

NEON STEAC Spring 2021 Meeting Report

The STEAC met for the bi-annual meeting on April 19, 2021 and April 21, 2021. The meeting was virtual, and the following topics were discussed: I. [Draft NEON Participation Guidelines](#); II. Indicating active NEON projects in the community to reduce redundancy; III. Incentivizing NEON data citation; IV. Domain engagement strategy updates; V. NEON-EDI; VI. NSF Q&A; VII. [Visiting scientist](#) program; VIII. [NEON Biorepository](#); VIII. NEON Staff; X. NEON Staff Training; IX. STEAC Nominations.

Major STEAC Recommendations:

- Participation guidelines: The participation guidelines can be improved by defining the NEON community, specifying when and where guidelines should be followed, describing kind, professional behavior, and having a designated contact to address complaints at all NEON events.
- Fostering communication to reduce project redundancy: Amplify and advertise collaborative workspaces (i.e. Slack channels) for NEON data users, for example through coordination with the TWGs and by linking to virtual workspaces on the NEON website.
- Data citation: Consider devoting at least some resources to encouragement/enforcement of proper citation of NEON data sets. Outreach to journals, NSF program officers/Data Management Plans, updated data citation guidelines, and modifications of the API and R packages could produce useful results without a major expenditure of time.
- Engagement: The STEAC commends the development of Domain Specific Engagement Plans and encourages NEON to track metrics that help identify which activities have the greatest impact.
- Visiting Science Program: The STEAC recommends that NEON proceed with the Visiting Scientist Program, look for opportunities to provide need-based funding, and consider opportunities for visiting scientists to reside at NEON facilities other than NEON Headquarters.
- Biorepository: The STEAC sees good progress and recognizes that in time the repository will be fully discovered and more highly used. The STEAC recommends that NEON think about how to reach out to additional communities likely to leverage these resources (e.g. systematists).
- Training materials: The STEAC is supportive of the idea of NEON sharing training materials with the wider community.

I. [Draft NEON Participation Guidelines](#):

NEON presented participation guidelines that communicate a commitment to developing an inclusive community. The guidelines require all NEON community members to agree and adhere to a code of conduct, in order to create a safe and positive community experience for all. Through these guidelines, individuals can report misconduct by phone or by submitting a form. The complaints will be reviewed and NEON and/or Battelle will determine the necessary response to the complaint.

The STEAC recommends a few improvements to participation guidelines: 1) Clearly state who is included in the NEON community; 2) Identify when and where the NEON community must follow these guidelines; 3) Under "Expected behaviors" describe kind, professional behavior; 4)

Make a statement that illegal behaviors will be reported to law enforcement; 5) Clearly state what will happen to violators of professional guidelines; 6) All NEON events should have a designated contact to address complaints; and 7) Consider [professional accountability](#) programs and [Bystander Training](#) to help create teams trained to intervene and de-escalate hostile situations.

II. Indicating active NEON projects in the community to reduce redundancy:

A persistent challenge in network-enabled science is preventing unnecessary duplication of effort, while at the same time recognizing that some replication and redundancy may benefit scientific discovery. This challenge is particularly important when thinking about early career scientists, who may have a lot to lose if their research goals are being pursued simultaneously by other teams, and especially if those other teams have more resources and can move more quickly. Many other networks have wrestled with this challenge, which is compounded for more bottom-up networks by the fact that data suppliers and end-users may be the same entities. Networks like FLUXNET have experimented with strategies for preventing duplication of research effort such as project registries, but these can have unintended side effects (i.e. project squatting).

In general, the STEAC does not view it as NEON's responsibility to carefully supervise the use of the data to prevent duplication of effort. However, the NEON project would benefit from supporting community-driven efforts to better connect the "network of individuals" using NEON data, for example through SLACK channels and direct or indirect support of workshops for disciplinary communities of practice. The STEAC notes that the recently formed Aquatic NEON Slack (which already has dozens of active users), and the well-organized NEON EFI community Slack (>350 users), provide substantial examples of successful mechanisms for connecting NEON users across career stages. NEON staff can support these efforts by either formally hosting slack channels, or advertising bottom-up efforts on their website. The majority of the STEAC supported the latter option, as NEON-sponsored Slack could create the expectation that NEON staff would have to moderate the channel. At the same time, the STEAC felt the existing NEON community message board has not been able to reach critical mass (similar message boards often needed >10k users to be effective) and NEON could consider descoping it in favor of amplifying community Slack channels. The TWGs may be able to play an useful role in creating and amplifying the work of virtual collaborative spaces. It will be important to understand the extent to which the new Open Environmental Data synthesis center can play a role in facilitating virtual and in-person platforms for collaboration within these communities.

III. Incentivizing NEON data citation:

Proper citation of publicly available datasets has been an ongoing challenge for research networks. The Long Term Ecological Research (LTER) network has been struggling with this issue for 40 years. The situation is poised to improve, however, as datasets are more clearly and readily available, document object identifiers (DOI) are assigned to all data sets, and journals are taking a more active role in encouraging/enforcing author citation of datasets. The American Geophysical Union (AGU) has a newly formed Community of Practice for Data Citation in the Earth, Space, and Environmental Sciences (and related sciences). This community of practice emerged from a discussion of "Why Is Citing Data Still Hard?"

Some organizations dedicate resources to encouraging/enforcing citation. For example, the Global Biodiversity Information Facility strongly encourages citation of dataset DOIs and actively

checks papers and interacts with authors to make sure this is done properly. Each LTER site has an Information Manager and many devote at least some time to checking papers for proper citation.

The STEAC discussed several ideas for fostering progress in data citation. The STEAC recognizes that working with journals is an obvious approach, particularly if done in partnership with other data providers and repositories, and NEON's size and duration make it well-positioned to take a leadership role in such an effort. The STEAC also encourages NEON to engage with the NSF on this issue. Data Management plans are required on all proposals and these may create opportunities for follow up, perhaps with some alteration of reporting.

More generally, there is a need to foster continued improvement in the "culture" of data citation. Similar to GBIF, NEON should consider dedicating person-hours to actively checking papers and following up with authors. More broadly, emphasizing the potential of proper data citation to increase the citation rate of papers (there is some anecdotal evidence for this), network analysis of dataset DOI numbers to determine which datasets are used most often together or function as crucial network nodes, and constant reminders in data portals may be useful for facilitating continued progress. Similarly, given that many journals limit the number of citations allowed in a paper, the NEON data citation guidelines <https://www.neonscience.org/data-samples/data-policies-citation> should be updated to provide information about alternative places to reference NEON data products and DOIs outside of the "Literature Cited". Given that most NEON data downloads are done through the API, NEON might also consider ways to make it easier for users to assemble the citations/DOIs for the datasets they've used. For example, this could take the form of a function in NEON's `neonUtilities` R package, or the community's `neonstore` package (the latter of which provides a greater degree of provenance tracking, and thus might be well positioned to report DOIs on the datasets actually used in any specific analysis).

IV. Domain engagement strategy updates:

Recent domain engagement efforts at NEON have focused on increasing local scientific interactions and determining how best to interface with other ecological networks (e.g. LTER, LTAR, and CZO). Local and regional science projects and outreach are domain-led, with headquarters facilitating as needed.

The STEAC commends the development of Domain Specific Engagement Plans. Domains conduct six activities each year. Workshop attendance, presentations, and panel participation have increased since the program began. One of the strongest engagement activities is the TEX Fellow Program, which is a great example of grassroots science that builds relationships with local communities while also providing career development opportunities within NEON.

The STEAC encourages NEON to continue emphasizing the importance of outreach and championing the TEX Fellow Program, but also to be cognizant of the additional workload (data collection needs to remain the top priority for domain staff) and track metrics that help identify which activities have the greatest impact. A communication channel that solicits periodic feedback from the domain managers to determine if there is anything they should be doing less of could help identify workload stresses as more responsibilities are added over time.

V. NEON-EDI:

NEON and the Environmental Data Initiative (EDI) have a memorandum of understanding to cooperate on three main goals that connect to FAIR (Findability, Accessibility, Interoperability,

and Reuse) principles: 1) build a diverse and inclusive community, 2) develop and recommend best practices and standards, and 3) improve interoperability and discovery of datasets. In that capacity, NEON and EDI have collaborated on the development of an R package and data standard, called *ecocomDP*, that facilitates data discovery and integration across biodiversity datasets from the LTER program and NEON. This particular cooperative effort arose from research needs that were highlighted at the NEON Science Summit in fall of 2019 by research groups. The package was based on an existing workflow for aquatic macroinvertebrate data, and extended to capture a much broader suite of biodiversity metrics for LTER and NEON data.

The STEAC agrees that data synthesis meetings like the NEON Science Summit are key opportunities where interactions between NEON staff and researchers can draw out the needs of the research community and help to steer the development of workflow products that facilitate NEON data use and integration. The STEAC was glad to hear that similar efforts are underway for the integration of instrumental data across LTER and NEON and that NEON staff are in communication with other synthesis groups about opportunities for collaboration.

VI. NSF Q&A:

STEAC met with Roland Roberts from NSF. There were two main topics:

- 1) While NEON data is made publicly available quickly, data from Assignable Assets are only delivered to the individual PI whose grant funded the collection, with no set policy for data release. This policy applies even when the funding source is public (e.g. the large majority of AA projects are currently NSF funded), and even though the data collection leverages NEON infrastructure and personnel. Also, at the moment there is no clear way that this data is FAIR (findable, accessible, interoperable, or reusable) on the NEON data site, yet this data may be very valuable and adds a lot of value added to the NEON site. Roland responded that there is an NSF expectation that the data will become available within some time period. The time period depends upon the area of research. He suggested we look at other networks. For example, the LTER network makes data public after 2 years, except for student collected data. Roland also commented that NSF requires a data management plan but does not often dictate the repository. Roland Roberts also said that NSF would be interested in the STEACs recommendation on repositories and metadata for the assignable assets. Multiple members of the STEAC clarified that the FAIRness of AA data was of greater concern than the length of the data embargo period.

While NEON/Battelle does not have authority to set/enforce AA data policies that usurp those of the funding agency, the STEAC recommends that NEON explore standards that AA PIs could voluntarily adopt that would allow such data to achieve a higher level of FAIRness. NEON should point users to EDI or some other open repository that is interoperable with NEON so that while NEON does not have the responsibility for storing the data, the existence and location of AA data could be made discoverable from within the NEON CI. A TWG could be tasked with developing standards for metadata that would facilitate the discoverability of any AA data collected at a NEON. Similarly, the default expectation on an AA project should be that data are delivered to PIs using formats, names, and units that maximize interoperability with NEON data. Embargo time periods should follow community standards, such as those at LTER sites.

- 2) Tracking NEON data continues to be a challenge. Community leadership should encourage data citation, make suggestions on how to cite data, and give examples. But while leadership can encourage, be transparent, and offer guidance, ultimately it is the

communities' responsibility to force/encourage people to cite the data. The community may be better able to set these standards.

At this moment the STEAC has no additional recommendation beyond the previous discussion on citing NEON data (See Sec III).

VII. [Visiting scientist](#) program:

Discussion at the Spring meeting did not focus on the specific language of the draft Visiting Scientist Call, but did focus on several topics that could inform the final language as well as the timing of the call.

The STEAC discussed whether the pending NSF Center for Advancement and Synthesis of Open Environmental Data (OED) reduced the need for a NEON Visiting Scientist Program or affected the timing. Should NEON look to the OED as a mechanism to engage scientists? Should the Visiting Scientist Program be tabled for now and re-evaluated once the new Center is established? The STEAC viewed the new Center as an important future partner with NEON but that NEON should not look to the Center to meet the all of the goals of the Visiting Scientist Program. Moreover, the Center will not begin operation until 2022.

The STEAC recommends that NEON proceed with the Visiting Scientist Program right away. In addition we recommend providing need-based funding that visiting scientists could apply for to help offset travel and housing costs. Furthermore, this funding mechanism could be targeted to increase the diversity of NEON data users. Gauge the level of interest and then re-evaluate whether additional NEON funding will be needed to attract a diversity of strong applicants from different career stages and types of academic institutions. The call should highlight opportunities to reside at NEON facilities other than NEON Headquarters, for example, at the NEON Biorepository. The call should also encourage Visiting Scientists to participate in synthesis working groups (e.g., new OED center, LTER, NCEAS, Powell Center), where they can share their experience and acquired knowledge of NEON data and operations.

VIII. [NEON Biorepository](#):

Nico Franz provided a virtual tour of the Alameda Building showing the NEON Biorepository. He also showed the new portal and the sampling and exploration tool developed to manage the flow of shipments and transfer manifest level into ["Darwin Core standard"](#) occurrence data.

The repository has also developed a new search tool to make the portal simpler and more visible by utilizing tokens to keep track of search parameters and finding samples housed outside of the repository at other NEON sites. The repository will make some of the upstream sample/data/loan assessment steps more mechanized at the portal through dashboards, news feeds, etc. There are currently 35 data sets with [GBIF](#). Nico Franz is confident the data linkages and data publishing cultures (Biorepository to {NEON main, iDigBio, GBIF, etc.}) will continue to be better harmonized. The data includes some basics of the NEON protocols in the occurrence records, and the protocols are listed and summarized in the collection descriptions that get published.

The STEAC sees good progress and recognizes that in time the repository will be fully discovered and more highly used. NEON should consider having visiting scientists at the Biorepository in addition to HQ. Publizing that NEON is part of the DOE FICUS call can also help. There also could be more outreach to the community of systematists as they might not be

looking at Macrosystems for funding. NEON may consider highlighting the repository on social media. Tool development should remain a priority.

VIII. NEON Staff:

As always, the STEAC members enjoyed the chance to hear from NEON domain and headquarters staff over lunch. Across all the groups, it seems that NEON staff morale was high, and NEON employees felt supported during COVID. One group had an interesting discussion about outreach and engagement at conferences and other professional meetings. Events and spaces that allow for deep engagement and cross-cutting interactions across multiple NEON staff and end-users may be more impactful at this stage than conference booths or more one-way presentations of information (e.g. from NEON staff to passive audiences).

The STEAC recognizes the importance of building communities and that networks of people are very important in this phase of NEON development. Having periodic meetups of NEON users at conferences etc. could help build these communities.

X. Training

NEON presented an update on their current training program for field crews, which is an essential part of NEON providing consistent data across sites and through time. The program is well developed and utilizes learning management software to organize and track training, as well as an extensive suite of trainings for trainers. The entire training program is impressive. One of the key questions presented to the STEAC was about interest from the ecology community to make these training materials available more broadly.

The STEAC is supportive of the idea of NEON sharing training materials with the wider community. Such an effort has multiple positive outcomes, such as encouraging the collection of more standardized, interoperable, and NEON-compatible data across the community, helping users better use and interpret NEON data (i.e. details about protocols), and being used more broadly in the training of future ecologists (e.g. in field or lab courses). NEON might also consider opportunities to offset the costs of training by charging for training. This could take many forms (e.g. in person, synchronous hybrid, asynchronous remote), either in community-dedicated sessions or piggybacking on staff training, and may include instructor training. We would encourage NEON to develop some initial test cases (e.g. options that are in high demand or topics that are low hanging fruit because they would be easy to convert); the STEAC looks forward to hearing more about these options in future meetings.

IX. STEAC Nominations:

The STEAC considered 13 nominations for STEAC membership. Per our Bylaws, each nominee was first discussed individually and the STEAC voted on whether each individual was suitable (n=9) or not (n=4). Suitability was determined based on conflicts of interest and the topical needs of the STEAC. After this the STEAC deliberated, and voted on, the rank order of the remaining candidates for each of the three focal areas where there are recruiting needs: remote sensing (4 nominees), organismal biology (3 nominees), and cyberinfrastructure (2 nominees). This list was sent to Paula Mabee and Mike Kuhlman, but is redacted from this report. The STEAC also noted that for future recruitment we would like to be more deliberate in reaching out more broadly to potential nominees. Finally, as we'd run over time, we decided to push to a future meeting our discussion about term lengths and STEAC members scheduled to rotate off.