: I. Approval of the minutes, II. Welcome and NEON Updates, III. Demographics - new data and benchmarking, IV. NEON Connect, Ambassadors as mentors & resources; creating a community of helpers, V. NEON Biorepository Sample Use, VI. Meet with Domain staff re: engagement, VII. Assignable Asset Program - updates and future directions, VIII. Google Cloud planning update, IX. OKN Proposal, X. Dust data product updates, and XI. Data Product catalog strategy.

I. Approval of the minutes. Minutes approved

II. Welcome and NEON Updates

At the meeting there were no new updates from NEON regarding the status of the competition for the management of the NEON facility. The STEAC was updated that the National Science Board had met the previous week but no new information was provided. Because of this, the STEAC was informed that any ongoing recruitment or similar activities remained on hold until the new management contract announcement is made. Similarly, it was decided that in the meantime STEAC members who were planning to rotate off in the fall could be asked to stay on to avoid creating gaps in STEAC membership. At the same time, NEON noted that in the event that NEON management were to transition away from Battelle, it is not clear what the new oversight would look like and whether the current STEAC would be dissolved.

NEON also provided an update related to an ongoing discussion with the STEAC related to the misuse of NEON provisional streamflow data in a publication. STEAC was informed that NEON had been in touch with the journal editor to discuss options to request an author correction to conclusions drawn in the paper, as well as to correct the data citation (DOI) in the publication to point to the provisional data used in the paper. It was noted that the authors had not provided the provisional data DOI, but instead the release data DOI, so manuscript reviewers would not have known the study was based on provisional, not final QA/QC, data. NEON noted that the journal also does not allow for published rebuttals so the options for correcting the publication record are limited but they are still exploring options for updating the published online version to include the correct provisional data DOI, in which the results are based, in the article. This request has been made by the journal editor to the article authors and NEON has asked that the data in the paper be referred to as "provisional" in the revised article corrections. The STEAC recommended that instead of focusing too much on this article, that NEON improve their communication of the use of provisional data and make it easier to identify when and if provisional data is included in a data download, particularly when using the NEON data API. NEON informed the members of the STEAC that they are updating their data availability and author usage statements to provide better guidance on when and how to report what data and data versions are used in future studies. NEON also informed the STEAC that they are working

to make reporting of provisional data more obvious to end-users. But the STEAC also commented that because there are many provisional data users it is important that access to these data remain open and instead that new information be included in the documentation when using provisional measurements. The STEAC also recommended that there could be a community effort, perhaps led by the STEAC, to publish a "best practices" article regarding how to properly use open environmental/network data, like NEON, to inform the community and new users as to the possible issues they should be aware of.

NEON informed the STEAC that they had recently provided new documentation to the NSF related to their plans for implementing new training, including sexual harassment training, and a Code of Conduct. The goal is to foster an internal culture to try and minimize the risks of harassment and retaliation. NEON expects this to be a 2-page document laying out the framework for fostering a safe work environment. One challenge that was noted is that existing Assignable Asset PIs will not be covered under NEONs Code of Conduct and would be expected to operate under their own institutions Code of Conduct and/or safe and inclusive work environment regulations. However, new Assignable Asset PIs will be provided with NEON's Code of Conduct and are expected to act in accord with it. The STEAC applauded this effort and noted that they expect this will help increase diversity in the field by opening up new safe and inclusive options for underrepresented communities who have traditionally felt excluded or uncomfortable in the field.

Finally, during the opening statements NEON also informed the members of the STEAC that they had been invited to participate in the upcoming Environmental Data Science Innovation & Inclusion Lab (ESIIL) Innovation Summit. NEON noted that they are putting in a lot of energy into the event so that NEON participates in the development of future inclusive research efforts and can incorporate new suggestions and best practices into their ongoing operations. The STEAC applauded NEONs participation and requested that NEON provide a detailed report back to the STEAC after the event.

III. Demographics - New data and benchmarking

The STEAC applauds NEON's initiative to critically evaluate data on the demographic diversity of their staff, using these data as benchmarks by which to measure progress on efforts to enhance DEI across the network. To an extent, NEON data are aligned with the demographics of ecology writ large. The field of ecology has a lot of work to do to increase diversity, inclusivity, and equity for underrepresented identities at all career stages, and to be mindful of closing persistent gender gaps in senior positions.

The STEAC offers the following recommendations to NEON to enhance demographic representativeness through the network:

 1)   The benchmarking of NEON staff demographics should be a dynamic (in time) initiative to evaluate how representativeness has changed historically, which could guide interpretation of future analyses of NEON staff diversity.

2)   NEON could consider disaggregating the diversity data to capture more dimensions of diversity (e.g., less binary classifications, and including axes for disability status, socioeconomic status, and other important diversity metrics). Here, continued coordination with agencies like the Ecological Society of America (ESA) would be particularly advantageous for evaluating how network-level metrics compare with the field at large.

 3)   In addition to an analysis of the demographics of the NEON staff, the STEAC urges

NEON to develop systems for benchmarking the demographics of the broader NEON community. For example, NEON can track the diversity of NEON data users, attendees at NEON-sponsored events, speakers in the NEON webinar series, scientists who engage with NEON's social media platforms, among other examples.

4)   Thinking systematically about "pathways" for NEON careers through which underrepresented staff (for example, hired through internship or seasonal tech positions) could advance professionally through the organization into permanent and/or full-time positions.

The members of the STEAC encourage NEON to recognize its role as a leader in the field of ecology. NEON is in a position to push the field forward by developing goals that seek to not only build a workforce that matches the demographics of the pool, but that also seeks to move the needle on representation of the pool itself.

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

NEON updated the members of the STEAC on their community engagement programs, including NEON Connect, the NEON Ambassadors program, NEON Code Hub, an upcoming data quality survey, and NEON internships. NEON Connect is a new community engagement and tracking platform that will provide a central location for interaction between NEON and the community. The platform has single sign-on capability which will allow data portal users to link their two accounts. Users may also keep their accounts separate if they do not wish to share their information. NEON Connect is in Phase II of production which includes testing and familiarizing NEON staff with use of the tool. The next phase - rollout to all users – is anticipated in Fall 2023. The members of the STEAC commend NEON for their progress in developing the platform and are willing to serve as platform testers prior to release. The members of the STEAC reiterate its earlier recommendation that NEON Connect highlight the benefits for registered users when this new engagement platform is launched (e.g., individualized data issue notifications, access to workshops and the assignable assets program).

The members of the STEAC recognize that the NEON Ambassadors program is off to a very successful start with a series of virtual workshops taking place soon. The first cohort of Ambassadors is planning to submit a NSF proposal in September for a follow-up in-person workshop focusing on derived data products. The second cohort will be recruited in September 2023.

Staff presented metrics on NEON Code Hub and noted an external evaluator may be brought in to help identify the pathway to understanding what code hub should be and how many people should be using it. STEAC concurs that an assessment of this sort would be valuable. The members of the STEAC encourage NEON to continue to consider ways to identify existing community derived data products and to build linkages between researchers. It might be helpful to have targeted outreach to known code developers or people who have NEON-related grants and to link from data product pages to code and derived products.

V. NEON Biorepository Sample Use

NEON provided an overview of the current status of the biorepository, which houses over 367,000 samples and encompasses more than 3,100 taxa. Among the various categories, bulk

terrestrial invertebrates are the most frequently utilized data, followed by carabids, soil microbes, small mammals, mosquitos, and bulk aquatic invertebrates. On the other hand, wet deposition samples are the least utilized due to their short shelf life, plant vouchers lack the necessary replication within species for applications, and aquatic sample types suffer from sparse spatial coverage and low replication.

Over the past few years, there has been an increase in the number of inquiries and samples used. However, it is not anticipated that there will be a further increase in 2023. The Global Biodiversity Information Facility (GBIF) is employed to track citations from research that has accessed this data. Presently, NEON has 229 citations on GBIF, with plant vouchers, mammals, and carabids being among the most frequently cited, possibly due to their improved taxonomic resolution.

It is worth noting that the majority (two-thirds) of sample use is not funded by the National Science Foundation (NSF). To ensure proper citation of this data, the NEON team has created a 'how to cite' page for training users. Additional updates include expanding the Symbiota Application Programming Interface (API) and enabling the search for biorepository products through the Environmental Data Initiative (EDI) portal. Moreover, the biosample services offered by the NEON Biorepository are now advertised on the BioKIC website (Arizona State University). Lastly, there is an unpublished manuscript describing the design and scope of the NEON Biorepository. A new federally funded Center on Biological Collections is being planned by the community as a result of the CHIPS Act.

The members of the STEAC discussed ways to improve sample use in an attempt to meet the performance benchmark within NEON's operations. The STEAC recommended engaging early career researchers and NEON ambassadors to organize workshops, presentations to REU (Research Experiences for Undergraduates) groups, bootcamps, and incorporating use of sample data into class projects. Furthermore, the potential for challenges using biorepository data were discussed, with the caveat that it may be more feasible when more image data becomes available. Finally, tracking and understanding data use could be used for targeted advertising to existing biorepository users but also to communities that might not be aware of its existence.

VI. Meet with Domain staff re: engagement

Bonnie Meinke and Kirsten Ruiz led a discussion with the goal of building potential networking bridges and facilitating engagement among STEAC members and their local NEON Domain sites. Four Domain Managers were also present: Beth Fugate (Domain 8), Becky Miller (Domain 6), Christin Godfrey (Domain 1), and Amy Jacobs (Domains 15 & 16). Many STEAC members present noted ongoing in-person use of a NEON site and several members currently use data from more than one domain. Bonnie shared a hope that STEAC members would further consider possible individual actions, including arranging a tour of a NEON field site, facilitating a meeting of local colleagues to a NEON site, or looping domain staff into relevant events hosted externally. Several STEAC members identified opportunities to interact with staff at their local NEON Domain sites. The STEAC discussed additional ways that NEON domain staff could extend outreach beyond their current local contacts, but concluded that effective and relevant activities would need to be identified at the domain level. Building on existing breadth of expertise and participation at LTER sites that are co-located or near NEON Domain sites should be prioritized. Domain staff are encouraged to participate in Long-Term ecological research (LTER) Annual Meetings at sites that are co-located with NEON sites. STEAC members were interested in learning more about how NEON Connect might help record and share success of

engagement opportunities in the future.

VII. Assignable Asset Program - updates and future directions

The members of the STEAC were pleased to learn about recent improvements and initiatives for the Assignable Asset Program, which facilitates the use of NEON infrastructure and the time and expertise of NEON personnel at cost for researchers external to the NEON program and at no profit to Battelle. This is an essential program for the scientific community to utilize the power of NEON's continental-scale resources and the program benefits NEON through additional data collection and research activities. Through time there has been an increasing number of requests to the Assignable Assets program as principal investigators include NEON resources in their grant applications.

Recent developments in the Assignable Assets program include development of the Small Project division, which well-aligned with previous STEAC feedback. The Small Project division balances resource accessibility and the work of the Assignable Assets staff for requests below $5k and where internal NEON funding may be combined with grant or university support. The small projects in particular make NEON resources more accessible for a wider range of researchers (including students) at relatively low cost to NSF/Battelle.

Another exciting development is a public-facing site and project database development that will detail past and current Assignable Assets projects. In the future these will be featured on the website in a searchable database, which will provide a wide range of examples for the scientific community and could prompt new ideas for projects. This database will also provide information regarding the landowners of NEON sites, which fall into a myriad of different categories (e.g., federal, state, private land ownership) and will help to facilitate the permitting process for non-NEON researchers to do research at NEON sites.

The STEAC was encouraged to hear that NSF continues to work with NEON to provide an accessible and efficient Assignable Assets program, since this is a unique resource in the ecological and environmental research community through which researchers can engage with NEON. The Assignable Assets staff carefully consider proposals as they are shepherded through the planning, agreement, and implementation phases, with substantial and consistent interactions with researchers. The STEAC recommends that a more flexible Assignable Assets cost estimate model (for example, providing an estimate at the proposal stage with some flexibility for adjustment by NSF given a thoroughly detailed estimate if the proposal is funded) with NSF and other major external funding agencies might both increase the number of proposals that use the Assignable Assets program while making the time of program staff more efficient. Another recommendation is to encourage the NSF to make "NEON-enabled" funding opportunities more clearly available across a broader set of programs beyond macrosystems biology. This might diversify the sort of projects that make effective use of NEON data, and would also avoid situations where the majority of NEON Assignable Assets funding comes from a single program within NSF.

VIII. Google Cloud planning update

NEON presented an update to the STEAC on their migration of services to the Google Cloud. This discussion focused primarily on three aspects of this migration from an end-user's perspective: Google Earth Engine for AOP data, Google-specific cloud services for tabular data, and "cloud-agnostic" technologies for tabular data. For Google Earth Engine, NEON reported a plan for an initial release of five AOP products and work on a tutorial series, webinar, and ESA short course. For Google-specific services, NEON is exploring internal uses of Google

BigQuery, including a Jupyter notebook demo in Google CloudRun, and hosting internal Shiny apps in Google's cloud. At the STEAC's previous recommendation use of BigQuery has not yet expanded to externally-facing data. For cloud-agnostic technologies, NEON reported that feedback from the TWGs (primarily the Ecological Forecasting TWG, plus 3 individuals from other TWGs) supported a focus on direct bucket access to tabular data. Currently NEON's raw file structure is difficult to navigate and is based on CSV files that are not cloud native. A longer-term solution would be to create a bucket that uses a cloud-native tabular format, such as parquet or arrow. There was also a small update about the migration of NEON's internal pipelines to the cloud and some of the cost trade-offs in how Google charges for some of its own internal services versus the cost of running open-source solutions in containers within Google's cloud.

Overall, the STEAC reiterates its prior 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. We also recommend communicating with other agencies (e.g., NASA, NOAA, USGS) and NSF major projects that have prior experience in cloud migration.

IX. OKN Proposal

Christine Laney from NEON presented a potential NSF proposal for the Prototype Open Knowledge Network (Proto OKN) request for proposals. The solicitation goals are to create interconnected network of knowledge graphs of application domains to support AI development, to support AI powered solutions to address complex challenges, to make public data more amenable to complex analyses to increase government and non-government users, and to tie in such analyses with the National Information Exchange Model (NIEM). There are various potential partner organizations for this solicitation (e.g., NASA, NSF, NIH, NI Justice, NOAA, USGS). The RFP describes three themes of awards. Theme 1 awards will develop a singular network graph or node to address a specific use case, uses well-identified data sets, follow a sustainable user design process, and work together with a Theme 2 fabric team. Theme 2 awards will provide an interconnecting fabric to link Theme 1 awards. Theme 3 awards provide education and public engagement of potential users from Theme 1 and 2 awards. In terms of the logistics, the proposal is due June 20th with projects starting in October of 2023. The awards are for $1.5 million dollars over 3 years.

NEON is considering putting in a proposal to fill a perceived gap in terms of environmental science. The proposal would be for machine actionable resources to understand and prepare for environmental change. The project would build a mapped ontology of search terms to discover ecological data across NEON and the Environmental Data Initiative with an emphasis on water resources. Existing work that could be leveraged for this project include the LTER/EDI working group variable naming from the ecocomDP data design package for community ecology data, NEON/LTER variable names mapped to each other and to existing ontologies, and NEON/Ameriflux variable name mapping. This would be a pilot proposal to generate a knowledge graph spanning data from numerous partners that is ready to be incorporated into NIEM and would have an emphasis on ethics given potential AI uses (e.g., sensitivity of traditional ecological knowledge and CARE data principles).

Christine asked the members of the STEAC for thoughts about this proposal regarding the usefulness and sustainability of this type of project, the short timeline for the proposal to come together (~one month), and other opportunities to do something like this in the future. Paula emphasized that this proposal would enable NEON to be discoverable by machines and people working with AI. The members of the STEAC provided some contacts to consider reaching out

to about such work (e.g., Jake Zwart with USGS, the Climate Forecast Convention, the DOE ESS-DIVE). The members of the STEAC also suggested that NEON keep abreast of other efforts to integrate environmental data into AI by private groups (e.g., the Musk Foundation and X-PRIZE). The STEAC also recommended that NEON consider how efforts invested in this proposal might be leveraged in the future if the funding for this proposal is not awarded.

X. Dust data product updates

Dust and particulate sampling is critical to ecology and human health. NEON's intention was to sample dust across the great plains. NEON deploys optical particulate matter analyzers at a total of 6 sites across three Domains (10, 13, and 15) to help the scientific community gain insight on the regional dust transport across the Rocky Mountain region. Aerosol dust can be composed of numerous inorganic and organic elements; everything from biological components such as pollen to the byproducts of incomplete combustion. The analyzers measure particulate mass and size. However, measurements by the instruments have been problematic, with frequent data quality flags based on QA/QC tests. Feedback from the community suggests that NEON's dust sampling is not particularly useful. Small dust particles pass through filters. There were also problems with loss of expensive lids that protect one of the sample size collectors. NEON's Dust Sampling team completed an assessment of data flagging cause and has recalibrated its procedures. NEON followed the recommendation of STEAC to reach out to user groups to discuss the issues, including EPA's Clean Air Status and Trends Network (CASTNET) program, which is interested in total deposition, and NADP NTN, which focuses on wet deposition. CASTNET's dust sampling is designed to capture pollutants. The sampling is currently undergoing scientific review. CASTNET expects a 10-20% reduction of sites due to funding cuts. CASTNET looks at areas of high total deposition uncertainty, coastal regions and wildlife tracts. NADP's Total Deposition (TDEP) Science Committee is most relevant to NEON's total mass data, including total nitrogen and phosphorus.

The discussion with CASTNET and NADP proved to be very helpful. It was discovered that there's no community standard for collecting dust samples and dust sampling is not straightforward. Particulate mass filters do not capture primary inputs of nitrogen. NEON would like to set up a technical working group to discuss dust and particulate dry and wet deposition to improve data quality for particulate size classes, review data and protocols, and collaborate with researchers to ensure usability. The members of the STEAC recommend that NEON suspend particulate mass sampling and encourage Assignable Assets sampling to address scientific needs.

XI. Data Product catalog strategy

NEON works with other networks and organizations that develop derived data products with NEON data. Derived data products may be hosted solely by a repository or jointly with NEON. Currently, NEON inconsistently represents partner-generated derived data products. A strategy for derived data products was presented to describe the scope of the products NEON is enabling and to maximize discoverability. For consistency among partners NEON will delete the 6 gap-filled AmeriFlux data products that are currently in the NEON data product catalog. Each external partner will be assigned a new data product in the catalog, that represents the full suite of data products each partner generates. Although this strategy will create consistency across external partners, in the short-term this will reduce the number of data products in the catalog to 183 (soon to be 184). The *pros of this approach is that it will p*rovide consistency, increase discoverability of curated products, provide credit to NEON via the incremental addition of partners through time, and reduce the complexities of trying to manage the specifics of the derived products partners generate.

The STEAC applauds NEON's attention to enhancing the development and accessibility of derived data products from partner organizations. Increasing the visibility of derived data products developed by partners allows NEON to leverage the work of the community. The STEAC recognizes that the NSF is focused on the number of data products, as NEON has used this as a metric of success. Therefore, a communication plan to communicate that this change has not impacted the data collected or the number of NEON derived products is key. With the removal of data products from the catalog, it is important to have a review and sunset policy in place.

The STEAC recommends that NEON use a naming convention that is descriptive and include tags from the code hub to link the product with relevant code. The STEAC expressed some concern with using a single catalog number for a data product that may change. Since NEON is not hosting the data, users have to get information on data changes from partner organizations. Partner organizations will maintain responsibility for communicating any changes in the product. The STEAC recommends that the strategy include methods for indicating major changes in derived data products.

The strategy for derived data products of partner organizations is an exciting step for NEON. The members of the STEAC hope this will help lead to a solution for 'freestyle' data products, i.e., ones developed by the community. NEON might consider starting a new technical working group to develop a strategy for derived data products developed by the wider community.